# Visual inspection for the diagnosis of cutaneous melanoma in adults

## Review information

**Review type: Diagnostic test accuracy**

**Review number: #164d**

### Authors

Jacqueline Dinnes[1], Jonathan J Deeks[1], Matthew J Grainge[2], Naomi Chuchu[1], Lavinia Ferrante di Ruffano[1], Rubeta N Matin[3], David R Thomson[4], Kai Yuen Wong[5], Roger Benjamin Aldridge[6], Rachel Abbott[7], Monica Fawzy[8], Susan E Bayliss[1], Yemisi Takwoingi[1], Clare Davenport[1], Kathie Godfrey[9], Fiona M Walter[10], Hywel C Williams[11], Cochrane Skin Cancer Diagnostic Test Accuracy Group[1]

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[2]Division of Epidemiology and Public Health, School of Medicine, Nottingham, UK
[3]Department of Dermatology, Churchill Hospital, Oxford, UK
[4]Department of Plastic Surgery, St George's Hospital, London, UK
[5]Department of Plastic and Reconstructive Surgery, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[6]Department of Plastic Surgery, NHS Lothian/University of Edinburgh, Edinburgh, UK
[7]Welsh Institute of Dermatology, University Hospital of Wales, Cardiff, UK
[8]Department of Plastic and Reconstructive Surgery, Norfolk and Norwich University Hospital NHS Trust, Norwich, UK
[9]c/o Cochrane Skin Group, The University of Nottingham, Nottingham, UK
[10]Public Health & Primary Care, University of Cambridge, Cambridge, UK
[11]Centre of Evidence Based Dermatology, University of Nottingham, Nottingham, UK

Citation example: Dinnes J, Deeks JJ, Grainge MJ, Chuchu N, Ferrante di Ruffano L, Matin RN, Thomson DR, Wong KY, Aldridge RB, Abbott R, Fawzy M, Bayliss SE, Takwoingi Y, Davenport C, Godfrey K, Walter FM, Williams HC, Cochrane Skin Cancer Diagnostic Test Accuracy Group. Visual inspection for the diagnosis of cutaneous melanoma in adults. 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 has one of the fastest rising incidence rates of any cancer. It accounts for a small percentage of skin cancer cases but is responsible for the majority of skin cancer deaths. History-taking and visual inspection of a suspicious lesion by a clinician is usually the first in a series of ‘tests’ to diagnose skin cancer. Establishing the accuracy of visual inspection alone is critical to understating the potential contribution of additional tests to assist in the diagnosis of melanoma.

### Objectives

To determine the diagnostic accuracy of visual inspection for the detection of cutaneous invasive melanoma and intraepidermal melanocytic variants in adults with limited prior testing and in those referred for further evaluation of a

suspicious lesion. Studies were separated according to whether the diagnosis was recorded face-to-face (in-person) or based on remote (image-based) assessment.

### Search methods

We undertook a comprehensive search of the following databases from inception up to August 2016: Cochrane Central Register of Controlled Trials; ; 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

Test accuracy studies of any design that evaluated visual inspection in adults with lesions suspicious for melanoma, compared with a reference standard of, either histological confirmation or clinical follow-up. Studies reporting data for 'clinical diagnosis' where dermoscopy may or may not have been used were excluded.

### Data collection and analysis

Two review authors independently extracted all data using a standardised data extraction and quality assessment form (based on QUADAS-2). We contacted authors of included studies where information related to the target condition or diagnostic threshold were missing. We estimated summary sensitivities and specificities per algorithm and threshold using the bivariate hierarchical model. We investigated the impact of: in-person test interpretation; use of a purposely developed algorithm to assist diagnosis; and observer expertise.

### Main results

Forty-nine publications reporting on a total of 51 study cohorts with 34,351 lesions (including 2499 cases) were included, providing 134 datasets for visual inspection. Across almost all study quality domains, insufficient information was provided in the majority of study reports to allow the risk of bias to be judged, while concerns regarding applicability of study findings were scored as 'High' in three of four domains assessed. Selective participant recruitment, lack of detail regarding the threshold for deciding on a positive test result, and lack of detail on observer expertise were particularly problematic.

Attempts to analyse studies by degree of prior testing were hampered by a lack of relevant information and by the restricted inclusion of lesions selected for biopsy or excision. Accuracy was generally much higher for in-person diagnosis compared to image-based evaluations (relative diagnostic odds ratio of 8.54, 95% CI 2.89, 25.3, P<0.001). Meta-analysis of in-person evaluations that could be clearly placed on the clinical pathway showed a general trade-off between sensitivity and specificity, with the highest sensitivity (92.4%, 95% CI 26.2, 99.8%) and lowest specificity (79.7%, 95% CI 73.7, 84.7%) observed in participants with limited prior testing (n = 3 datasets). Summary sensitivities were lower for those referred for specialist assessment but with much higher specificities (e.g. sensitivity 76.7% (95% CI 61.7, 87.1%) and specificity 95.7% (95% CI 89.7, 98.3%) for lesions selected for excision, n = 8 datasets). These differences may be related to differences in the spectrum of included lesions, differences in the definition of a positive test result, or to variations in observer expertise. We did not find clear evidence that accuracy is improved by the use of any algorithm to assist diagnosis in all settings. Attempts to examine the effect of observer expertise in melanoma diagnosis were hindered due to poor reporting.

### Authors' conclusions

Visual inspection is a fundamental component of the assessment of a suspicious skin lesion; however, the evidence suggests that melanomas will be missed if visual inspection is used on its own. The evidence to support its accuracy in the range of settings in which it is used is flawed and very poorly reported. Although published algorithms do not appear to improve accuracy, there is insufficient evidence to suggest that the 'no algorithm' approach should be preferred in all settings. Despite the volume of research evaluating visual inspection, further prospective evaluation of the potential added value of using established algorithms according to the prior testing or diagnostic difficulty of lesions may be warranted.

## Plain language summary

### What is the diagnostic accuracy of visual inspection of skin lesions with the naked eye for diagnosis of melanoma in adults?

#### What is the aim of the review?

Melanoma is one of the most dangerous forms of skin cancer. The aim of this Cochrane Review was to find out the accuracy of visual inspection of suspicious skin lesions with the naked eye to diagnose melanoma. The Review also investigated whether diagnostic accuracy using visual inspection of suspicious skin lesions on a patient in-person differed to diagnostic accuracy of visual inspection of images of suspicious skin lesions. Researchers in Cochrane included 19 studies to answer this question.

#### Why is the diagnostic accuracy of visual examination of skin lesions suspected to be melanomas important?

Not recognising a melanoma when it is present (a false negative test result) delays surgery to remove it, risking cancer spreading to other organs in the body and possibly death. Diagnosing a skin lesion as a melanoma when it is not (a false positive result) may result in unnecessary surgery, further investigations, and patient anxiety. Visual inspection of suspicious skin lesions by a clinician using the naked eye is usually the first of a series of 'tests' to diagnose melanoma. Knowing the diagnostic accuracy of visual inspection alone is important to decide whether additional tests are needed to improve accuracy to an acceptable level.

#### What was studied in the review?

Researchers sought to find out the diagnostic accuracy of visual inspection of suspicious skin lesions on a patient in-person and the diagnostic accuracy of visual inspection of images of suspicious skin lesions. Researchers also sought to find out whether diagnostic accuracy was improved by use of a visual inspection checklist or by an increase in level of clinical expertise; they considered the diagnostic accuracy of the first visual inspection of a lesion, for example, by a GP, and of lesions which had been referred for further evaluation, for example, by a dermatologist.

**What are the main results of the review?**

Only 19 studies (17 in-person studies and 2 image-based studies) were clear whether the test was the first visual inspection of a lesion or was a visual inspection following referral (for example, when patients are referred by a general practitioner to skin specialists for visual inspection).

First visual inspection in-person (3 studies)

The results of 3 studies of 1339 suspicious skin lesions suggest that in a group of 1000 lesions, of which 90 (9%) actually are melanoma:

- An estimated 268 will have a visual inspection result indicating melanoma is present. Of these, 185 will not be melanoma and will result in an unnecessary biopsy (false positive results).

- An estimated 732 will have a visual inspection result indicating that melanoma is not present. Of these, 7 will actually have melanoma and would not be sent for biopsy (false negative results).

Two further studies restricted to 4228 suspicious skin lesions that were all selected to be excised found similar results.

Visual inspection in-person after referral, all lesions selected to be excised (8 studies)

The results of eight studies of 5331 suspicious skin lesions suggest that in a group of 1000 lesions, of which 90 (9%) actually are melanoma:

- An estimated 108 will have a visual inspection result indicating melanoma is present, and of these, 39 will not be melanoma and will result in an unnecessary biopsy (false positive results).

- Of the 892 lesions with a visual inspection result indicating that melanoma is not present, 21 will actually be melanoma and would not be sent for biopsy (false negative results).

Overall, the number of false positive results (diagnosing a skin lesion as a melanoma when it is not) was observed to be higher and the number of false negative results (not recognising a melanoma when it is present) lower for first visual inspections of suspicious skin lesions compared to visual inspection following referral.

Visual inspection of images of suspicious skin lesions (2 studies)

Accuracy was much lower for visual inspection of images of lesions compared to visual inspection in-person.

Value of visual inspection checklists

There was no evidence that use of a visual inspection checklist or the level of clinical expertise changed diagnostic accuracy.

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

In the majority of included studies, the diagnosis of melanoma was made by lesion biopsy and the absence of melanoma was confirmed by biopsy or by follow up over time to make sure the skin lesion remained negative for melanoma*. Biopsy or follow-up are likely to have been reliable methods for deciding whether patients really had melanoma. In a few studies, the absence of melanoma was made by expert diagnosis, which is less likely to have been a reliable method for deciding whether patients really had melanoma. There was lots of variation in the results of the studies in this review. Poor reporting of study conduct made assessment of the reliability of studies difficult. Selective inclusion of particular types of skin lesion, lack of detail about how the diagnosis of melanoma was made, and lack of detail on the expertise of the doctor doing the visual inspection were particularly problematic.

**Who do the results of this review apply to?**

Thirteen studies were undertaken in Europe (68%), with the remainder undertaken in Asia (n = 1), Oceania (n = 4), North America (n = 1). Mean age ranged from 30 to 73.6 years (reported in 10 studies). The percentage of individuals with melanoma ranged between 4% and 20% in first visualised lesions and between 1% and 50% in studies of referred lesions. In the majority of studies, the lesions were unlikely to be representative of the range of those seen in practice, for example, only including skin lesions of a certain size or with a specific appearance. In addition, variation in the expertise of clinicians performing visual inspection and in the definition used to decide whether or not melanoma was present across studies makes it unclear as to how visual inspection should be carried out and by whom in order to achieve the accuracy observed in studies.

**What are the implications of this review?**

Error rates from visual inspection are too high for it to be relied upon alone. Although not evaluated in this review, other technologies need to be used to ensure accurate diagnosis of skin cancer. There is considerable variation and uncertainty about the diagnostic accuracy of visual inspection alone for the diagnosis of melanoma. There is no evidence to suggest that visual inspection checklists reliably improve the diagnostic accuracy of visual inspection, so recommendations cannot be made about when they should be used. Despite the existence of numerous research studies, further, well-reported studies assessing the diagnostic accuracy of visual inspection with and without visual inspection checklists and by clinicians with

different levels of expertise are needed.

**How up-to-date is this review?**

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

*In these studies, biopsy, clinical follow up, or specialist clinician diagnosis were the reference standards.

# Background

This review is one of a series of Cochrane Diagnostic Test Accuracy (DTA) reviews on the diagnosis and staging of melanoma and keratinocyte skin cancers conducted for the National Institute for Health Research (NIHR) Cochrane Systematic Reviews Programme. Appendix 1 shows the content and structure of the programme. Appendix 2 provides a glossary of terms used, and a table of acronyms used is provided in Appendix 3.

## Target condition being diagnosed

Melanoma is one of the most aggressive 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 a small percentage of skin cancer cases but is responsible for up to 75% of skin cancer deaths (Boring 1994; Cancer Research UK 2017).

Melanoma arises from uncontrolled proliferation of melanocytes - the epidermal cells that produce pigment or melanin. It most commonly arises in the skin but can occur in any organ that contains melanocytes, including mucosal surfaces, the back of the eye, and lining around the spinal cord and brain. Cutaneous melanoma refers to a 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. 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 rate of 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.

The incidence of melanoma rose to over 200,000 newly diagnosed cases worldwide in 2012 (Erdmann 2013; Ferlay 2015), with an estimated 55,000 deaths (Ferlay 2015). In the UK, melanoma has one of the fastest rising incidence rates of any cancer and has 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 (http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer/incidence). While overall incidence rates are higher in women than in men, the rate of incidence in the latter is increasing faster than in women (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 ultraviolet (UV) exposure, in conjunction with possible earlier detection (Linos 2009; Belbasis 2016). Putative risk factors are reviewed in detail elsewhere (Belbasis 2016), but can be broadly divided into host or environmental factors. Host factors include fair 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 2005), clinically atypical naevi (Gandini 2005), or large congenital naevi (Swerdlow 1995); genetically inherited skin disorders e.g. xeroderma pigmentosum (Lehmann 2011); and a family history of melanoma (Gandini 2005). Environmental factors include recreational, occupational, and work-related exposure to sunlight (both cumulative and episodic burning) (Gandini 2005; Armstrong 2017); artificial tanning (Boniol 2012); and immunosuppression, e.g. 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).

A database of over 40,000 US patients from 1998 onwards which assisted the development of the eighth American Joint Committee on Cancer (AJCC) Staging System indicated a five-year survival of 99% for stage IA melanoma (melanoma <=1mm thick without ulceration, mitosis or involvement of the lymph nodes), dropping to anything between 32% and 93% in stage III disease (melanoma of any thickness with metastasis to the lymph nodes) depending on tumour thickness, the presence of ulceration and number of involved nodes (Gershenwald 2017). Before the advent of targeted and immuno-therapies, stage IV melanoma (melanoma disseminated to distant sites / visceral organs) was associated with median survival of six to nine months, one year survival rate of 25%, and three year survival of 15% (Balch 2009; Korn 2008).

Between 1975 and 2010, five year relative survival for melanoma (i.e. not including deaths from other causes) in the US increased from 80% to 94%, with survival for localised, regional, and distant disease estimated at 99%, 70%, and 18%, respectively in 2010 (Cho 2014). Overall, mortality rates however 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 much of the observed improvement in survival may be due to earlier detection and heightened vigilance (Cho 2014). New targeted therapies for stage IV melanoma (e.g. BRAF inhibitors) have improved survival and immunotherapies are evolving such that long term survival is being documented (Pasquali 2018). No new data regarding the survival prospects for patients with stage IV disease were analysed for the AJCC 8 staging guidelines due to lack of contemporary data (Gershenwald 2017).

### *Treatment of melanoma*

For primary melanoma, the mainstay of definitive treatment is early detection and excision of the lesion, to remove both the tumour and any malignant cells that might have spread into the surrounding skin (Garbe 2016; Marsden 2010; NICE 2015; SIGN 2017; Sladden 2009). Recommended surgical margins vary according to tumour thickness (Garbe 2016 ) and stage of disease at presentation (NICE 2015).

## Index test(s)

For the purposes of our series of reviews, each component of the diagnostic process, including visual inspection or clinical examination, 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.

Clinical history taking to identify risk factors and visual inspection of the lesion, surrounding skin and comparison with other lesions on the rest of the body is fundamental to the diagnosis of skin cancer. The strongest common phenotypic risk factor is the presence of atypical naevi; typically the presence of over a hundred moles or naevi of abnormal appearance that may pose diagnostic challenges (Goodson 2010; Rademaker 2010; Salerni 2012). In the UK, clinical examination is typically done at two decision points – first in the general practice (GP) surgery where a decision is made to refer or not to refer, and then a second time by a dermatologist or other secondary care clinician where a decision is made to biopsy or not. Specialist advice can also be sought using teledermatology, where lesion images are forwarded with variable clinical information (such as age, gender, and location of lesion) to specialist clinics or to commercial organisations for interpretation. The accuracy of these diagnostic encounters (defined as the proportion of 'correct' diagnoses, i.e. true positive plus true negative diagnoses out of the total number of diagnoses) is known to vary according to qualifications and experience (Westerhoff 2000; Morton 1998); the accuracy of 'image-based' as opposed to face-to-face diagnosis is less clear.

Research into the cognitive processes involved in dermatological diagnoses suggests that two main strategies are employed simultaneously and iteratively (Elstein 2002; Norman 1989; Norman 2009). Non-analytical pattern recognition formulates an initial hypothesis; identification is made implicitly, without conscious thought or reference to specific rules and hidden from the conscious view of the diagnostician (Norman 2009). Analytical pattern recognition using more explicit rules based on conscious analytical reasoning is then employed to test the initial hypothesis. Analytical pattern recognition has been described as the "careful and systematic gathering of data and weighing the elicited information against mental rules" (Norman 2009). The balance between non-analytical and analytical reasoning varies between clinicians, according to factors such as experience and familiarity with the diagnostic question.

Various attempts have been made to formalise the "mental rules" involved in analytical pattern recognition for melanoma, ranging from setting out criteria that should be considered (e.g. 'pattern analysis'; Sober 1979; Friedman 1985) to formal scoring systems with explicit numerical thresholds (MacKie 1985; MacKie 1990). The most commonly used algorithms are described in detail in Appendix 4.

The ABCD (**a**symmetry, **b**order irregularity, **c**olour variegation, **d**iameter > 6 mm) algorithm of clinical warning signs was developed in 1985 to help distinguish melanoma from a benign naevus (Friedman 1985), and then extended to include an E for 'enlargement' criterion (Thomas 1998). As a result of its simplicity, ABCD(E) is now widely advocated for use by non-experts or lay persons (American Academy of Dermatology 2015). The approach has been criticised for its inability to capture nodular and amelanotic melanomas, which account for a relatively small proportion (~15% to 20%) of incident melanomas but a large proportion (~50%) of melanoma-related deaths (Moreau 2013; Shaikh 2012). In addition, up to a third of melanomas may be < 6 mm in diameter (Maley 2014), a proportion which is likely to increase due to improved skin surveillance. The validity of ABCD(E) as a useful tool for the lay public has also been called into question (Girardi 2006, Liu 2005, Aldridge 2011). Subsequent modifications have been suggested, including altering the meaning of the ABCD acronym for use in paediatric populations (Cordoro 2013); changing 'D' to 'dark' (Goldsmith 2014)); or changing the acronym altogether (e.g. CCC for **c**olour, **c**ontour, and **c**hange (Moynihan 1994); or "Do UC" the melanoma for **d**ifferent, **u**neven, **c**hanging (Yagerman 2014)). To date, the latter three have not been evaluated in populations with lesions suggestive for melanoma.

The seven-point checklist assessing change in size, shape, colour, inflammation, crusting or bleeding, sensory change, or diameter ≥ 7 mm was developed by UK researchers as a guide to help non-dermatologists detect possible melanoma (MacKie 1985; MacKie 1990). The revised weighted version (MacKie 1990) is currently recommended for general practitioner (GP) use in the evaluation of pigmented lesions (NICE 2015). A primary care based evaluation found moderately good performance for the identification of clinically significant lesions (including malignant and premalignant lesions as disease positive) in primary care (sensitivity and specificity for the presence of ≥ 3 features were 62.7% and 65.0%, respectively), with higher sensitivity for the detection of melanoma (80.6%) at the expense of low specificity (61.7%) (Walter 2013).

Unlike most formalised rules, the 'ugly duckling' sign is based on differential pattern recognition where abnormal lesion identification is achieved by noticing the odd one out, i.e. a melanoma will be the pigmented lesion that does not match the rest of a person's naevi, for example a very dark or pale/pink lesion that is different in colour compared to the rest of the pigmented naevi (Grob 1998). Although ugly duckling is inherently a form of subjective pattern recognition, sensitivity has been reported to be 100% for pigmented-lesion experts and 85% for non-clinicians (Scope 2008). The assumption that an individual has a "normal" naevus phenotype is debatable, however. Many individuals have multiple 'atypical' pigmented lesions which, although very similar morphologically, allow malignancy to easily

disguise itself amidst an abnormal complex of pigmented lesions (also referred to as 'The Little Red Riding Hood' phenomenon) (Mascaro 1998).

## Clinical Pathway

The diagnosis of melanoma can take place in primary, secondary, and tertiary care settings by both generalist and specialist healthcare providers. In the UK, people with concerns about a new or changing lesion will usually present first to their general practitioner or less commonly directly to a specialist in secondary care, which could include a dermatologist, plastic surgeon, general surgeon or other specialist surgeon (such as an ear, nose, and throat (ENT) specialist or maxillofacial surgeon), or ophthalmologist (Figure 2). 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 seven-point checklist (MacKie 1990); lesions suspected to be melanoma should be referred urgently for appropriate specialist assessment within two weeks (Chao 2013; Marsden 2010; NICE 2015c; SIGN 2017).

Teledermatology consultations can aid more appropriate triage of lesions into urgent referral; non-urgent secondary care referral (e.g. for suspected basal cell carcinoma); or where available, referral to an intermediate care setting, e.g. clinics run by GPs with a special interest in dermatology. The distinction between setting and examiner qualifications and experience is important as specialist clinicians might work in primary care settings (for example, in the UK, general practitioners (GPs) with a special interest in dermatology and skin surgery who have undergone appropriate training), and generalists might practice in secondary care settings (for example, plastic surgeons who do not specialise in skin cancer). The level of skill and experience in skin cancer diagnosis will vary for both generalist and specialist care providers and will also impact on test accuracy.

The specialist clinician will also use history-taking and visual inspection of the lesion (in comparison with other lesions on the skin), usually in conjunction with dermoscopic examination, to inform a clinical decision. If melanoma is suspected, then urgent excision biopsy is recommended; for suspected cutaneous squamous cell carcinoma (cSCC) urgent excision with predetermined surgical margins. Other lesions such as basal cell carcinoma (BCC) or pre-malignant lesions such as lentigo maligna may also be referred for a diagnostic biopsy, followed by appropriate treatment or further surveillance or reassurance and discharge.

### *Prior test(s)*

Although smartphone applications and community-based teledermatology services can increasingly be directly accessed by people who have concerns about a skin lesion (Chuchu 2018), visual inspection of a suspicious lesion by a clinician is usually the first in a series of tests to diagnose skin cancer. In the UK first visual inspection of a suspicious lesion usually takes place in primary care; however, in some countries, people with suspicious lesions can present directly to a secondary care setting. Considering the degree of prior testing that study participants have undergone is key to interpretation of resulting test accuracy indices, which are known to vary according to the spectrum or case-mix of included participants (Lachs 1992; Moons 1997; Leeflang 2013; Usher-Smith 2016). Studies of people with suspicious lesions at the initial clinical presentation stage ('test naïve') are likely to have a wider range of differential diagnoses and include a higher proportion of people with benign diagnoses compared with studies of participants who have been referred for a specialist opinion on the basis of visual inspection (with or without dermoscopy) by a generalist practitioner. Furthermore, studies in more specialist settings may focus on equivocal or difficult to diagnose lesions rather than lesions with a more general level of clinical suspicion. A simple categorisation of studies according to primary, secondary, or specialist setting may not always adequately reflect differences in spectrum.

### *Role of index test(s)*

Visual inspection and history-taking are key to diagnosing skin cancer and are always undertaken as part of a clinical examination regardless of examiner experience and whatever additional technologies are available. For the generalist practitioner, the key is to minimise the proportion of people who are referred unnecessarily and identify those lesions that require urgent referral. For the specialist, the aim is not only to identify those in need of urgent excision due to invasive cancer, but also to identify high risk lesions with considerable potential to progress to invasive disease, such as those with severe dysplasia or *in situ* disease e.g. lentigo maligna, for example. Given differences in setting, prior testing, observer qualifications, experience and training, the anticipated performance in terms of accuracy is likely to vary.

When diagnosing potentially life-threatening conditions such as melanoma, the consequences of falsely reassuring a person that they do not have skin cancer can be serious and potentially fatal, as the resulting delay to diagnosis means that the window for successful early treatment may be missed. To minimise these false-negative diagnoses, a good diagnostic test will demonstrate high sensitivity and a high negative predictive value (NPV), where very few of those with a negative test result will actually have a melanoma. Giving falsely positive test results (meaning the test has poor specificity and a high false-positive rate) resulting in the removal of lesions that turn out to be benign is arguably less of an error than missing a potentially fatal melanoma, but is not cost free. False-positive diagnoses not only cause unnecessary scarring from the biopsy or excision procedure, but also increase patient anxiety whilst they await the definite histology results and increase healthcare costs as the number needed to remove to yield one melanoma diagnosis increases.

### *Alternative test(s)*

A number of other tests have been reviewed as part of our series of Cochrane DTA reviews on the diagnosis of melanoma. In particular, dermoscopy has become an essential tool for the specialist clinician and is increasingly being taken up in primary care settings. Dermoscopy (also referred to as dermatoscopy or epiluminescence microscopy or ELM) uses a hand-held microscope and incident light (with or without oil immersion) to reveal subsurface images of the

skin at increased magnification of x 10 to x 100 (Kittler 2011). Used alongside clinical examination, dermoscopy has been shown in some studies to increase the sensitivity of clinical diagnosis of melanoma from around 60% to as much as 90% (Kittler 1999; Carli 2002; Bono 2006; Stanganelli 2000) with much smaller effects in others (Benelli 1999; Bono 2002). The accuracy of dermoscopy depends on the experience of the examiner (Kittler 2011), with accuracy when used by untrained or less experienced examiners potentially no better than clinical inspection alone (Binder 1997; Kittler 2002).

Pattern analysis (Steiner 1987; Pehamberger 1993) is thought to be the most specific and reliable technique to aid dermoscopy interpretation when used by specialists (Maley 2014); however, dermoscopic histological correlations have been established and diagnostic algorithms developed based on colour, aspect, pigmentation pattern, and skin vessels (e.g. the ABCD rule for dermoscopy (Nachbar 1994; Stolz 1994), the Menzies (Menzies 1996) and the seven-point dermoscopy checklist (Annessi 2007; Argenziano 1998; Argenziano 2001; Gereli 2010; amongst others). Dermoscopy used in addition to visual inspection (in-person evaluations) or used alone (dermoscopic image interpretation remotely from the patient concerned) are the subject of a separate systematic review (Dinnes 2018).

A number of other tests which may have a role for the diagnosis of melanoma in a specialist setting have been reviewed as part of our series of systematic reviews, including teledermatology, mobile phone applications, reflectance confocal microscopy, optical coherence tomography, computer-aided diagnosis or artificial intelligence-based techniques, and high frequency ultrasound (Dinnes 2015). 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.

We also considered and excluded a number of tests from review including tests used in the context of monitoring people, such as total body photography of those with large numbers of typical or atypical naevi, and finally histopathological confirmation following lesion excision. The latter is the established reference standard for melanoma diagnosis and will be one of the standards against which the index tests are evaluated in these reviews.

### Rationale

Our series of reviews of diagnostic tests used to assist clinical diagnosis in either clinical practice or in a research setting 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 diagnostic and treatment decisions. With increasing rates of melanoma and a trend to adopt the use of dermoscopy and other high-resolution image analysis in primary care, 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 contribute to morbidity and mortality or whether newer technologies run the risk of increasing false-positive diagnoses. 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). The trend toward remote interpretation of dermatology images (whether clinical or dermoscopic images) and the use of remote technologies that do not involve clinicians without substantive evidence could further disrupt clinical pathways and healthcare payments as they may attract custom from the worried well, leaving an ever decreasing pool of qualified doctors to pick up any resulting problems.

There are few available systematic reviews in the field. The literature searches for the most comprehensive systematic reviews of visual inspection were carried out up to 2007 (Vestergaard 2008) or are focused on specific clinical questions, for example, specific healthcare professionals (Corbo 2012 including only direct comparisons of the accuracy of primary care physicians versus dermatologists, and Loescher 2011 reviewing the skin cancer detection skills of advanced practice nurses) or settings (Herschorn 2012 including direct comparisons of visual inspection versus dermoscopy in primary care). More recently, Harrington and colleagues (Harrington 2017) published a systematic review of clinical prediction rules (or published algorithms) used to assist the diagnosis of melanoma; however, the requirement for a clinical prediction rule does not allow comparison of accuracy with and without the use of an algorithm.

The critical question about the accuracy of visual inspection alone and the impact of examiner, prior patient testing, underlying risk status, and the use of images for diagnosis needs to be answered before the potential contribution of additional diagnostic tests can be set in context and appropriately placed in the diagnostic pathway.

This review follows a generic protocol which covers the full series of Cochrane DTA reviews for the diagnosis of melanoma (Dinnes 2015). The Background and Methods sections of this review therefore use some text that was originally published in the protocol (Dinnes 2015) and text that overlaps some of our other reviews (Dinnes 2018).

## Objectives

To determine the diagnostic accuracy of visual inspection for the detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants in adults.

Accuracy was estimated separately according to the prior testing undergone by study participants:

- i. those with limited prior testing, i.e. primary presentation
- ii. those referred for further evaluation of a suspicious lesion, i.e. referred participants

Accuracy was also estimated separately according to whether the diagnosis was recorded based on a face-to-face (in-person) encounter or based on remote (image-based) assessment.

### Secondary objectives

For the identification of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants:

- i. To determine the diagnostic accuracy of individual algorithms used to assist visual inspection; and
- ii. To determine the effect of observer experience on diagnostic accuracy

For the alternative definitions of the target condition:

- iii. To determine the diagnostic accuracy of visual inspection for the detection of invasive melanoma alone in adults.
- iv. To determine the diagnostic accuracy of visual inspection for the detection of any skin cancer or skin lesion with a high risk of progression to melanoma in adults (i.e. requiring excision).

### Investigation of sources of heterogeneity

We set out to address a range of potential sources of heterogeneity for investigation across our series of reviews, as outlined in our generic protocol (Dinnes 2015) and described in Appendix 5; however, our ability to investigate these was necessarily limited by the available data on each individual test reviewed.

The sources of heterogeneity that were investigated for visual inspection were:

- In-person versus image-based evaluations;
- study setting: primary, community or private care versus secondary versus specialist clinics;
- use of a diagnostic algorithm: no algorithm reported versus any named algorithm used;
- type of reference standard: histology alone versus histology plus clinical follow-up or other reference standard; and
- disease prevalence: <=10% versus >10%. The 10% cut-off was chosen based on advice from clinical co-authors (RB, HW).

# 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); however, we did not include studies that compared results for malignant lesions to those for healthy skin (i.e. with no lesion present).
- 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 if they included less than five melanoma cases or less than five benign lesions. The size threshold of five is arbitrary. However, such small studies are unlikely to add precision to the estimate of accuracy.

### *Participants*

We included studies in adults with pigmented skin lesions or lesions suspicious for melanoma or 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.

We excluded studies that recruited only participants with malignant or benign diagnoses.

We excluded studies conducted in children or which clearly reported inclusion of more than 50% of participants aged 16 and under.

### *Index tests*

Studies reporting accuracy data for visual inspection alone, with either image-based or in-person diagnosis, were eligible for inclusion. For in-person visual inspection, diagnosis is undertaken in a clinic setting with the patient present (face-to-face diagnosis). For these studies we assumed that patient history-taking would have taken place and is likely to have contributed to lesion diagnosis; however, we did not specifically extract details of patient history-taking due to anticipated poor reporting in the primary studies. For image-based studies, diagnosis is based on clinical or ‘macro’ images (photographs), remotely from the study participant. For these studies, any additional patient information that was provided to assist diagnosis was extracted.

All established algorithms or checklists to assist diagnosis by visual inspection were included. Studies developing new algorithms or methods of diagnosis (i.e. derivation studies) were included if they:

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

- investigated lesion characteristics that had previously been suggested as associated with melanoma and the study reported accuracy based on the presence or absence of particular combinations of characteristics.

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)
- evaluated the accuracy of the presence or absence of individual lesion characteristics or morphological features, with no overall diagnosis of malignancy
- reported accuracy data for ‘clinical diagnosis’ with no clear description as to whether the reported data related to visual inspection alone
- were based on the experience of a particular skin cancer clinic, where dermoscopy may or may not have been used on an individual patient basis.

Although primary care clinicians can in practice be specialists in skin cancer, we considered primary care physicians as generalist practitioners and dermatologists as specialists. Within each group, we extracted any reporting of special interest or accreditation in skin cancer.

### *Target conditions*

The primary target condition was defined as the detection of:

- any form of invasive cutaneous melanoma or atypical intraepidermal melanocytic variants (i.e. including melanoma in situ, or lentigo maligna, which has a risk of progression to invasive melanoma).

Two additional definitions of the target condition were considered in secondary analyses, namely the detection of:

- any form of invasive cutaneous melanoma alone;
- any skin lesion requiring excision. This latter definition includes melanoma plus other forms of skin cancer, such as basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (cSCC), as well as melanoma in situ, lentigo maligna, and lesions with severe melanocytic dysplasia.

The diagnosis of the keratinocyte skin cancers, basal cell carcinoma, and squamous cell carcinoma as primary target conditions are the subject of a separate series of reviews (Dinnes 2015a).

### *Reference standards*

The ideal reference standard is histopathological diagnosis in all eligible 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 reporting of a minimum dataset as a necessary inclusion criterion, but extracted any pertinent information.

Partial verification (applying the reference test only to a subset of those undergoing the index test) was of concern given that lesion excision or biopsy are unlikely to be carried out for all benign-appearing lesions within a representative population sample. Therefore, to reflect what happens in reality, we accepted clinical follow-up of benign-appearing lesions as an eligible reference standard, whilst recognising the risk of differential verification bias (as misclassification rates of histopathology and follow-up will differ).

Additional eligible reference standards included cancer registry follow-up and 'expert opinion' with no histology or clinical follow-up. Cancer registry follow-up is considered less desirable than active clinical follow-up, as follow-up is not carried out within the control of the study investigators. Furthermore, if participant-based analyses as opposed to lesion-based analyses are presented, it may be difficult to determine whether the detection of a malignant lesion during follow-up is the same lesion that originally tested negative on the index test.

All of the above were considered eligible reference standards 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.

## 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 6), 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.

### *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, BCC, or cSCC were included at initial screening. Inclusion criteria (Appendix 7) 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 methodologist 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 final lesion diagnoses or diagnostic threshold were missing. In particular, invasive cSCC (included as disease positive for one of our secondary objectives) is not always differentiated from '*in situ*' variants such as Bowen's disease (which we did not consider as disease positive for any of our definitions of the target condition). 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 this 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 review topic (see Appendix 8). The modified QUADAS-2 tool was piloted on a small number of included full text articles. One clinical (as detailed above) and one methodologist 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*

Separate analyses were conducted according to the point that study participants have reached in the clinical pathway (numbered from 1 to 7 in Figure 3), the clarity with which the pathway could be determined (clear or unclear), and the evaluation of in-person versus image-based diagnosis.

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 rather than systemically (thus it is important to be able to correctly identify cancerous lesions for each person), 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 the same people contribute data for multiple lesions, most studies include very few people 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. Where multiple algorithms were assessed in an individual study, datasets were selected on the following preferential basis:

i. ‘no algorithm’ reported; data presented for clinician’s overall diagnosis or management decision
ii. pattern analysis or pattern recognition
iii. ABCD algorithm (or derivatives of)
iv. 7-point checklist (also referred to as Glasgow/MacKie checklist)

Where multiple thresholds per algorithm were reported, we included the standard or most commonly used threshold. If data for multiple observers was reported, we used data for the most experienced observer, using single observer diagnosis in preference to a consensus or average across observers. If we were unable to choose a dataset based on the above ‘rules’, a random selection of one dataset per study was made. Data on the accuracy of dermoscopy, to allow comparisons of tests, was included in a separate review in our series (Dinnes 2018).

For each analysis, estimates of sensitivity and specificity were plotted on coupled forest plots and in receiver operating characteristic (ROC) space. For tests where commonly used thresholds were reported we estimated summary operating points (summary sensitivities and specificities) with 95% confidence and prediction regions using the bivariate hierarchical model (Chu 2006; Reitsma 2005). Where inadequate data were available for the model to converge the model was simplified, first by assuming no correlation between estimates of sensitivity and specificity and secondly by setting estimates of near zero variance terms to zero (Takwoingi 2015). Where all studies reported 100% sensitivity (or 100% specificity) the number with disease (or no disease) was summed across studies and used to compute a binomial exact 95% confidence interval.

For computation of likely numbers of true positive, false positive, false negative and true negative findings in the summary of findings tables, these indicative values were applied to lower quartile, median and upper quartiles of the prevalence observed in the study groups. In summary of findings tables we report these numbers for the average operating point on the SROC curve.

### *Investigations of heterogeneity*

Investigations of heterogeneity, comparisons between algorithms and according to observer experience were made by comparing summary ROC curves using the hierarchical summary receiver-operator curves (HSROC) model (Rutter 2001). HSROC curves allow incorporation of data at different thresholds and from different algorithms or checklists. We used an HSROC model that assumed a constant SROC shape between tests and subgroups, but allowed for differences in threshold and accuracy by addition of covariates. The significance of the differences between tests or subgroups was assessed by the likelihood ratio test assessing differences in both accuracy and threshold, and by a Wald test on the parameter estimate testing for differences in accuracy alone. Simpler models were fitted when convergence was not achieved due to small numbers of studies, first assuming symmetric SROC curves (setting the shape term to zero), and then setting random effects variance estimates to zero. Estimates of accuracy from HSROC models are presented as diagnostic odds ratios (DORs) (estimated where the SROC curve crosses the sensitivity=specificity line) with 95% confidence intervals. Differences between tests and subgroups from HSROC analyses are presented as relative diagnostic odds ratios (RDORs) with 95% confidence intervals.

Bivariate models were fitted using the xtmelogit command in STATA 15 and HSROC models fitted using the NLMIXED procedure in the SAS statistical software package (SAS 2012, version 9.3; SAS Institute, Cary, NC, USA) and the metadas macro (Takwoingi 2010).

### *Sensitivity analyses*

Sensitivity analyses, restricting analyses to studies at least risk of bias were planned; however, these were not conducted due to insufficient study numbers.

### *Assessment of reporting bias*

Because of uncertainty about the determinants of publication bias for diagnostic accuracy studies and the inadequacy of tests for detecting funnel plot asymmetry (Deeks 2005), we did not perform tests to detect publication bias.

# 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 for any one of the suite of reviews of tests to assist in the diagnosis of melanoma or keratinocyte skin cancer. Of the 1051 full text papers assessed, 848 were excluded from all reviews in our series (see Figure 4 PRISMA flow diagram of search and eligibility results).

Of the 232 studies tagged as potentially eligible for this review of visual inspection, 49 publications were included reporting 49 individual studies. Exclusions were mainly due to the inability to construct a 2x2 contingency table based on the data presented (n = 54); the use of ineligible index tests (n = 39) (for example: reporting of data for visual inspection and dermoscopy only (n = 12), reporting of data for 'clinical diagnosis' (n = 11), or for serial use of the index test in a follow-up context (n = 7)); or not meeting our requirements for an eligible reference standard (n = 23). Other reasons for exclusion included ineligible study populations (n = 20) (for example, recruiting only malignant or only benign lesions (n = 18)), inadequate sample size (n = 14), ineligible definition of the target condition (n = 14) or with test interpretation by medical students or laypersons (n = 6). A list of the 183 publications 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 separate pdf.

The authors of 14 publications were contacted for the purposes of this review of visual inspection and, to date, responses have been received with regard to 7 publications. One response allowed the inclusion of the study in the review (Walter 2012), five provided clarifications on methods used on studies included (Bono 2006; Bourne 2012; Rosendahl 2011; Stanganelli 2000; Walter 2012); one replied with the information needed but the two studies could not be included due to the evaluation of 'clinical diagnosis' (Youl 2007; Youl 2007a); and five replied but were not able to provide the information requested in relation to eight study publications, one which could still be included (Menzies 2009 ) and seven which could not (Fabbrocini 2008; Freeman 1963; Heal 2008; Menzies 2009; Warshaw 2009; Warshaw 2009a; Warshaw 2010).

The 49 included study publications report on a total of 51 cohorts of lesions and 134 datasets with 34,351 lesions and 2499 malignancies. The total number of study participants with suspicious lesions cannot be estimated due to lack of reporting in study publications. Two thirds of studies (n = 32; 65%) also reported accuracy data for diagnosis using dermoscopy; these comparisons are reported in Dinnes 2018. Seven studies reported data for additional tests including: teledermatology (n = 1) and computer-assisted diagnosis techniques (n = 6).

## Methodological quality of included studies

The overall methodological quality of all included studies (n = 49) is summarised in Figure 5 and Figure 6.

Across almost all study quality domains, insufficient information was provided in the majority of study reports to allow the risk of bias to be judged, while applicability of study findings were scored as of 'High' concern in three of four domains assessed.

For participant selection, only 22% of studies (n = 11) were judged at low risk of bias and 27% (n = 13) were considered high risk. Ten studies (20%) either used a case-control type design with separate selection of melanoma cases and lesions with benign diagnoses (n = 6) or did not clearly describe the study design used (n = 5). Over half (55%; n = 27) reported random or consecutive participant recruitment reported; the remaining 45% did not describe recruitment methods. Over half of studies (53%) did not describe whether any exclusion criteria were applied and were judged as at unclear risk of bias. Seven studies (14%) applied inappropriate participant exclusions, excluding 'difficult to diagnose' lesions such as awkwardly located lesions (Bono 2002; Morales Callaghan 2008; Unlu 2014); those with disagreement on histopathology (de Giorgi 2012; Ek 2005; Zaumseil 1983); or dermoscopically 'peculiar' lesions (Carli 2003b).

Almost all cohorts (96%; n = 47) were considered at high concern for applicability of participants. In the majority of cases (n = 41), high concern was due to restricted study populations: inclusion of only melanocytic (n = 10) or amelanotic (n = 1) lesions; restriction by lesion diameter (Bono 2002b; Bono 2006; Steiner 1987); or, most commonly, inclusion of lesions selected for excision based on the clinical or dermoscopic diagnosis or selected retrospectively from histopathology databases (n = 37). Only four cohorts were judged to have included a representative patient population (Grimaldi 2009; Menzies 2009; Stanganelli 2000; Walter 2012). Fourteen cohorts also included multiple lesions per participant, with only 8 clearly including similar number of participants and lesions (Bono 2002; Bono 2002b; Bono 2006; Bourne 2012; Collas 1999; Krahn 1998; Pizzichetta 2004; Unlu 2014).

For the index test domain, studies were considered separately according to whether they reported in-person evaluations of visual inspection (n = 33) or evaluations based on interpretation of clinical images (image-based evaluations; n = 16). For the in-person evaluations, 24% (n = 8) were considered at low risk of bias, and 9% (n = 3) were judged high risk; 22 (67%) did not provide sufficient information to allow the risk of bias to be fully judged. All studies were considered to have made the diagnosis blinded to the reference standard result: 24% (n = 8) also clearly reported pre-specification of the diagnostic threshold (5 of the 8 using named algorithms (Argenziano 2006; Cristofolini 1994; Stanganelli 2000; Walter 2012; Zaumseil 1983 and three by the same author team (Bono 2002; Bono 2002b; Bono 2006) describing the process by which the diagnosis was reached. Three studies developed new algorithms (Thomas 1998) or evaluated multiple thresholds for test positivity (Benelli 2001; McGovern 1992).

Reporting was poorer for the image-based evaluations, with over three quarters of studies (n = 13) not providing sufficient information to allow the risk of bias to be fully judged, one study (6%) judged at low risk of bias and two (12%) at high risk. All studies were again considered to have made the diagnosis blinded to the reference standard result, with one prospectively testing two pre-specified diagnostic thresholds (Benelli 2001) and two (de Giorgi 2012; Scope 2008) testing multiple diagnostic thresholds.

High concern for the applicability of the index tests was recorded for 85% (n = 28) of in-person evaluations. High concern was primarily due to a lack of description of the diagnostic thresholds used (n = 24), but also as a result of presentation of average (Argenziano 2006) or consensus diagnoses (Barzegari 2005; Benelli 1999; Carli 2002; Cristofolini 1997; Morales Callaghan 2008; Steiner 1987) as opposed to the diagnosis of a single observer. Two studies were also judged to have reported diagnosis by non-expert observers (Menzies 2009; Walter 2012), both of which reported diagnoses by large groups of primary care practitioners. In reality, specific expertise in diagnosing pigmented lesions does vary amongst examiners, for example Menzies 2009 requiring a history of excision or referral of at least 10 pigmented skin lesions over the previous 12-month period but excluding those already using dermoscopy or digital monitoring of lesions, and Walter 2012 excluding those with specialist dermatology training but reporting some training in dermatology for almost a quarter of participating GPs. Almost three quarters of studies (n = 24) were judged to have applied and interpreted the 'test' in a clinically applicable manner, nine (27%) provided sufficient detail of the threshold used and 11 (33%) described the observers as expert or experienced. All image-based studies were of high concern for applicability, due to the image-based nature of interpretation limiting the clinical applicability of findings but also the lack of detail on the thresholds used (n = 13). A higher proportion (62%; n = 10) described the observers as expert or experienced.

Of the 49 included cohorts, 85% were judged at low risk of bias for the reference standard due to the use of an acceptable reference standard (n = 42). Six did not meet our criteria for an acceptable reference standard, with more than 20% of the benign lesions having only expert diagnosis with no clinical follow-up (Bono 1996; Green 1991; Grimaldi 2009; Menzies 2009; Stanganelli 2000; Walter 2012), three of which were primary care based studies (Grimaldi 2009; Menzies 2009; Walter 2012). Blinding of the reference standard to the index test (in this case the pathology referral diagnosis) was recorded but did not contribute to the overall risk of bias for the reference standard domain. No blinding of the reference standard was implemented in three studies (Menzies 2009 and Walter 2012 referring patients for excision under standard practice and Thomas 1998 describing a form recording the presence or absence of each ABCDE criterion to the usual pathology form) and blinding was not described in the remaining 46 studies (94%). The applicability of the reference standard was of low concern in 9 studies (18%), high in 7 (14%), and unclear for 33 (67%). In all cases, high concern was due to the use of expert opinion for classifying the final diagnosis of some lesions. The majority of studies (40; 82%) did not report histopathology interpretation by an experienced histopathologist or by a dermatopathologist.

In terms of flow and timing, 20 cohorts were judged at high risk of bias, 7 at low risk and 22 did not provide enough information on which to judge this domain. Of those at high risk, 11 cohorts did not use the same reference standard for all participants (differential verification), and 15 did not include all participants in the analysis either due to incomplete information (n = 7; Argenziano 2006; Bono 1996; Ek 2005; McGovern 1992; Menzies 2009; Pizzichetta 2004; Walter 2012 ); inadequate images (Chang 2013; Dolianitis 2005; Green 1994; Lorentzen 1999; Pizzichetta 2004; Rosendahl 2011; Scope 2008); and exclusion of particular lesion groups following recruitment (Bourne 2012; Dummer 1993; Menzies 2009). A further 37 cohorts were unclear on the interval between the application of the index test and excision for histology with 12 reporting consecutive diagnosis and excision or biopsy.

## Findings

### *1. Target condition: invasive melanoma and melanocytic intra-epidermal variants*

Thirty-seven studies reported accuracy data for the detection of invasive melanoma and melanocytic intra-epidermal variants, one of which reported data for three different sets of lesions (Morton 1998a; Morton 1998b; Morton 1998c) giving a total of 39 datasets; 28 evaluations were conducted in-person and 11 were image-based.

Summary details of the in-person studies are summarised in Appendix 9, with quality assessments in Appendix 10. Summary details of the image-based studies are summarised in Appendix 11 with quality assessments in Appendix 12. Details of established algorithms used to assist diagnosis are described in detail in Appendix 4. Results for the primary analyses are presented in Table 1. Forest plots of study data for each analysis in Table 1 are given for each analysis in Figure 7 and Figure 8; summary estimates are depicted in Figure 9 and Figure 10. Table 2 reports heterogeneity investigations, Table 3 compares test algorithms and Table 4 compares observers.

#### In-person evaluations

Of the 28 evaluations conducted on an in-person basis, 17 contained enough information to describe where on the clinical pathway participants were assessed (coded as 'clear' on pathway), and 11 were considered not to have provided sufficient information to allow the pathway to be identified (coded 'unclear' on pathway). These evaluations are considered according to position on the pathway and clear versus unclear pathway classification (Table 1). Figure 7 presents the results of the individual studies grouped by their position on the pathway; Figure 9 depicts the summary estimates at each point on the pathway.

##### Studies in participants with limited prior testing

Six in-person evaluations of visual inspection recruited series of participants with pigmented lesions who were presenting for a first structured clinical assessment of a suspicious lesion (Grimaldi 2009; Menzies 2009; Walter 2012; Collas 1999; Gachon 2005; McGovern 1992) (Appendix 9; Appendix 10). All studies included participants with pigmented

lesions; Gachon 2005 restricted inclusion to melanocytic lesions only. The prevalence of disease ranged from 4 to 6% in four studies, with Collas 1999 (11%) and Grimaldi 2009 (20%) reporting higher prevalence of melanoma.

Three studies prospectively included all participants presenting in primary care within a given time frame and were clearly positioned on the clinical pathway (Pathway 2-c in Figure 9):

- summary sensitivity was 92.4% (95% CI: 26.2, 99.8%) and specificity 79.7% (95% CI: 73.7, 84.7%) [1339 lesions and 55 melanomas (Grimaldi 2009; Menzies 2009; Walter 2012)].

Histological diagnosis was supplemented with clinical follow-up of at least 3 months for lesions considered benign (all three studies) and expert clinical diagnosis only without follow-up for some benign lesions (Menzies 2009; Walter 2012).

Three studies included only participants with lesions selected for excision (Pathway 3-c and 3-u in Figure 9): two were conducted in private dermatology clinics (Collas 1999; Gachon 2005) and one at an open access veterans' dermatology clinic (McGovern 1992) (Appendix 9):

- summary sensitivity was 90.1% (95% CI: 70.0, 97.3%) and specificity 81.3% (95% CI: 67.5, 90.0%) for two studies clearly positioned on the clinical pathway(Pathway 3-c) [4228 lesions and 160 melanomas (Gachon 2005; McGovern 1992)].
- sensitivity was 78.9% (95% CI: 62.7, 90.4%) and specificity 94.0% (95% CI: 90.7, 96.3%) [353 lesions and 38 melanomas (Collas 1999)] in the single study that could not be clearly positioned on the clinical pathway (Pathway 3-u).

Diagnosis was recorded by primary care physicians with a range of experience (Grimaldi 2009; Menzies 2009; Walter 2012 ) or by dermatologists (Collas 1999; Gachon 2005; McGovern 1992) with no obvious differences in sensitivity or specificity. No formal algorithm to assist diagnosis was reported in four studies; two of which classified lesions 'suspicious for malignancy' as test positive (Grimaldi 2009; Gachon 2005) and two reported data for 'correct' or 'primary' diagnosis of melanoma' (Menzies 2009; Collas 1999). Walter 2012 reported data for MacKie's revised 7-point checklist (MacKie 1990) at a threshold of >= 3, and McGovern 1992 used the BCD algorithm at >=2 characteristics present (this study also reported data using the original 7-point checklist, see Analyses by algorithm reported below).

**Studies in referred participants**

Twenty-two in-person evaluations of visual inspection were conducted in participants referred for specialist assessment; 12 could be clearly positioned on the clinical pathway (three evaluations from a single study) and 10 did not provide sufficient information for a clear assessment to be made (Figure 7; Appendix 9; Appendix 10).

Two studies were judged to include all participants referred for further assessment (Pathway 4-c in Figure 9) and both were clearly positioned on the clinical pathway:

- summary sensitivity was 74.6% (95% CI: 48.9, 90.0%) and specificity 98.6% (95% CI: 94.7, 99.6%) [3494 lesions and 61 melanomas; (Barzegari 2005; Stanganelli 2000)].

Fifteen studies providing 17 datasets included only those with any lesion selected for excision (Pathway 5-c and 5-u in Figure 9):

- summary sensitivity was 76.7% (95% CI: 61.7, 87.1%) and specificity 95.7% (95% CI: 89.7, 98.3%) (5331 lesions and 258 melanomas) for six studies (with 8 datasets) clearly positioned on the clinical pathway (Pathway 5-c) (Bono 2002; Bono 2002b; Bono 2006; Ek 2005; Green 1991; Morton 1998a; Morton 1998b; Morton 1998c).
- summary sensitivity was 82.8% (95% CI: 74.4, 88.9%) and specificity 89.2% (95% CI: 71.1, 96.5%) (9611 lesions and 1015 melanomas) for nine studies that could not be clearly positioned on the clinical pathway (Pathway 5-u) (Benelli 1999; Carli 2002; Cristofolini 1994; Cristofolini 1997; Langley 2001; Morales Callaghan 2008; Thomas 1998; Unlu 2014; Zaumseil 1983).

Three studies were considered to report data only for those participants with equivocal or difficult to diagnose lesions selected for excision (Pathway 5*-c and 5*-u in Figure 7 and Figure 9):

- summary sensitivity was 84.7% (95% CI: 55.5, 96.1%) and specificity 89.5% (95% CI: 79.5, 95.0%) (930 lesions and 88 melanomas) for two studies clearly positioned on the clinical pathway (Pathway 5*-c) (Dummer 1993; Soyer 1995).
- sensitivity was 61.4% (95% CI: 49.0, 72.9%) and specificity 87.3% (95% CI: 82.5, 91.2%) (318 lesions and 73 melanomas) in one study not clearly positioned on the clinical pathway (Pathway 5*-u) (Steiner 1987).

Studies included pigmented lesions referred for further evaluation at a dermatology or pigmented lesion clinic, two restricting to melanocytic lesions only (Morales Callaghan 2008; Unlu 2014) and four restricting by lesion diameter (<= 3 mm (Bono 2006), <= 6 mm (Bono 2002b), < 10 mm (Steiner 1987), or <= 15 mm (Barzegari 2005)). The prevalence of disease ranged from 1% (Ek 2005) to 41% (Soyer 1995). Disease prevalence was generally lower in studies clearly positioned on the clinical pathway (11% or less in 7 of 10 datasets) compared to those that could not be clearly positioned (7 of 9 datasets reporting disease prevalence of 15% or over (Appendix 9). The prevalence of melanoma in studies of equivocal lesions was 3% (Dummer 1993), 23% (Steiner 1987) and 41% (Soyer 1995).

Diagnoses were recorded by dermatologists or dermatology residents (or were assumed to be by dermatologists based on authors' institutions or study settings), by surgical oncologists or by plastic surgeons (Appendix 9). Observer experience was poorly reported, with only 7 studies referring to 'experienced' or 'expert' observers; three studies were clearly positioned on the pathway and 4 not clearly positioned. Observer diagnosis with no formal algorithm was reported in all studies apart from five using ABCD or ABCDE algorithms (Benelli 1999; Cristofolini 1994; Cristofolini 1997; Thomas 1998; Stanganelli 2000). Diagnosis was more often based on the opinion of a single observer as

opposed to a consensus or average decision in studies clearly positioned on the pathway (10/12 datasets) (Stanganelli 2000; Bono 2002; Bono 2002b; Bono 2006; Green 1991; Morton 1998a; Morton 1998b; Morton 1998c; Dummer 1993; Soyer 1995 ) compared to those not clearly positioned (3/10) (Thomas 1998; Unlu 2014; Zaumseil 1983).

## Image-based evaluations

Of the 11 image-based evaluations, two contained enough information to describe where on the clinical pathway participants were assessed (coded as 'clear' on pathway), and 9 were considered not to have provided sufficient information to allow the pathway to be identified (coded 'unclear' on pathway) (Appendix 11 and Appendix 12). Results are present in Table 1. Figure 8 presents the results of the individual studies grouped by their position on the pathway; Figure 10 depicts the summary estimates at each point on the pathway.

### Studies in participants with limited prior testing

Two studies retrospectively reviewed clinical images from participants with lesions excised in primary care settings (Pathway 3-c and 3-u in Figure 10):

- sensitivity was 22.2% (95% CI: 2.8, 60.0%) and specificity 70.7% (95% CI: 54.4, 83.9%) (50 lesions and 9 melanomas) in one study clearly positioned on the clinical pathway (Pathway 3-c) *(* Bourne 2012).
- sensitivity was 20.7% (95% CI: 8.0, 39.7%) and specificity 96.8% (95% CI: 94.6, 98.2%) (463 lesions and 29 melanomas) in the study not be clearly positioned on the clinical pathway (Pathway 3-u) (Rosendahl 2011). The study report was unclear as to whether the excisions were undertaken at the primary care practice or in a referral setting.

The prevalence of melanoma was 6% (Rosendahl 2011) and 20% (Bourne 2012) and both studies included a range of different types of lesions. Lesion images were reviewed by three GPs and a clinical nurse with varying levels of dermoscopy experience (Bourne 2012) and by an expert dermatologist (Rosendahl 2011). Diagnosis was made without the aid of a published algorithm.

### Studies in referred participants

Nine evaluations of clinical images were conducted in participants referred for specialist assessment; one could be clearly positioned on the clinical pathway and eight did not provide sufficient information for a clear assessment to be made.

One study clearly positioned on the clinical pathway was considered to have included all participants referred for further assessment (Pathway 4-c in Figure 10):

- sensitivity was 74.2% (95% CI: 55.4, 88.1%) and specificity 82.5% (95% CI: 73.8, 89.3%) (134 lesions and 31 melanomas (Stanganelli 2005)).

The remaining 8 studies did not provide sufficient information to clearly position them on the clinical pathway but were assumed to have obtained lesion images from referral settings (Pathway 5-u and 5*u in Figure 10):

- summary sensitivity was 60.3% (95% CI: 49.2, 70.5%) and specificity 77.0% (95% CI: 63.9, 86.4%) (293 lesions and 96 melanomas) for six studies that included all lesions selected for excision (Pathway 5-u) (Benelli 2001; Carli 2002b; Dolianitis 2005; Pizzichetta 2004; Stanganelli 1998; Winkelmann 2016).
- summary sensitivity was 61.9% (95% CI: 46.7, 75.0%) and specificity 81.8% (95% CI: 75.2, 87.0%) (303 lesions and 98 melanomas) across two studies that included participants with equivocal lesions selected for excision (Pathway 5*u) (Carli 2003b; de Giorgi 2012).

Studies were retrospective case series apart from two case control type studies (Winkelmann 2016; Dolianitis 2005) and one with an unclear design (Benelli 2001). Three studies (Benelli 2001; Dolianitis 2005; Stanganelli 1998) evaluated observer accuracy before and after dermoscopy training. Images of pigmented or melanocytic lesions were reviewed in all studies apart from one focused on hypomelanotic (<=30% pigmentation) or amelanotic lesions (Pizzichetta 2004). The prevalence of melanoma ranged from 19% (Carli 2002b) to 50% (Dolianitis 2005); four studies included only melanomas (including in situ) and benign naevi (Carli 2003b; de Giorgi 2012; Stanganelli 2005; Winkelmann 2016).

Lesion diagnosis was undertaken by dermatologists or by observers with mixed qualifications; observer experience was poorly reported (Appendix 11). Stanganelli 2005 also provided accuracy data for the average of three GPs (data reported in section 1.3.2). Most studies presented average accuracy across observers; only two reported accuracy for a single observer (Benelli 2001; Pizzichetta 2004). Diagnoses were made without the use of diagnostic algorithms in all studies except Benelli 2001 (ABCDE algorithm) and de Giorgi 2012 (ABCD).

## Secondary analyses

Secondary analyses were conducted for the detection of invasive melanoma and melanocytic intra-epidermal variants regardless of classification by clinical pathway.

### Covariate investigations

A preliminary analysis across the 39 datasets contributing to the primary analyses described above found a large difference in accuracy for in-person evaluations compared to those based on the assessment of clinical images (relative diagnostic odds ratio (RDOR) 8.54, 95% CI 2.89, 25.3, P < 0.001) (Table 2; Figure 11). The magnitude and importance of the observed difference is so large, raising serious concerns about the applicability of visual inspection studies done via image observation only, that we elected to undertake all subsequent covariate investigations based on in-person evaluations only (n = 28).

For the 28 in-person evaluations, only one of the four covariate investigations approached statistical significance (Table 2); observed accuracy was lower in studies where disease prevalence of melanoma (percentage of cases in the study which tested positive for the reference standard) was over 10% compared to those with disease prevalence of 10% or less (RDOR 0.31, 95% CI 0.09, 1.00; P = 0.05). The RDOR for study setting (secondary care or specialist clinic compared to primary care) was 1.51 (95% CI 0.32, 7.09; P = 0.59; Figure 12), for use of a named algorithm to aid diagnosis compared to no algorithm reported was 1.03 (95% CI 0.25, 4.34; P = 0.96; Figure 13), and for use of histology plus clinical follow-up or other reference standard compared to histology alone was 0.76 (95% CI 0.14, 4.02; P = 0.74; Figure 14).

**Analyses by algorithms used to assist visual inspection**

Of the 28 in-person evaluations only seven reported using an algorithm to assist visual inspection, limiting the ability of meaningful comparisons between algorithms to be made (Table 3). Observer diagnosis without the use of a formal algorithm (n = 21 datasets) had the highest diagnostic accuracy (DOR 46.2, 95% CI 21.9, 97.5), with an average sensitivity of 78% (95% CI 68, 85%) and average specificity 93% (95% CI 88, 96%). Pooled sensitivity was slightly higher and specificity slightly lower for variations on the (A)BCD(E) algorithm (n = 6 datasets), but with overlapping confidence intervals (summary sensitivity 83% (95% CI 75, 88%); summary specificity 88% (95% CI 64, 97%)). Two datasets reported data for either the original seven point checklist at a number of thresholds (McGovern 1992) or for the revised 7 point checklist (Walter 2012). At the standard threshold of >=3 for both algorithms, the highest observed sensitivity and specificity was 94% (95% CI 73, 100) and 80% (95% CI 77, 83%) for the revised version (Walter 2012).

The image-based evaluations reported data for either no algorithm or for variations of ABCD(E); a similar pattern was observed with much lower levels of overall accuracy (Table 3).

**Analyses by observer experience**

Analyses by observer expertise were restricted by the limited amount of information provided in the study reports (Table 4; Appendix 9 and Appendix 11). Our analyses are therefore based primarily on study subgroups by observer qualifications (consultant/registrar/mixed qualifications/primary care practitioners), with the ‘consultant’ category separated into ‘Expert consultant’ (for any study describing observers as expert or experienced) and ‘Consultant’ where experience or expertise was not otherwise reported (for example, for those that described observers as dermatologists) (Table 4; Figure 15).

No clear pattern according to observer experience could be discerned for in-person evaluations. Relative DORs in comparison to the ‘Expert consultant’ group (9 studies) ranged from 0.45 (95% CI 0.05, 3.67; P=0.44) for observers at resident/registrar level (2 studies) to 7.28 (95% CI 0.69, 76.3; P=0.09) for GPs (3 studies).

For image-based evaluations, accuracy was highest for the ‘Expert consultant’ group (DOR 20.5 (95% CI 4.82, 86.9); relative DORs in comparison to the ‘expert’ group ranged from 0.18 (95% CI 0.04, 0.90; P=0.04) for observers described as ‘dermatologists’ (4 studies) to 0.56 (95% CI 0.04, 7.51; P=0.63) for mixed secondary and primary care observers (1 study).

Across all definitions of the target condition, 7 studies provided comparative data according to observer qualifications or experience (Table 5). Most were image-based assessments, using no prescribed algorithm to aid diagnosis and reporting average results across groups of observers. Some evidence of increased sensitivity and smaller increases in specificity were observed with increasing experience; however, wide variations in accuracy remained with sensitivity ranging from 58% to 91% for expert dermatologists and specificities from 53% to 99%.

### *2. Target condition: invasive melanoma only*

In this section, we present the results for studies of visual inspection for the identification of invasive melanoma, according to the approach taken for diagnosis: in-person or image-based evaluations. Summary characteristics of studies are presented in Appendix 13 and results of meta-analyses in Table 6, and Figure 16. Table 7 compares results in studies reporting data for invasive melanoma alone and for invasive melanoma plus atypical intraepidermal melanocytic variants.

Seven datasets evaluated the accuracy of in-person visual inspection for the detection of invasive melanoma (Bono 1996; Green 1994; Kopf 1975; Krahn 1998; McGovern 1992; Viglizzo 2004; Walter 2012), only two of which also reported data for the primary target condition (McGovern 1992; Walter 2012). All studies were based in secondary care or specialist units apart from Walter 2012 (primary care) and McGovern 1992 (army medical centre dermatology clinic). Studies used a modified version of the ABCD checklist (McGovern 1992), the revised 7-point-checklist (Walter 2012 ), or no algorithm (n = 5; 71%) to assist diagnosis. The prevalence of melanoma ranged from 2% (Kopf 1975; Walter 2012 ) to 49% (Krahn 1998). Two studies supplemented a histological reference standard with clinical follow-up (Walter 2012 ) and expert diagnosis of some benign lesions (Walter 2012; Bono 1996).

Sensitivities ranged from 67% to 100% and specificities ranged from 76% to 100%. In meta-analysis the DOR was 62.4 (95% CI 17.6, 222) (6857 lesions and 208 melanoma cases). Sensitivity and specificity at the average operating point on the SROC curve were 86% (95% CI: 68, 94%) and 91% (95% CI: 81, 96%) respectively. For the two in-person evaluations which also reported data for the primary target condition (Table 7), specificity estimates were hardly affected due to small numbers of included melanoma in situ lesions (5 in McGovern 1992 and 2 in Walter 2012). Sensitivity however, was higher for detection of invasive melanoma alone in McGovern 1992 (100% versus 73% for detection of invasive melanoma or atypical intraepidermal melanocytic variants) due to correct diagnosis of only two of five *in situ* melanomas, and was marginally lower in Walter 2012 (93.8% versus 94.4% for detection of invasive melanoma or atypical intraepidermal melanocytic variants) due to correct identification of both *in situ* melanomas with one invasive melanoma missed.

Five datasets reported the accuracy of image-based visual inspection for the detection of invasive melanoma (Lorentzen

1999; Rao 1997; Scope 2008; Troyanova 2003; Westerhoff 2000), but none of which reported data for the primary target condition. Only two studies used images from normal practice settings (Lorentzen 1999; Rao 1997); one obtained images from a teledermatology company (Scope 2008) and two selected images of melanoma cases and controls for use in dermoscopy training studies (Troyanova 2003; Westerhoff 2000). The prevalence of melanoma ranged from 3% (Scope 2008) to 50% (Troyanova 2003; Westerhoff 2000). The ABCD checklist (Rao 1997), the ugly duckling approach (Scope 2008), or no algorithm (n = 3) was used to assist diagnosis. Four evaluations clearly presented only the clinical image with no further patient information (80%), and one (Rao 1997) may have presented observers with a concurrent dermoscopic image of the lesion as blinding between images was not clearly described.

Sensitivities ranged from 62% to 86%; specificities ranged from 54% to 95%. In meta-analysis the DOR was 14.8 (95% CI 3.56, 61.9) [599 lesions and 150 melanoma cases]. Sensitivity and specificity at the average operating point on the SROC curve were 76% (95% CI: 50, 91%) and 83% (95% CI: 62, 93%) respectively.

Accuracy was non-significantly higher for in-person compared to image-based evaluations (RDOR 4.21; 95% CI 0.62, 28.6; P = 0.13).

### *3. Target condition: any skin lesion requiring excision*

In this section, we present the results for studies of visual inspection for the identification of any skin lesion requiring excision (for each study, data could only be extracted for the detection of any skin cancer), according to the approach taken for diagnosis: in-person or image-based evaluations. Summary characteristics of studies are presented in Appendix 14 and results of meta-analyses in Table 6 and Figure 17. Table 7 compares results in studies reporting data for invasive melanoma alone and for invasive melanoma plus atypical intraepidermal melanocytic variants.

Seven datasets evaluated the accuracy of in-person visual inspection for the detection of any skin lesion requiring excision (Argenziano 2006; Chang 2013; Ek 2005; McGovern 1992; Stanganelli 2000; Steiner 1987; Walter 2012), five of which also reported data for the primary target condition (Ek 2005; McGovern 1992; Stanganelli 2000; Steiner 1987; Walter 2012). Three studies were based in primary care (Argenziano 2006; Walter 2012) or community dermatology clinics (McGovern 1992), the others were based in secondary care or specialist referral clinics. The prevalence of skin cancer ranged from 3% (Walter 2012) to 68% (Ek 2005). Studies used the ABCD algorithm (Argenziano 2006; McGovern 1992; Stanganelli 2000), the revised 7-point-checklist (Walter 2012), or no algorithm (n = 3) to assist diagnosis. Two studies supplemented a histological reference standard with clinical follow-up (Stanganelli 2000; Walter 2012) and expert diagnosis of some benign lesions (Walter 2012).

Sensitivities ranged from 57% to 98%; specificities ranged from 13% to 99%. In meta-analysis the DOR was 20.5 (95% CI 7.11, 59.3) (8091 lesions and 2187 skin cancer cases). Sensitivity and specificity at the average operating point on the SROC curve were 81% (95% CI: 68, 90%) and 81% (95% CI: 56, 93%) respectively. For the in-person evaluations which also reported data for the primary target condition (Table 7), specificity estimates were not affected in four of the five studies due to the relatively small percentage of other skin cancers in the study populations (BCCs making up 2% of all lesions in McGovern 1992; 1% in Stanganelli 2000 and Walter 2012; and 6% in Steiner 1987). Sensitivities increased in two studies due to a majority of BCCs correctly identified (Stanganelli 2000; Steiner 1987); sensitivity fell in Walter 2012 due to 3 of 4 BCCs not picked up by the revised seven point checklist; and remained the same in McGovern 1992. A large increase in sensitivity and fall in specificity was observed in Ek 2005, however, as BCCs made up 47% of the total study population and invasive SCCs comprised 20%. When these two lesion groups were considered as disease positive, sensitivity increased from 48% to 98% and specificity fell from 99% to 13% due to the largely correct identification of BCC and SCC as malignant and high false positives in the remaining group of lesions considered disease negative (including large proportions with Bowens disease, solar keratoses, or seborrhoeic keratoses).

Three datasets reported the accuracy of image-based visual inspection for the detection of any skin lesion requiring excision (Carli 2002b; Rosendahl 2011; Stanganelli 1998), all of which also reported data for the primary condition. All studies selected images from normal practice settings, two in secondary care (Carli 2002b; Stanganelli 1998) and one from a primary care practice (Rosendahl 2011). The prevalence of lesions suitable for excision ranged from 22% (Rosendahl 2011) to 47% (Stanganelli 1998); the latter selecting images for use in a dermoscopy training study. Data were presented for a single dermatologist (Rosendahl 2011), a consensus of two dermatologists (Carli 2002b ), or the average across 20 dermatologists (Stanganelli 1998). No algorithm was used to assist diagnosis (n = 3) and no further patient information was presented to assist diagnosis.

Sensitivities ranged from 64% to 80%; specificities ranged from 74% to 85%. In meta-analysis the DOR was 11.9 (95% CI 2.22, 65.3) (547 lesions and 138 skin cancer cases). Sensitivity and specificity at the average operating point on the SROC curve were 75% (95% CI: 49, 90%) and 79% (95% CI: 38, 96%) respectively. For the three studies which also reported data for the primary target condition (Table 7), sensitivities increased in two due to correct identification of BCCs (Rosendahl 2011; Stanganelli 1998). Specificity decreased in Carli 2002b due to small sample size and high prevalence of malignancy (20/53; 38%) and decreased in Rosendahl 2011 due to the use of a different threshold for the primary target condition 'is this lesion a melanoma?' compared to ‘should this lesion be excised?’ for the target condition of any lesion requiring excision.

No significant difference in accuracy between in-person and image-based evaluations was identified (RDOR 1.70; 95% CI 0.24, 12.3; P=0.55).

# Discussion

## Summary of main results

## Summary of main results

Visual inspection has been evaluated in a range of study populations, on an in-person basis and using clinical images, and both with and without the use of published algorithms to assist diagnosis. Wide variations in sensitivity and specificity were observed for all definitions of the target condition.

There are five main findings from our review:

1) There is an almost universal problem with poor reporting in the primary studies, hindering attempts to analyse studies according to their position on the clinical pathway and to fully assess sources of heterogeneity and methodological quality.

Less than two thirds of in-person evaluations of visual inspection contained enough information to describe where on the clinical pathway participants were assessed. This was particularly the case for studies apparently conducted in referred populations, where almost half of studies neither described participants as 'referred', nor provided any description of participants' prior testing or pathway followed prior to presentation for specialist review. Observer experience and expertise in pigmented lesion diagnosis is likely to impact on test accuracy; however, this information was rarely provided in any detail making it difficult to assess any differences in accuracy according to clinician experience. Analyses by reported observer qualifications and descriptions of observers as 'expert' or 'experienced' showed no significant differences between groups.

In terms of methodological quality, studies were at unclear risk of bias due to poor reporting of key items around participant selection, pre-specification of thresholds used, and timing of diagnosis in relation to reference standard diagnosis. Concern around applicability of studies was almost universally poor due to restricted inclusion of lesions and lack of reproducibility of diagnostic thresholds. Given these limitations and the heterogeneity in various aspects of the primary studies, our results cannot be considered conclusive regarding the accuracy of visual inspection for melanoma diagnosis.

2) Prior testing of participants or study position on the clinical pathway does appear to matter.

Focusing on in-person evaluations that could be clearly positioned on the clinical pathway (Summary of findings table 1), the highest sensitivity (92.4%, 95% CI 26.2, 99.8%) and lowest specificity (79.7%, 95% CI 73.7, 84.7%) for the primary target condition of invasive melanoma or atypical intraepidermal melanocytic variants was observed in three datasets from participants with limited prior testing; however, confidence intervals were wide and heterogeneity high, particularly for sensitivity. Data for referred participants suggest that summary sensitivities fall to around 75%, but with much higher specificities (e.g. sensitivity 76.7% (95% CI 61.7, 87.1%) and specificity 95.7% (95% CI 89.7, 98.3%) for lesions selected for excision, n = 8 datasets). Sensitivity was higher for equivocal lesion populations but with very wide confidence intervals (84.7%, 95% CI 55.5, 96.1%) with summary specificity of 89.5% (95% CI 79.5, 95.0%) (2 datasets).

The general trade-off between sensitivity and specificity along the pathway could be due to differences in the spectrum or 'case mix' of included lesions, differences in the definition of a positive test result, or may be linked to variations in observer expertise. Spectrum effects can be observed when tests that are developed further down the referral pathway have lower sensitivity and higher specificity when applied in settings with participants with limited prior testing (Usher-Smith 2016). Classic examples include the use of dipstick tests for detection of urinary tract infection (UTI) (Lachs 1992) and the D-dimer test to detect pulmonary embolism (PE) (Ginsberg 1993). In both studies, as the prior probability of having UTI or PE increases (and so prevalence of disease increased), test sensitivity increased (from 79% to 93% in Ginsberg 1993, and from 58% to 92% in Lachs 1992) while specificities decreased (from 76% to 45% in Ginsberg 1993 and from 77% to 42% in Lachs 1992). However, this direction of effect is not consistent across tests and diseases as Leeflang 2013 clearly demonstrates; the mechanisms in action are often more complex than prevalence alone and can be difficult to identify.

Using disease prevalence as a proxy for disease spectrum, our classification of studies did result in a somewhat lower prevalence of disease (suggesting a wider spectrum of lesion types) in limited prior testing studies (median prevalence 5%, interquartile range (IQR) 3%, 9%) compared to referral settings (median prevalence 15%, IQR 10, 21%), but with overlapping ranges (2% to 11%, and 1% to 41%, respectively). The lower specificity observed in limited prior testing studies is likely related to the presence of a wider range of benign lesions with similar characteristics to melanoma leading to more referrals. Observers in primary care are also likely to have a lower threshold for considering benign lesions as possibly malignant due to the risk of missing true cases of melanoma, contributing both to higher sensitivity and a higher false positive rate. Referred populations on the other hand may have a higher proportion of equivocal or 'difficult to diagnose' melanomas that are difficult to identify.

In terms of eligibility criteria, varying degrees of clinical suspicion of malignancy were required for lesion inclusion in limited prior testing populations ranging from lesions that could not immediately be diagnosed as benign, to there being a requirement for a teledermatology second opinion. In referral populations, eligibility was frequently based on lesion excision, the basis or rationale for which was not described. The restriction to lesions deemed to be suitable for excision will decrease specificity, as more obviously benign lesions would be excluded. The spectrum of lesion types in the disease negative groups also varied across studies, with a number of studies restricting inclusion only to those with melanocytic lesions (such that all benign lesions were benign melanocytic naevi) and others reporting high proportions of other types of skin cancers (BCC or SCC), or of benign keratotic lesions such as seborrhoeic or actinic keratoses, or of Spitz naevi which may be difficult to differentiate from melanoma.

3) Visual inspection alone is not sufficiently sensitive for the detection of melanoma, and there is no clear evidence that accuracy is improved by the use of any named or published algorithm to assist diagnosis in all settings.

Test sensitivity was greater than 90% (i.e. less than 1 in 10 melanomas missed) in only 6 of the 28 in-person based evaluations of the primary target condition and confidence intervals for the pooled estimates were wide, raising the question

as to whether visual inspection can be relied on to rule out the presence of melanoma. Applying the sensitivity and specificity estimates for the limited prior testing studies cited above to a hypothetical cohort of 1000 lesions at disease prevalence of 4%, 9%, and 16% (see Summary of findings table 1) shows that visual inspection would *on average* miss 3, 7 or 12 melanomas, with 195, 185 and 171 false positive results (potentially leading to unnecessary excisions or lesion referral or follow-up depending on the anticipated clinical action following a positive result). The wide confidence intervals however mean that the number of melanomas missed could range from between 0 and 118, with false positives from 129 to 252. For a cohort of 1000 lesions in a referred population at prevalence of 4%, 9%, and 16% ( Summary of findings table 1), the pooled sensitivity of 76.7% and specificity 95.7% translate to 9, 21, and 37 melanomas missed on average (range: 5 to 61) and 41, 39, and 36 false positive results (range: 14 to 99).

The evidence to support the use of available algorithms to assist visual inspection was limited, and results are likely to be confounded by patient spectrum and observer experience. Considerable variation in definitions of test positivity across studies that did not report using any algorithm was also observed, i.e. where observer diagnosis was based on observers' own interpretation of lesion characteristics. Where reported, visual inspection was considered to be positive for observers 'correct diagnosis of melanoma', 'suspicion of malignancy', or 'selection for excision', each of which is likely to result in varying proportions of test positive or negative for any given population.

Nevertheless, covariate investigations for the primary analysis across all study settings suggested no difference in accuracy according to the reported use of any named or published algorithm to assist diagnosis. This result was supported by limited subgroup analysis according to algorithm used. Only one eligible study directly compared the accuracy of visual inspection with and without the use of an algorithm (Collas 1999); however, the authors developed their own new algorithm for the study and found sensitivity to be higher without the use of the algorithm. Comparing different algorithms, McGovern 1992 reported highest sensitivities from the BCD algorithm (any one characteristic present) and the original seven point checklist (at least two characteristics present). Current guidelines in the UK support the use of the revised seven point checklist in primary care (NICE 2015). A number of studies assessing the revised seven point checklist algorithm did not meet the stringent inclusion criteria for our review (Healsmith 1994; Higgins 1992; Osborne 1999; Walter 2013); however, the single eligible study using the revised seven point checklist as part of a large randomised controlled trial reported high sensitivity (94%) when used by GPs (Walter 2012).

4) The definition of the target condition has an effect on diagnostic accuracy.

Results from studies reporting data for more than one definition of the target condition show that sensitivity in particular is affected by the inclusion of, and percentage of, melanoma in situ and BCC lesions considered disease positive. The direction of effect depends on observers' ability to correctly identify these lesions as malignant. It is likely that similar effects impact on results observed across all included studies. Clear identification of the target condition was not provided in 11 of the 28 datasets included in our primary analyses so that the inclusion of melanoma in situ lesions as disease positive was assumed on the basis that the disease positive group was described as 'melanoma' and not as 'invasive melanoma' or 'malignant melanoma'. Of those studies that clearly reported including in situ lesions, the percentage of the disease positive group (invasive melanoma and atypical intraepidermal melanocytic variants) described as being in situ ranged from 10% to 50%. Where studies included other invasive skin cancers (mainly BCCs or SCCs) in the study population (lesions considered disease negative for detection of the primary target), we attempted to class any that were correctly identified by observers as malignant as 'true negative' results as opposed to 'false positives' (thereby increasing observed specificities), on the basis that removal of any skin cancer in the attempt to identify melanomas would not be a negative consequence of the test. Our ability to reclassify lesions relied on studies providing a disaggregation of test results according to final lesion classification and was not always possible, particularly when invasive SCCs were not separated from 'in situ' lesions such as Bowen's disease.

5) There are substantial differences in diagnostic accuracy between in-person and image-based assessments.

Accuracy was much lower and reporting was poorer for evaluations of a diagnosis based on the interpretation of clinical images as opposed to in-person evaluations. Other than possible differences in patient spectrum between in-person and image-based studies, one possible explanation for the observed difference is that even using the highest quality clinical image, a remote assessment is not equivalent to a physical, face-to-face patient to clinician interaction, which will include patient history-taking as well as a total body examination. We were unable to examine any impact from history-taking over and above inspection of the lesion itself; however, history-taking and in particular, assessment of and knowledge of patients' other lesions could have a significant impact on the decision as to whether or not a patient has melanoma (Grob 1998; Aldridge 2013). Subtle differences in assessing the lesion shape and colour can be done in an in–person consultation, e.g. by stretching the lesion in the axis perpendicular to the skin creases, which may distort the lesion shape, and by altering the light intensity and direction used during lesion inspection. Palpation of the lesion (and regional lymph nodes) is also possible during in-person examination. The fact that image quality is likely to vary between studies, the time taken to review each image is likely to vary, and the considerable variation in supplementary information provided to observers (ranging from no clinical information, to clinical details regarding patient age, gender or lesion site and information on lesion change over time) will have further contributed to variation in accuracy and lower accuracy estimates in comparison to in-person evaluations. Furthermore, the diagnostic context may have a key influence on observer decisions. In a face-to-face diagnostic encounter and for the examination of lesion images for a teledermatology consultation, the clinicians concerned know that their assessment has a direct consequence on patient management and potentially on patient outcomes. The image-based evaluations included in our primary analysis however were not conducted for teledermatology purposes, but were studies using lesion images to compare accuracy between clinical image diagnosis and dermoscopic image diagnosis, or to compare observer or algorithm performance, for example. Observers would have been aware that their assessment of the lesion image was done in an experimental setting, and would not impact on patients; this could

potentially have affected interpretation.

### Strengths and weaknesses of the review

The strengths of our 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 according to approach to diagnosis, the definition of the target condition, and the patient pathway was adopted to estimate test accuracy in different study populations. A detailed and replicable analysis of methodologic quality was undertaken.

In comparison to other available systematic reviews, our review extends the time period searched for eligible studies to August 2016 (from 2007 in Vestergaard 2008 and from March 2015 in Harrington 2017), and we include all eligible studies regardless of availability of a direct comparison with dermoscopic examination (as required in Vestergaard 2008) or requirement for an algorithm or clinical prediction rule to be included (Harrington 2017). Our stringent application of review inclusion criteria meant that several otherwise eligible studies were excluded. For example, those reporting accuracy data for 'clinical diagnosis' where dermoscopy may or may not have been used to assist diagnosis were excluded on the basis that the contribution of visual inspection of the lesion could not be discerned.

Studies evaluating eligible algorithms (that were included in Harrington 2017) were excluded from our review, due to lack of data to construct a 2x2 contingency table, the serial use of the algorithm in the context of lesion follow-up, or use of inadequate reference standards. Without these restrictions, the observed data would likely have been considerably more heterogeneous and of poorer methodological quality. At the same time, our inclusion of all studies reporting data for visual inspection means that an overall assessment of observer accuracy could be made regardless of the use of a named algorithm. Harrington and colleagues rightly point out that lower sensitivity associated with the use of a clinical prediction rule "should not prevent [its] use unless usual decisions, made without the rule, are demonstrably better"; however, unless the accuracy of 'usual decisions' is examined, any benefit from the use of an algorithm cannot be established.

The main concerns for the review are a result of the poor reporting of primary studies, in particular forcing some assumptions to be made to allow studies to be split by pathway and in separating studies by the different definitions of the target condition. Our inability to clearly separate studies by pathway is of real concern given the evidence for the effect on accuracy according to the spectrum or case-mix of included participants (Lachs 1992; Moons 1997; Leeflang 2013).

Finally, observer expertise is key for any diagnostic process based on visual inspection, with both non-analytical pattern recognition (implicit identification) and analytical pattern recognition (using more explicit 'rules' based on conscious analytical reasoning) employed to varying extents between clinicians, according to factors such as experience and familiarity with the diagnostic question (Norman 2009). A lack of clear reporting of observer training and experience made analysis difficult.

### Applicability of findings to the review question

Varying definitions of the eligible study populations and lack of clarity regarding the patient pathway and any prior testing may restrict the applicability of our findings to the clinical setting. Varying definitions of test positivity and lack of reproducibility of diagnostic thresholds, variability in the use of published algorithms, and in observer qualifications and experience, further restrict the transferability of results to a clinical setting.

## Authors' conclusions

### Implications for practice

Visual inspection is an essential, fundamental component of the assessment of a suspicious skin lesion; however, the evidence suggests that melanomas will be missed if visual inspection is used on its own. The evidence to support its accuracy in the range of settings in which it is used is both flawed and poorly reported, resulting in an inability to produce meaningful summary results and clear pointers as to where visual inspection is most useful. Overall, the use of published algorithms to assist diagnosis does not appear to improve accuracy; however, neither is there sufficient evidence to suggest that the 'no algorithm' approach should be preferred in all settings, e.g. for training junior staff. Further investigation may lend support to the theory that expert observers are more reliant on non-analytical pattern recognition, while attempts to assist analytical pattern recognition are of more benefit for less experienced or more generalist observers.

### Implications for research

Despite the vast volume of research that has been funded to evaluate visual inspection, further prospective evaluation of the added value of established algorithms according to the prior testing or diagnostic difficulty of lesions may be warranted. Prospective recruitment of consecutive series of participants and with systematic follow-up of non-excised lesions to avoid over-reliance on a histological reference standard would allow results to be more generalisable to routine practice. A clear identification of the level of training and experience required to achieve good results is also required. Any future research study needs to be clear about the diagnostic pathway followed by study participants prior to study enrolment, and 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 advisor and co-author 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 Michael Bigby, who was the Dermatology Editor for this review; and the clinical referees, Andrew Affleck and Chris Bower. We also wish to thank the Cochrane DTA editorial base and colleagues.

## Contributions of authors

JD was the contact person with the editorial base.
JD co-ordinated contributions from the co-authors and wrote the final draft of the review.
JD, NC, LFR, DT, KYW, RBA, RA, and MF screened papers against eligibility criteria.
JD and NC obtained data on ongoing and unpublished studies.
JD, NC, LFR, DT, KYW, RBA, RA, and MF appraised the quality of papers.
JD, NC, LFR, DT, KYW, RBA, RA, and MF extracted data for the review and sought additional information about papers.
JD entered data into RevMan.
JD, MJG and JJD analysed and interpreted data.
JD, JJD, NC, LFR, YT and CD worked on the methods sections.
JD, FW, DT, KYW, RBA, RA, MF, RNM and HCW drafted the clinical sections of the background and responded to the clinical comments of the referees.
JD, JJD, CD and YT 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

Jacqueline Dinnes: I am employed by the University of Birmingham under a National Institute for Health Research (NIHR) Cochrane Programme Grant (13-89-15) to produce this review.
Jonathan J Deeks: The project was funded by the NIHR Cochrane Programme Grant. Jon Deeks received additional support as an NIHR Senior Investigator and from the NIHR Birmingham Inflammation Biomedical Research Centre.
Matthew J Grainge: nothing to declare.
Naomi Chuchu: nothing to declare.
Lavinia Ferrante di Ruffano: The NIHR Cochrane Systematic Review Programme Grant funded the project but to my knowledge did not influence its planning or conduct.
Rubeta N Matin: nothing to declare.
David R Thomson: nothing to declare.
Kai Yuen Wong: nothing to declare.
Roger Benjamin Aldridge: nothing to declare.
Rachel Abbott: nothing to declare.
Monica Fawzy: nothing to declare.
Susan E Bayliss: nothing to declare.
Yemisi Takwoingi: nothing to declare.
Clare Davenport: nothing to declare.
Kathie Godfrey: I have received reimbursement of travel expenses incurred by attending meetings.
Fiona M Walter: nothing to declare.
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.

## Differences between protocol and review

We set out to review visual inspection and dermoscopy for the detection of melanoma in a single review; however, due to the volume of evidence identified, two separate reviews were prepared: one for visual inspection alone and one for dermoscopy, the latter including direct comparisons with visual inspection where both tests were evaluated in the same studies.

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

Secondary objectives have been tailored to the individual test, with two objectives added: to determine the diagnostic accuracy of individual algorithms for visual inspection; and to determine the effect of observer experience.

Sources of heterogeneity that could be investigated (as listed under Secondary objectives) were restricted due to lack of data.

Studies using cross-validation, such as 'leave-one-out' cross-validation were excluded rather than included as these methods are not sufficiently robust and are likely to produce unrealistic estimates of test accuracy.

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 algorithms or methods of diagnosis (i.e. derivation studies) were included if they:

- used a separate independent 'test set' of participants or images to evaluate the new approach, or
- investigated lesion characteristics that had previously been suggested as associated with melanoma and the study reported accuracy based on the presence or absence of particular combinations of characteristics.

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)
- evaluated the accuracy of the presence or absence of individual lesion characteristics or morphological features, with no overall diagnosis of malignancy
- reported accuracy data for 'clinical diagnosis' with no clear description as to whether the reported data related to visual inspection alone
- were based on the experience of a particular skin cancer clinic, where dermoscopy may or may not have been used on an individual patient basis."

Although we extracted any reporting of special interest or accreditation in skin cancer according to observer expertise, we were unable to analyse the effect on accuracy.

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. Sensitivity analyses were not performed as planned due to lack of data.

# Published notes

# Characteristics of studies

## Characteristics of included studies

### *Argenziano 2006*

#### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Randomised controlled trial allocating primary care physicians to use either visual inspection alone or visual inspection plus dermoscopy (only excised lesions can be included for each arm).<br>**Data collection:** Prospective<br>**Period of data collection** May 2003 to Sept 2004<br>**Country** Italy and Spain |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Patients asking for screening or exhibiting one or more skin tumours as seen during routine physical examination (patient-finding screening) were considered for inclusion; those undergoing excision were included in this review (i.e. those deemed sufficiently suspicious by the Expert evaluation). PCPs were invited to participate in the trial; only those who attended the training sessions and who then screened patients and referred them to the Pigmented Lesion Clinics were randomised.<br>**Setting:** Primary<br>**Prior testing:** No prior testing<br>**Setting for prior testing:** N/A<br>**Exclusion criteria:** NR<br>**Sample size (patients):** No. eligible: 3271 patients screened; 1325 patients allocated to Naked Eye observation and 1197 patients allocated to dermoscopy observation; No. included: 162 received histology after Expert evaluation at the PLC<br>**Sample size (lesions):** 85 in VI arm and 77 in Dermoscopy arm underwent excision<br>**Participant characteristics:** Based on full sample: mean age 40, range 2-90 (visual inspection group)/ 41, range 3-94 (dermoscopy group). Male 498 (38%): VI group / 451 (38%) dermoscopy<br>**Lesion characteristics** NR |
| 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 | **Visual inspection (VI)** ABCD (control arm of RCT comparing naked eye examination to naked eye plus dermoscopy)<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** Qualitative NR; Described in Intro as: simple morphologic features summarized by the asymmetry, border irregularity, color variegation, and diameter 5 mm (ABCD)<br>**Diagnosis based on:** Average (n=37)<br>**Observer qualifications:** Primary care physicians<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described<br>**Other detail:** Pre-randomisation all participating PCPs underwent training in ABCD rule for clinical diagnosis and 3-point checklist for dermoscopy.<br>**Dermoscopy:** evaluated in intervention arm of trial only. |

## Visual Inspection - in-person

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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 |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>*Details:* All lesions considered suggestive of skin cancer at the PLC were excised and subsequently diagnosed histopathologically. Equivocal lesions by histopathologic examination were reviewed by a second independent pathologist and a final diagnosis made.<br>Disease positive: 92 malignant tumours; Disease negative: 70 benign tumours<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 12; BCC: 66; cSCC: 14<br>Sebhorrheic keratosis: 13; Melanocytic nevi 51; Other: 6 |
| 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 |
| **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 | **Excluded participants:** Data can only be extracted for those with histology (i.e. patients considered to have lesions suggestive of skin cancer); remainder had expert diagnosis (not included in the final 2x2 data extracted)<br>**Time interval to reference test:** Not reported |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

## Notes

| Notes | - |
|---|---|

## *Barzegari 2005*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Not reported<br>**Period of data collection** NR<br>**Country** Iran |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented skin lesions with a clinical diagnosis of melanocytic lesion <=15mm diameter referred to dermatology clinic for diagnostic evaluation or cosmetic reasons<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion; Patient request for evaluation/excision<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. included: 91<br>**Sample size (lesions):** No. included: 122<br>**Participant characteristics:** Mean age 32.3 (6-94y); Male: 30; 33%<br>**Lesion characteristics** NR |
| Are the included patients and chosen study setting appropriate? | Unclear |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

### Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** Qualitative Melanoma likely (i.e. melanoma first in list of considered diagnoses)/ Melanoma possible (melanoma one of a number of diagnoses)<br>**Diagnosis based on:** Consensus (2 observers); n=2<br>**Observer qualifications:** Dermatology registrar (Dermatology resident (3rd year)); Dermatologist<br>**Experience in practice:** Mixed experience (low and high experience combined)<br>**Experience with index test:** Mixed (low and high experience combined) |

### Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

### Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 6; Disease negative: 116<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 3; Melanoma (in situ): 3<br>Sebhorrheic keratosis: 2; Benign naevus: 104; Dysplastic naevus 7 DF 1 AK |
| 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 |
| **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? | 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 | Excluded participants: None;<br>Time interval between index and reference: Unclear<br>Time interval between index test(s): Consecutive |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | No |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Benelli 1999*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** 01/09/1997 to 30/09/1998<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** All pigmented skin lesions observed and excised at the Dermatologic Surgery Department<br>**Setting:** Dermatologic Surgery Department<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Dermatologic Surgery Department<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** NR<br>**Sample size (lesions):** No. included: 401<br>**Participant characteristics:** NR<br>**Lesion characteristics:** Thickness 42 < 0.75 mm thick, 80.76-1.5 mm thick. 4 1.5-4 mm thick (mean 0.60 mm, median 0.55 mm. max 1.9 mm, min 0.10 mm, SD 0.45). |
| 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 | **Visual inspection (VI)** ABCDE<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Lesions assessed by both dermatologists clinically and dermoscopically<br>**Diagnostic threshold:** Data given for accuracy of each potential score (1-5); score estimation described in detail<br>**Diagnosis based on:** Consensus (2 observers); n= 2<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described<br>**Dermoscopy** 7FFM also assessed by same observers |

## Visual Inspection - in-person

| 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? | No |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | High 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 |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

### Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 60 (15%) lesions; Disease negative: 340 (non melanoma) + 1 BCC<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 54 (13.5%); Melanoma (in situ): 6 (1.5%); BCC: 1 (0.4%)<br>Sebhorrheic keratosis: 1 (0.4%); Melanocytic nevi: 316; Epithelioid and/or spindle cell nevi: 18 (4.5%); Lentigo simplex: 5 (1.2%) |
| 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 |
| **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? | 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 | Excluded participants: NR;<br>Time interval to reference test: same day |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | Yes |
| Were all patients included in the analysis? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Low risk |

### Notes

| Notes | - |
|---|---|

## *Benelli 2001*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Unclear<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** Not reported - only dates of training course and agreement study given (April-May 1999)<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Unclear |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Slides of pigmented skin tumours were selected for evaluation during a training course on dermoscopy. Lesions not located on head, palms or soles histological slide available<br>**Setting:** Training images; Authors institution. Institute of Dermatologic Sciences, University of Milan<br>**Prior testing:** Slides of pigmented skin tumours were selected for evaluation during a training course on dermoscopy<br>**Setting for prior testing:** Unspecified<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** Not reported<br>**Sample size (lesions):** No. included: 49 (paper reports 50 but only 49 accounted for in text)<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI)** ABCDE<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** No further information used<br>**Diagnostic threshold:** ABCDE Score >/=2; presence of 2 criteria; ABCDE Score >/=3; presence of 3 criteria. All criteria described in full<br>**Diagnosis based on:** Single (n=1); Average (n=65; attending one of three courses in dermoscopy held to inform dermatologists about a new dermatoscopic diagnostic method (7FFM))<br>**Observer qualifications:** Dermatologists<br>**Experience in practice:** Expert author; Not described for participating dermatologists<br>**Experience with dermoscopy:** Expert author; Prior experience not described for participating dermatologists; all underwent dermoscopy training for study purposes<br>**Dermoscopy:** 7FFM; ABCDE also evaluated in study |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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 |
| Was the test interpretation carried out by an experienced examiner? | 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<br>Disease positive: 12/49 melanomas (paper reports 50 but only 49 accounted for in text)<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 10; Melanoma (in situ): 2; BCC: 2 pigmented BCC<br>3 seborrhoeic keratoses, 2 pigmented basal cell carcinoma, 1 blue nevus, 2 angiokeratoma, 5 Spitz nevus, 5 junctional nevi, 9 compound nevi, 10 nevi undergoing regression. |
| 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 |
| **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? | 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 | Excluded participants: None reported<br>Time interval to reference test: Unclear |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

## Notes

| Notes | - |
|---|---|

### *Bono 1996*

## Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Unclear<br>**Data collection:** Not reported<br>**Period of data collection** between March 1993 and Oct 1994<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Unclear |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented skin lesions at the Instituto Nazionale Tumori of Milan.<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic) Instituto Nazionale Tumori of Milan<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 45<br>**Sample size (lesions):** No. eligible: 54/ No. included: 43<br>**Participant characteristics:** NR<br>**Lesion characteristics:** Site - Face/Ears: 3 (6%)/ Trunk: 39 (72%)/ Limbs: 12(22%); 10 MM ≤1mm depth; median size: 10mm (4 to 40mm) |
| Are the included patients and chosen study setting appropriate? | Unclear |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** Not reported; 'clinical diagnosis'<br>**Diagnosis based on:** Single observer; n= NR<br>**Observer qualifications:** treating surgeon<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard**<br>Histological diagnosis Disease positive: 18; Disease negative: 25<br>*Expert opinion* - Disease negative: 11<br>**TARGET CONDITION (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 18<br>Mild/moderate dysplasia: 8 dysplactic nevi<br>Benign naevus: 17 common melanocytic nevi |
| Is the reference standards likely to correctly classify the target condition? | No |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Unclear |
| **Could the reference standard, its conduct, or its interpretation have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Expert opinion (with no histological confirmation) was not used as a reference standard | No |
| 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?** | High |

## Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | **Excluded participants**: only 43 lesions had complete clinical and histological information. 11 lesions not surgically removed had only clinical diagnosis (benign) and were not included in the final accuracy analysis<br>**Time interval to reference test**: not reported<br>**Time interval between index test(s)**: not reported |
| Was there an appropriate interval between index test and reference standard? | Unclear |
| Did all patients receive the same reference standard? | No |
| 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? | Yes |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

## Notes

| Notes | - |
|---|---|

### *Bono 2002*

## Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** June 1998-March 2000<br>**Country** Italy<br>**Test set derived** A training set was separately derived using data obtained from 237 previously studies lesions (Farina 2000) |
| 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:** Cutaneous pigmented lesions with clinical and/or dermatoscopic features that suggested a more or less important suspicion for CM<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Clinical and/or dermatoscopic suspicion<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** Location/site of lesion - Awkwardly situated lesions eg interdigital space, ears, nose or eyelids. Lesions on scalp excluded due to hair interference with reflectance - lesion size obvious large, thick melanomas<br>**Sample size (patients):** No. included: 298<br>**Sample size (lesions):** No. included: 313<br>**Participant characteristics:** Mean age: 40y (10-86y); Male: 122; 41%<br>**Lesion characteristics:** Lesion site: Head/Neck: 3%; Trunk: 61%; Limbs: 36%; Thickness ≤1mm: 70% (46/66); for 55 invasive MM: median thickness 0.64mm, range 0.17-3.24mm. Median diameter: 11mm (3-31mm) |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | Yes |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm (Training in the unit is based on ABCD but subjective experience of the clinician used for diagnosis)<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Same clinician undertook clinical diagnosis and diagnosis using dermoscopy<br>**Diagnostic threshold:** Clinical diagnostic criteria based on subjective experience; emphasise lesion colour over dimensions. Diagnosis of suspect CM made when the level of suspicion was 'roughly 50% or more'. ABCD (asymmetry, border, colour, dimension) criteria have been the basis of training at the unit, but is not implemented in diagnosis; preferred emphasis on colour rather than dimensional character<br>**Diagnosis based on:** Single observer; (n=1)<br>**Observer qualifications:** Surgical oncologists<br>**Experience in practice:** High experience or 'Expert'; over 5 years<br>**Dermoscopy:** also evaluated in same study (no algorithm) |

## Visual Inspection - in-person

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | No |
| Was the test interpretation carried out by an experienced examiner? | Yes |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>**TARGET CONDITION (Final diagnoses)**<br>Melanoma (invasive): 55; Melanoma (in situ): 11; BCC: 6<br>'Benign' diagnoses: 241;151 compound naevus, 24 junctional naevus, 12 dermal naevus, 12 lentigo simplex, 10 dysplastic naevus, 8 spindle-cell naevus, 8 sebhorrheic keratosis, 5 blue naevus, 3 spitz naevus, 8 other. |
| 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 |
| **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? | 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 | Excluded participants: NR<br>Interval between index and reference: NR |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

## Notes

| | |
|---|---|
| **Notes** | - |

### *Bono 2002b*

Patient Selection

| **A. Risk of Bias** | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** Dec 2000 and Aug 2001<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| **B. Concerns regarding applicability** | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Consecutive cutaneous pigmented lesions that were <=6mm in diameter and required surgical biopsy for diagnosis based on clinical or dermoscopic suspicion of CMM<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Clinical and/or dermatoscopic suspicion<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** lesion size >6mm; non-pigmented<br>**Sample size (patients):** No. eligible: 349/ No. included: 157<br>**Sample size (lesions):** No. eligible: 375/ No. included: 161<br>**Participant characteristics:** Mean age 38y (14-82); Male: 61 (39%)<br>**Lesion characteristics:** Site: head/Neck: 14 (9%); trunk: 88 (55%); limbs: 59 (36%)<br>**Lesion size:** median: 5mm (1mm to 6mm) |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | Yes |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI).** No algorithm (ABCD (asymmetry, border, colour, dimension) criteria have been the basis of training at the unit, but is not implemented in diagnosis; preferred emphasis on colour rather than dimensional character)<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Other test data:** Dermoscopy evaluated in same study by same observer(s)<br>**Diagnostic threshold:** A diagnosis of suspect CM is made when the level of suspicion is roughly 50% or more; lesions at a lower index of suspicion were considered benign for the purposes of this study.<br>**Diagnosis based on:** Single observer diagnostic criteria based on the subjective experience of the single clinician examining the pigmented lesion (n=2)<br>**Observer qualifications:** Surgical oncologists<br>**Experience in practice:** High experience or 'Expert'; observers described as "expert in the recognition of pigmented lesions"<br>**Other detail:** Diagnostic criteria were based on the subjective experience of the single clinician examining the pigmented lesion, although the ABCD criteria have been the basis of training at the unit, they did not consider the ABCD mnemonic an essential formula for diagnosis of CM. They did not take into consideration the dimensional character and attributed great importance to the colour of a given lesion.<br>**Dermoscopy:** performed by the same two clinicians who firstly made and registered the clinical diagnosis |

## Visual Inspection - in-person

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | No |
| Was the test interpretation carried out by an experienced examiner? | Yes |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 13 CM; Disease negative: 148<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 10; Melanoma (in situ): 3; BCC: 2 (1.2%)<br>Mild/moderate dysplasia: 26 (16.1%); Sebhorrheic keratosis: 4 (2.5%); Benign naevus: compound nevus 57 (35.4%), junctional nevus 38 (23.6%), spindle-cell nevus 6 (3.7%), spitz nevus 5 (3.1%), blue nevus 2 (1.2%), other 6 (3.7%), Lentigo simplex 2 (1.2%) |
| 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 |
| **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? | 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 | Excluded participants: none reported<br>Time interval to reference test: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Bono 2006*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective<br>**Period of data collection:** Jan 2003 - Dec 2004<br>**Country:** Italy |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Consecutive patients with pigmented skin lesions with a maximum diameter of <=3mm undergoing excision. The decision for diagnostic excision was based on clinical and/or dermoscopic features suggesting a more or less important suspicion for CM<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic) Istituto NazionaleTumori of Milan<br>**Prior testing:** Clinical and/or dermatoscopic suspicion<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** - lesion size >3mm<br>**Sample size (patients):** No. eligible: 204/ No. included: 204<br>**Sample size (lesions):** No. eligible: 206/ No. included: 206<br>**Participant characteristics:** Median age: 40 (6-74); Male: 71 (35%)<br>**Lesion characteristics** Head/Neck: 8 (4%); Trunk: 84 (41%); Limbs: 114 (55%). Median size: 2mm (1 to 3mm) |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | Yes |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Other test data**: Dermoscopy evaluated in same study by same observer(s)<br>**Diagnostic threshold:** A diagnosis of suspicious CM is made when the level of suspicion is roughly 50% or more; lesions at a lower index of suspicion were considered not CM<br>**Diagnosis based on:** Single observer; n= 1<br>**Observer qualifications:** Not reported (assumed Oncologist as per Bono 2002 and Bono 2002b); "single clinician examining the pigmented lesion"<br>**Experience in practice:** Not described<br>**Experience with dermoscopy:** Not described<br>**Dermoscopy:** evaluated in same study; Menzies criteria<br>**Any other detail:** ABCD (asymmetry, border, colour, dimension) criteria have been the basis of training at the unit, but is not implemented in diagnosis; preferred emphasis on colour rather than dimensional character |

## Visual Inspection - in-person

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | No |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Details: The slides were evaluated according to widely accepted criteria for the histopathological diagnosis of the various pigmented lesions.<br>Disease positive: 23; Disease negative: 183<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 19 (9.2%); Melanoma (in situ): 4 (2.0%)<br>Mild/moderate dysplasia: dysplastic naevus 10 (4.9%); junctional naevus 76 (36.9%); compound naevus 50 (24.3%); dermal naevus 12 (5.8%); blue naevus 11 (5.3%); reed naevus 7 (3.4%); spitz naevus 3 (1.5%); halo naevus 3 (1.5%); lentigo simplex 7 (3.4%); other 4 (1.9%) |
| 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 |
| **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? | 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 | **Excluded participants:** none<br>**Time interval to reference test**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

## Notes

| Notes | - |
|---|---|

### *Bourne 2012*

## Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** June 1 - July 6 2009<br>**Country** Australia |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** All skin lesions consecutively excised at a skin cancer practice to exclude skin cancer and common lesions assessed as clearly benign and not biopsied were included<br>**Setting:** Primary<br>**Prior testing:** Clinical and/or dermatoscopic suspicion. Prior testing to assemble the test set occurs in secondary care by an experienced skin cancer doctor, then the images are tested on primary care professionals<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** Clinically obvious basal cell carcinomas which could be easily diagnosed without dermoscopy were not included in the collection set.<br>**Sample size (patients):** No. eligible: 46/ No. included: 46<br>**Sample size (lesions):** No. eligible: 50/No. included: 50<br>**Participant characteristics:** Mean age: 58 (30 to 60); Male: 22<br>**Lesion characteristics:** Face = 8; Neck = 1; Chest = 3; Back = 21; Shoulder = 2; Arm = 3; Thigh = 4; Leg = 7; Foot plantar = 1 |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | Yes |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** No further information used; Image assessments were done on four occasions, each time using a different diagnostic approach.<br>**Diagnostic threshold:** Not reported clinicians provided with Excel answer sheets for each method listing the various criteria used in that algorithm but no algorithm was cited for VI<br>**Diagnosis based on:** Average (n=4)<br>**Observer qualifications:** 3 GPs and 1 clinical nurse<br>**Experience in practice:** Mixed; described as varying levels of dermatoscopic experience<br>**Dermoscopy:** evaluated in same study; 3-point rule; Menzies criteria |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **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 plus other<br>Histopathological examination (n=46);<br>Expert diagnosis as benign (n=3); Digital follow up (n=1)<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 1; Melanoma (in situ): 7; BCC: 6; Lentigo maligna 1<br>Sebhorrheic keratosis: 5; 'Benign' diagnoses: Banal nevus 10, Blue naevus 1, Nevus and seborrhoeic keratosis/solar lentigo collison 3, Solar lentigo 4, LPLK 4, Dermatofibroma 1, Psoriasis 1, Solar keratosis 2, Intraepidermal carcinoma 3, Regressed keratoacanthoma 1 |
| 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 |
| **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 | No |
| 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?** | High |

## Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | Excluded participants: As two of the methods (Menzies and 3 point checklist) related to only pigmented lesions, the 5 non-pigmented specimens in the set of 50 were excluded from the contingency tables for these methods.<br>Time interval to reference test: "all skin lesions consecutively excised to exclude skin cancer were recorded" |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | No |
| 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? | Unclear |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | Yes |
| **Could the patient flow have introduced bias?** | High risk |

Notes

| Notes | - |
|---|---|

*Carli 2002*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective for clinical examination and in-vivo dermoscopy; Retrospective image selection / Prospective interpretation for ex-vivo dermoscopic evaluation<br>**Period of data collection:** June 1997 - December 1998<br>**Country:** Italy |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Clinically equivocal and suspicious pigmented skin lesions subjected to excisional biopsy at the Institute of Dermatology<br>**Setting:** Secondary (not further specified)<br>**Prior testing:** Clinical and/or dermatoscopic suspicion<br>**Setting for prior testing:** Secondary<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** NR<br>**Sample size (lesions):** 256<br>**Participant characteristics:** None reported<br>**Lesion characteristics** Of the cutaneous melanomas, 14 (25.9%) were in situ melanoma (Clark level I), 18 (33.3%) were invasive with less than 0.75 mm thickness, 19 (35.3%) were of intermediate thickness (0.76–1.50 mm) and three (5.5%) were thicker than 1.5 mm. The median thickness of invasive melanomas was 0.94 mm ± 0.5 (SD) (range 0.2–2.6). |
| 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Unclear<br>**Other test data:** clinical examination and in vivo dermoscopy were performed before excision by two trained dermatologists and diagnosis reached<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Consensus (2 observers); final clinical diagnosis was based on agreement between the two observers. In case of disagreement, the opinion of a third observer (B.G.) was considered to be the judge for the diagnosis<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** High experience or 'Expert'; described as "dermatologists with extensive experience in both clinical and dermoscopic diagnosis of pigmented skin lesions"<br>**Dermoscopy:** evaluated in same study; pattern analysis |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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 |
| Was the test interpretation carried out by an experienced examiner? | Yes |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>**Target condition (Final diagnoses)** Melanoma (invasive): 40; Melanoma (in situ): 14<br>BCC: 5<br>Sebhorrheic keratosis: 4; Benign naevus: 90 common melanocytic naevi; 78 melanocytic naevi; 9 blue naevi; 16 Spitz reed naevi |
| 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 |
| **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? | 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 | Excluded participants: none reported<br>Time interval to reference test: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Carli 2002b*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Not reported<br>**Period of data collection:** NR<br>**Country:** Italy |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Clinically suspicious or equivocal pigmented skin lesions undergoing excision for diagnostic purposes; only lesions with a diameter of 14 mm or less were included<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion<br>**Setting for prior testing:** Secondary (general dermatology)<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. included: NR<br>**Sample size (lesions):** No. included: 57<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** thickness ≤1mm: 11 cases (5 in situ 6 invasive); All <=14mm diameter |
| 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs; Fixed focus distance of 10cm; images observed using a viewer in two separate diagnostic sessions<br>**Prior test data:** No further information used; Contact (dermoscopic) images viewed first and then distant images (clinical), without knowing the classification of the contact image of the individual lesions.<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Consensus (2 observers); n=2<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** High experience or ‘Expert’; States 'with experience in the field of PSL'<br>**Other detail:** Used an AF micro Nikkor 60 lens objective mounted on a NIKON f50 camera, with a fixed focus distance of 10cm<br>**Dermoscopy:** evaluated in same study; no algorithm |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | 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** *Histology (not further described)*<br>Disease positive: 21; Disease negative: 36<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 6; Melanoma (in situ): 5; BCC: 10<br>'Benign' diagnoses: 36 |
| 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 |
| **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? | 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 | **Excluded participants:** No exclusions reported<br>**Time interval to reference test:** Photographic procedures performed consecutively prior to surgery |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | Yes |
| Were all patients included in the analysis? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Low risk |

### Notes

| Notes | - |
|---|---|

## *Carli 2003b*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection:** 1999-2001<br>**Country:** Italy |
| 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:** Clinically difficult to diagnose or equivocal melanocytic lesions randomly selected from image database; all melanomas less than 1mm thickness.<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion<br>**Setting for prior testing:** Secondary (general dermatology)<br>**Exclusion criteria:** >/=1mm thick melanomas, dermoscopically peculiar lesions (eg Blue nevi or Spitz nevi)<br>**Sample size (patients):** NR<br>**Sample size (lesions):** No. included: 200<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** Diameter <6mm 58, 6-10mm 87, >=10mm 55 (results reported per subgroup). Lesions ≤1mm thickness: 64; median thickness 0.3mm, 25th-75th centile 0.00-0.58mm; Mean diameter 7.4 (SD2.79) mm; Median: 7mm (2-16mm)<br>**Any other detail:** Same lesions appear to be reported in De Giorgi 2011 but with a different set of 8 observers (De Giorgi 2011 excluded from review on this basis) |
| 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** No further information used; dermoscopic images interpreted subsequent to clinical images<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Average; n= 8<br>**Observer qualifications:** Dermatology registrar; 2 final year residents. Dermatologist 6<br>**Experience in practice:** Mixed - 2 senior experts, 4 practicing dermatologists, 2 last year resident dermatologists. Classified as 'high' due to expertise/training in dermoscopy use<br>**Other detail:** Clinical photos using Nikon F40 with macro lens at 15cm.<br>**Dermoscopy:** evaluated in same study; no algorithm (own choice) |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | 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<br>Disease positive: 64; Disease negative: 136<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 40; Melanoma (in situ): 24<br>Other: 136 melanocytic nevi |
| 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 |
| **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 | **Excluded participants:** No exclusions reported<br>**Time interval to reference test:** Interval not described |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Chang 2013*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective<br>**Period of data collection:** Jan 2006 to Jul 2009<br>**Country:** Taiwan |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Potentially malignant biopsied or excised skin lesions (non-tumour specimens excluded)<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Secondary (general dermatology)<br>**Exclusion criteria:** prior surgery; image mis-registered or poor quality images (unfocused or containing a motion artefact) (considered under Flow and Timing)<br>**Sample size (patients):** No. eligible: 3964; No. included: 676<br>**Sample size (lesions):** No. eligible: 4192; No. included: 769<br>**Participant characteristics:** Mean age: 47.6 (SD 21.0); Male: 296; 43.8%<br>**Lesion characteristics:** None reported |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** Not reported; clinicians' impressions prior to biopsy were classified as ''benign'', ''malignant'', or ''indeterminate''. When the clinicians were not confident enough to make a definite benign or malignant diagnosis, the clinical impression was considered as ''indeterminate'' data extracted for malignant vs rest and malignant/indeterminate vs rest<br>**Diagnosis based on:** Single observer; board-certified staff dermatologists from institute; n= 25<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Board certified; High |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

### Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** *Histology (not further described)*<br>Disease positive: 174; Disease negative: 595<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 4; Melanoma (in situ): 4; BCC: 110; cSCC: 20<br>'Benign' diagnoses: 595 |
| 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 |
| **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? | 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 | **Excluded participants:** mis-registered or poor quality images (unfocused or containing a motion artifact) as a study inclusion criterion<br>**Time interval to reference test:** Not described |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

### Notes

| Notes | - |
|---|---|

## *Collas 1999*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** January 1996 and August 1997<br>**Country:** France |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented skin lesions undergoing excision by dermatologists in private practice, and by hospital dermatologists<br>**Setting:** Secondary (general dermatology); Private care<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Secondary (general dermatology); Private care<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. included: 353<br>**Sample size (lesions):** No. included: 353<br>**Participant characteristics:** Male: 46%; 162<br>**Lesion characteristics:** None reported |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | Yes |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm. Own new algorithm;<br>Diagnosis based on features from ABCD and 7-point checklist but neither one specifically followed.<br>Authors' select own combination of lesion characteristics based on observed data<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Unclear<br>**Diagnostic threshold:** Data can be extracted at a number of thresholds<br>1. Primary diagnosis of melanoma; 2. Certainty of melanoma diagnosis; 3. Various combinations of assessed features (based on logistic regression)<br>Recorded: Most likely clinical diagnosis; degree of melanoma suspicion and clinical sign(s) that led to the removal decision based on ABCD rule (McCarthy 1995) and the seven-point checklist (Healsmith 1994).<br>**Diagnosis based on:** Single observer; n= NR<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described<br>**Other detail:** Most predictive features derived by logistic regression from the following list: 1) irregular contours , 2) abnormal pigmentation, 3) blurred, 4) frank tumor appearance, 5) erosion, ulceration or bleeding, 6) regression signs. 7) lesion recently amended, 8) lesion appeared recentl , 9) pruritic lesion, 10) other |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | No |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Was the test applied and interpreted in a clinically applicable manner? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | Yes |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | Unclear |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** *Histology (not further described)*<br>Disease positive: 38; Disease negative: 315<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 38<br>Other: 160 |
| 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 |
| **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? | 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 | **Excluded participants:** No exclusions reported<br>**Time interval to reference test:** Consecutive; "When the dermatologist decided to resection a pigmented lesion, he fulfilled a pre-printed sheet" |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | Yes |
| Were all patients included in the analysis? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | Yes |
| **Could the patient flow have introduced bias?** | Low risk |

## Notes

| | |
|---|---|
| **Notes** | - |

### *Cristofolini 1994*

#### Patient Selection

| **A. Risk of Bias** | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection:** October 1990-June 1991<br>**Country:** Italy |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| **B. Concerns regarding applicability** | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Patients with pigmented lesions presenting during a campaign for the early diagnosis of cutaneous melanoma at the Dermatology Department in Trento<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** Lesions that were not taken into consideration included benign lesions, naevi of Unna and Miescher types and naevi that showed no inclusion criteria at the ABCDE clinical examination<br>**Sample size (patients):** No. eligible: 700 people; No. included: not reported<br>**Sample size (lesions):** No. eligible: 220; No. included: 220<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI)** ABCDE<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Other test data**: Dermoscopy undertaken by same clinician(s) subsequent to clinical evaluation<br>**Diagnostic threshold:** lesions showing at least two of the ABCDE criteria all of which were shown the same diagnostic importance, were considered positive.<br>**Diagnosis based on:** Unclear; n=4<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** High experience or 'Expert'; all trained in the recognition of pigmented lesions during a training course about the clinical diagnosis of naevi and melanomas; all working in a department where the early diagnosis of melanoma had been dealt with for over 10 years.<br>**Experience with dermoscopy:** High experience /'Expert' users<br>**Other detail:** ABCDE criteria are (asymmetry in shape, border irregular and notched, colour mottled-haphazard display, dimension >6mm, evolution changes in pigmentation)<br>**Dermoscopy:** evaluated in same study; Pattern analysis |

## Visual Inspection - in-person

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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? | Unclear |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | Yes |
| Was the test interpretation carried out by an experienced examiner? | Yes |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | Unclear |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 33<br>Mild//moderate dysplasia: 23 dysplastic naevi; Sebhorrheic keratosis: 4; Benign naevus: 158 common naevus<br>Other: 2 thrombosed angiomas |
| 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 |
| **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? | 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 | **Excluded participants:** No exclusions reported<br>**Time interval to reference test:** Not described<br>**Time interval between index tests:** clinical evaluation directly followed by dermoscopy |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

Notes

| Notes | - |
|---|---|

*Cristofolini 1997*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** Nov 1992 to Sept 1993<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Patients with small and flat common and atypical pigmented skin lesions recruited during a health campaign for the early diagnosis of CM underwent clinical diagnosis, computerised analysis by SVS and subsequent skin biopsy.<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** No prior testing<br>**Setting for prior testing:** Secondary (general dermatology)<br>**Exclusion criteria:** None reported<br>**Sample size (patients)**: 176<br>**Sample size (lesions)**: 176<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI)** ABCD<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Clinical examination and/or case notes<br>**Diagnostic threshold:** Not reported; examined individual ABCD characteristics but no 'rule' as to when to diagnose melanoma; appears to be subjective dx<br>**Diagnosis based on:** Consensus (3 observers) (n=3)<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Not described in paper but judged as 'High'; States that “All lesions were examined by three dermatologists according to the ABCD system, if they disagreed a fourth dermatologist an expert in the diagnosis of pigmented lesions was consulted.” Cristofolini 1994 describes four dermatologists "trained in the recognition of pigmented lesions", three of the four are in common with Cristofolini 1997. |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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 |
| Was the test interpretation carried out by an experienced examiner? | Yes |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>**TARGET CONDITION (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 35<br>Other: 141 melanocytic nevi |
| 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 |
| **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? | 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 | Excluded participants: not reported<br>Time interval to reference test: 'subsequent skin biopsy'<br>Time interval between index test(s): not reported-appears to be simultaneous |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | Yes |
| Were all patients included in the analysis? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Low risk |

Notes

| Notes | - |
|---|---|

### *de Giorgi 2012*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** between Oct 2006 and Sept 2010<br>**Country** Italy |
| 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:** Pigmented melanocytic skin lesions with a maximum diameter of 6mm excised at Dept Dermatology<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** Location/site of lesion - palmar and plantar regions, mucosal lesions and pigmented melanocytic lesions of the nails excluded<br>**Sample size (patients):** NR<br>**Sample size (lesions):** No. included: 103<br>**Participant characteristics:** Mean age: Melanoma group male (50.4yrs) female (48.4 yrs); Benign group male (36yrs) female (36.8yrs)<br>**Lesion characteristics:** Head/Neck: 3; Trunk: 21; Upper limbs/shoulder: 16; Lower limbs/hip: 26; back= 34; dorsal acral =3. Thickness: <= 1mm 15; >1mm= 1 MM |
| 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 | **Visual inspection (VI)** ABCD<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** Unclear<br>**Other test data**: Dermoscopic images also presented separately to observer (only presence/absence of particular dermoscopic features recorded; not an overall diagnostic assessment)<br>**Diagnostic threshold:** ABCD criteria =>2 criteria present<br>**Diagnosis based on:** Consensus (3 observers); n= 3<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** High experience or 'Expert'; "the four dermatologists had the same level of training and experience in dermatology, with more than 5 years of practice in dermoscopy" |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | No |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | High 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 |
| Was the test interpretation carried out by an experienced examiner? | 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<br>Disease positive: 34; Disease negative: 69<br>**Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 34<br>'Benign' diagnoses: 69 benign melanocytic nevus |
| 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 |
| **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 | **Excluded participants:** none reported;<br>**Time interval to reference test**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Dolianitis 2005*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case control<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** July 2001 to June 2002<br>**Country** Australia |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | No |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Dermoscopy training study using a CD with five test sets of images, each with 40 images of melanocytic skin lesions. Only good-quality macroscopic and dermoscopic images were included.<br>**Setting:** Training images; author inst Dept Dermatology, University of Melbourne<br>**Prior testing:** unclear<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** Nonmelanocytic lesions; poor quality index test image - only good-quality macroscopic and dermoscopic images were included, where the whole lesion was visible, including the entire periphery (considered under flow/timing)<br>**Sample size (patients):** Not reported<br>**Sample size (lesions):** No. eligible: 40; No. included: 40<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** ≤1mm thickness: 14 invasive melanomas; median 0.50 mm |
| 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 | **Visual inspection (VI):** No algorithm<br>**Method of diagnosis:** Clinical photographs alone<br>**Prior test data:** No further information used<br>**Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone.<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Average; 61 participants (invited to participate in a study comparing dermoscopic algorithms; advertised at several medical meetings and on a Web site for primary care physicians).<br>**Observer qualifications:** 10 dermatologists, 16 dermatology trainees, 35 GPs<br>**Experience in practice:** Mixed. Participant (volunteers) "had a range of experience levels with assessment of skin lesions [outlined in detail in the paper].. and a significant number were novices in dermoscopy". Paper reports 82% of participants responded that they assessed at least 2-4 PSL per week. Participants were given explanatory written material and CDs containing educational material on dermoscopy and test images.<br>**Dermoscopy:** evaluated in same study based on dermoscopic images alone; pattern analysis; 7-point checklist; ABCD; Menzies criteria |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **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 plus other (one lesion described as having no biopsy performed)<br>*Histology (not further described)* Disease positive: 20; Disease negative: 19<br>*Expert dx:* 1<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 18; Lentigo maligna 2<br>Benign naevus: 7 dysplactic nevi; 3 spitz nevi; 3 junctional nevi; 2 compound nevi; 4 other (ink-spot lentigo, blue nevus, solar lentigo, ephelis) |
| 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 |
| **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 | No |
| 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?** | High |

## Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported<br>**Time interval between index test(s)**: not reported |
| Was there an appropriate interval between index test and reference standard? | Unclear |
| Did all patients receive the same reference standard? | No |
| Were all patients included in the analysis? | Yes |
| 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? | Unclear |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

## Notes

| Notes | - |
|---|---|

*Dummer 1993*

## Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection:** 12 month period (year/dates NR)<br>**Country:** Germany |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Patients with skin lesions difficult to diagnose clinically<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) a type of specialist care- dermatology based clinic<br>**Exclusion criteria:** Patients who had excisions performed in individual practices or where there was no histology or cases that were so obvious they didn't need to have further investigation (clearly benign)<br>**Sample size (patients):** Not reported<br>**Sample size (lesions):** No. eligible: 824; No. included: 771<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection:** No algorithm<br>**Method of diagnosis:** In person<br>**Prior test data:** In person<br>**Other test data:** Dermoscopic images viewed separately<br>**Diagnostic threshold:** NR<br>**Diagnosis based on:** Single observer; (n=2 or 3)<br>**Observer qualifications:** Unclear; clinician based in Dermatology clinic<br>**Experience in practice:** Unclear<br>**Experience with index test:** Unclear |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

### Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 23MM; Disease negative: 748 benign<br>**Target condition (Final diagnoses)**<br>Invasive melanoma: 23<br>Benign naevus 706; Sebhorrheic keratosis 4; Benign non-melanocytic naevus 32 |
| 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 |
| **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? | 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 | **Excluded participants**: 53 non-melanocytic lesions not included in the final analysis (no melanomas present in this group)<br>**Time interval to reference test**: Not reported<br>**Time interval between index test(s)**: Not reported |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

### Notes

| Notes | - |
|---|---|

***Ek 2005***

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** January 2001 to December 2002<br>**Country** Australia |
| 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:** Lesions excised at tertiary referral centre for the management of cancers; only those lesions in which malignancy could not be excluded were included<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** Punch, shave or incisional biopsies and palliative excisions. Equivocal pathology report (n=56).<br>**Sample size (patients):** No. eligible: 1302; No. included: 1223<br>**Sample size (lesions):** No. eligible: 2678; No. included: 2582<br>**Participant characteristics:** Mean age: 73.6y (16–102y). Male: 784 (64.1%); History of melanoma/skin cancer (%) 224; 8.7% recurrent lesions<br>**Lesion characteristics:** Head/Neck: 61%; Trunk: 14.4%; Limbs: 24.6% |
| Are the included patients and chosen study setting appropriate? | Unclear |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** Not reported pre-operative diagnosis<br>**Diagnosis based on:** Unclear; Likely single (n= 5).<br>**Observer qualifications:** Three consultants, a plastic surgery trainee and a clinical assistant.<br>**Experience in practice:** Mixed (low and high experience combined); Plastic surgery trainee usually 1st year, on 6 month rotation; clinical assistant described as having “many years of experience”.<br>**Other detail:** Some results are presented for consultant, senior registrar and registrar but underlying patient numbers are not provided per observer to allow separate 2x2 estimation. The discussion does describe the “six MM misdiagnosed as benign … as .. assessed by non-consultants”. |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 23<br>BCC: 1214; cSCC: 517<br>'Benign' diagnoses: 188 (7.3%) SCC in situ (Bowen's disease), 330 (12.8%) solar keratoses, 63 (2.4%) seborrhoeic keratoses 247 (9.6%) were other benign lesions |
| 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 |
| **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? | 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 | **Excluded participants**: Lesions with incomplete or incorrectly entered proformas were excluded (n=40).<br>**Index to reference interval**: Consecutive; used pre-operative clinical diagnosis of lesions undergoing biopsy |
| Was there an appropriate interval between index test and reference standard? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

Notes

| Notes | |
|---|---|

*Gachon 2005*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series (dermatologists recruited and asked to use standardized questionnaire form whenever he or she decided to remove a nevus or MM for any reason, e.g., suspicion of MM, aesthetics,comfort, prevention).<br>**Data collection:** Prospective<br>**Period of data collection** NR<br>**Country** France |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Melanocytic skin lesions removed for any reason (e.g. suspicion of melanoma, aesthetics, comfort, prevention) by volunteer dermatologists.<br>**Setting:** Secondary (general dermatology) and private care; mostly "community dermatologists working in a private setting, and only 2 were academic dermatologists"<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion / Patient request for evaluation/excision; 1199 (29.7%) excised because they were considered suspicious by the dermatologist, and 869 (21.5%) because they were considered as precursors by the dermatologist; 1634 (40.7%) removed due to aesthetic or functional reasons, and 535 (13.3%) “only to reassure the patient"<br>**Setting for prior testing:** N/A<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** Not reported<br>**Sample size (lesions):** No. included: 4036<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** 36 (24.1%) of 149 melanoma were in situ or other invasive lesions with a median Breslow thickness of 0.60 mm |
| 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 | **Visual inspection (VI);** No algorithm Accuracy presented only for clinician's first clinical impression of lesions; after recording likelihood of melanoma, assessments were made as to the contributions of pattern recognition, ABCD criteria and ugly duckling (differential recognition)<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** 'considered suspicious' by dermatologist<br>**Diagnosis based on:** Single observer; (n= 135 of 200 volunteers)<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Not described; Most were community dermatologists working in a private setting, and 2 were academic dermatologists.<br>**Experience with index test:** Not described |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 149; Disease negative: 3887<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 149 (36 were *in situ* or other invasive lesions with a median Breslow thickness of 0.60 mm)<br>'Benign' diagnoses: 3629 nevi (89.9%); 4 uncertain MMs/nevi (0.1%); and 254 non-melanocytic lesions clinically considered to be nevi or MMs (6.3%). |
| 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 |
| **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? | 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 | **Excluded participants: None reported**<br>**Time interval to reference test:** NR |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

Notes

| Notes | - |
|---|---|

### *Green 1991*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** Feb 1989 - Aug 1990<br>**Country** Australia |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented lesions with complete clinical and histological data<br>**Setting:** Secondary (referred from surgery, dermatology, casualty)<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Depts surgery, dermatology, casualty<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 81 / No. included: unclear<br>**Sample size (lesions):** No. eligible: 89; No. included: 70<br>**Participant characteristics:** Median age 32 yrs; Male 36 (44%)<br>**Lesion characteristics:** site Trunk: 80%; - Limbs: 10%; - face and neck 10% |
| 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 | **Visual inspection (VI);** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** In person<br>**Diagnostic threshold:** Not reported clinical diagnosis recorded plus assessment of diameter, colour, regularity of outline, diffuseness of edge and palpability<br>**Diagnosis based on:** Single observer; (n= NR)<br>**Observer qualifications:** Mixed; "in the majority of cases a surgeon or a dermatologist"<br>**Experience in practice:** Not described |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard;** Histological diagnosis and Expert diagnosis<br>*Histology:* 62/70 lesions<br>Expert diagnosis: 8/70 lesions; Eight lesions had clinical diagnoses assigned (all benign) in the absence of available histology reports<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 5<br>BCC: 2; Sebhorrheic keratosis: 7; Benign naevus: 53 Other: 2 skin tags, 1 'lentigo' |
| Is the reference standards likely to correctly classify the target condition? | No |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Unclear |
| **Could the reference standard, its conduct, or its interpretation have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Expert opinion (with no histological confirmation) was not used as a reference standard | No |
| 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?** | High |

## Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | **Excluded participants:** 19/89 lesions excludedue to incomplete clinical and histology records.<br>**Time interval to reference test:** Assumed consecutive; pathology referral form used to ascertain clinical diagnosis |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | No |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

Notes

| Notes | - |
|---|---|

### *Green 1994*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Not reported; appears to use previously acquired images to develop a new CAD classifier (not included as derivation), and compare results to clinical Dx of clinicians as recorded in notes. Unclear whether set up prospectively or was retrospective assessment.<br>**Period of data collection** August 1990 to April 1992<br>**Country** Australia |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented lesions for excision<br>**Setting:** Secondary (Dept Surgery)<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. included: 129<br>**Sample size (lesions):** No. eligible: 204; No. included: 164<br>**Participant characteristics:** Mean age 36y, range 6 to 87y; Male: 42.6%<br>**Lesion characteristics:** site - Face/Neck: 10% Trunk: 66%- Limbs: 24% |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI):** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** No further information used<br>**Diagnostic threshold:** Not reported; clinical diagnosis recorded plus assessment of diameter, colour, regularity of outline, diffuseness of edge and palpability (same as for Green 1991)<br>**Diagnosis based on:** Single observer; (n= NR)<br>**Observer qualifications:** Not reported<br>**Experience in practice:** Not described |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** *Histology (not further described)*<br>Disease positive: 18; Disease negative: 146<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 18; Melanoma (in situ): 3<br>128 melanocytic nevi; 15 miscellaneous pigmented lesions including seborrhoeic keratoses, basal cell carcinomas, and lentigines |
| 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 |
| **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? | 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 | **Excluded participants**: 33 lesions excluded due to problems using the images with the CAD software, e.g. lesion 'too big'; image 'obscured by hairs or surgeons pen marks' or 'software was unable to contend with the lesion characteristics, mainly because the lesion was too light or too fragmented' or 'avoidable operator error'<br>**Time interval to reference test**: NR |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

### Notes

| Notes | - |
|---|---|

## *Grimaldi 2009*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** Oct 2005 - Mar 2006<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Cutaneous pigmented lesions with digital images forwarded by primary care physicians to a referral centre for confirmation of diagnosis.<br>**Setting:** Primary; Lesions selected for referral by GPs; accuracy of GP diagnosis assessed<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** Lesions whose removal had been explicitly demanded by the patients for aesthetic reasons, as well as those irritated or subjected to trauma<br>**Sample size (patients):** No. included: 197<br>**Sample size (lesions):** No. included: 235<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None reported |
| Are the included patients and chosen study setting appropriate? | Yes |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI);** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Other test data:** "two-step judgment (before and after dermoscopy) formulated by the sending physician, who labelled each lesion as ‘benign’ or ‘suspicious for malignancy’."<br>**Diagnostic threshold:** Not reported "Each physician was asked to formulate a written first judgment of every lesion before digital acquisition and to re-evaluate it after dermoscopy"<br>**Diagnosis based on:** Single observer; (n= 13)<br>**Observer qualifications:** GP; From approximately 250 primary care clinicians attending a conference, 13 volunteered to participate<br>**Experience in practice:** Not clearly described; assumed to be Low experience with pigmented lesions<br>**Experience in dermoscopy:** Unclear; classified as 'trained' - “simple protocols for diagnosis were made up and given to the participants via e-learning courses, direct meetings, and involving self assessment procedures”<br>**Dermoscopy:** evaluated in same study; no algorithm (ABCD used for telediagnosis at reference centre) |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis plus follow up (Reference is expert diagnosis for Teledermatology component of study)<br>*Histology (not further described): n=16;* Disease positive: 5; Disease negative: 11<br>*Clinical FU (6 months) plus histology of suspicious lesions: n*=219; Disease positive: 0; Disease negative: 208<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 5<br>Other: 230 benign |
| Is the reference standards likely to correctly classify the target condition? | No |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Unclear |
| **Could the reference standard, its conduct, or its interpretation have introduced bias?** | High 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 | **Excluded participants**: NR<br>**Time interval to reference test**: NR<br>**Time interval between index test(s)**: NR |
| Was there an appropriate interval between index test and reference standard? | Unclear |
| Did all patients receive the same reference standard? | No |
| Were all patients included in the analysis? | Yes |
| 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? | Yes |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

### Notes

| Notes | - |
|---|---|

## *Kopf 1975*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective<br>**Period of data collection:** 1955 to 1967<br>**Country:** US |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** All lesions subject to biopsy at the Oncology Section of the Skin and Cancer Unit.<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. included: NR<br>**Sample size (lesions):** No. included: 5538<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI);** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Unclear<br>**Diagnostic threshold:** Not reported; clinical diagnosis<br>**Diagnosis based on:** Single observer; in clinic diagnosis (n= NR)<br>**Observer qualifications:** Oncologist<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described |

### Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

### Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 99; Disease negative: 5439<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 99 (described as 'malignant melanoma')<br>Diagnoses listed only for false positives; included: 3 pigmented BCC, 3 dermatofibromas, 2 junction nevi, 2 compound nevi, and 1 each of: Kaposi sarcoma, hemangioma, seborrheic keratosis, leiomyoma, cellular blue nevus, sclerosing hemangioma, SCC, verrucous nevus, and intradermal nevus FNs included: 6 clinically diagnosed as pigmented BCC; 2 "other forms" of BCC; 3 junction nevi; 3 pyogenic granulomas; 2 compound nevi; 2 squamous cell carcinomas; 2 halo nevi; 1 Bowen disease; 1 seborrheic keratosis; and 1 lentigo. Seventeen of these lesions were pigmented and six were not. |
| 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 |
| **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? | 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 | **Excluded participants:** None reported<br>**Time interval to reference test:** Not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Krahn 1998*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection:** NR<br>**Country:** Germany |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Excised pigmented skin lesions<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. included: 80<br>**Sample size (lesions):** No. included: 80<br>**Participant characteristics:** None reported<br>**Lesion characteristics** range in thickness (melanomas) 0.18-1.9mm; 29/39 <0.76mm; 7/39 0.76-1.5mm; 3/39 >1.5mm |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | Yes |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm reported<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Unclear<br>**Other test data**: Dermoscopy undertaken by same clinician(s) subsequent to clinical evaluation<br>**Diagnostic threshold:** Not reported; no details<br>**Diagnosis based on:** Single observer (n=1)<br>**Observer qualifications:** Not reported likely Dermatologist<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described<br>**Dermoscopy:** evaluated in same study; no algorithm |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone including histometrics<br>Disease positive: 39; Disease negative: 41<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 39 (SSM, lentigo MM, nodular M)<br>Benign naevus: 37 common nevus; 3 dysplastic nevus, 1 Spitz nevus |
| 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 |
| **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 | **Excluded participants**: none reported<br>**Time interval to reference test:** not reported<br>**Time interval between index test(s)**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Langley 2001*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection:** Not reported<br>**Country:** USA |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Patients with lesions scheduled for excision at the pigmented lesion clinic to either remove atypical nevi or to rule out melanoma or for cosmetic reasons<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Selected for excision; to remove atypical nevi or rule out melanoma or for cosmetic reasons<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. included: 29<br>**Sample size (lesions):** No. eligible: 40; No. included: 38<br>**Participant characteristics:** Mean age 39 yrs, range 19 to 95 years; Male: 14 (48%)<br>**Lesion characteristics:** None reported |
| Are the included patients and chosen study setting appropriate? | Unclear |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** Not reported NR; clinical diagnosis<br>**Diagnosis based on:** Unclear likely in clinic diagnoses (n= NR)<br>**Observer qualifications:** Not reported likely dermatologists<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** - Histological diagnosis plus other<br>*Histology details (n=38):* "After excision, the samples were processed in paraffin and stained with H&E for routine light microscopy. Correlation was performed by examining the confocal images and the pathology sections to compare nuclear, cellular, and morphologic detail and to identify potential significance of the in vivo CSLM observations. For the histologic diagnosis of dysplastic nevi, we used the criteria that are defined in the World HealthOrganization consensus study."<br>*Expert diagnosis (n=2):* Two lesions did not undergo histology; expert diagnosis only (both benign)<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 3; Melanoma (in situ): 1; Lentigo maligna 2<br>Dysplastic nevi: 17; Benign naevus: 15 |
| 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 |
| **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 | No |
| 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?** | High |

Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported<br>**Time interval between index test(s)**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

Notes

| Notes | - |
|---|---|

### *Lorentzen 1999*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection:** Between 1994 and 1997<br>**Country:** Denmark |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Patients with lesions suspicious for CMM referred to outpatients clinic<br>**Excluded participants**: none reported<br>**Time interval to reference test**: not reported<br>**Setting:** Not reported<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** Poor quality index test image (considered under flow/timing)<br>**Sample size (patients):** No. eligible: 242; No. included: 232<br>**Sample size (lesions):** No. eligible: 242; No. included: 232*<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None reported<br>***NB** Not all cases were assessed by all observers; 2x2 are based on presented sensitivity and specificity estimates for full dataset of lesions; "the dermatoscopy experts assessed almost all cases (98 ± 100%), whereas the non-expert group completed fewer assessments, from 76 to 98%. |
| 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** No further information used; no option to change clinical diagnosis after viewing dermoscopic image<br>**Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone; clinical images presented before dermoscopic images<br>**Diagnostic threshold:** Not reported; clinical diagnosis<br>**Diagnosis based on:** Average; n= 9<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** High; Moderate; Mixed (average reported); 4 'experienced dermatologists' (4-5 years daily experience) & 5 'non-expert dermatology residents' (1-2 years interest and formal training in dermatoscopy]<br>**Experience with index test:** High; Moderate; Mixed |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | 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<br>Disease positive: 65 ; Disease negative: 167<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 49 'malignant melanoma'<br>BCC: 16<br>Sebhorrheic keratosis: 12; Benign naevus: 137 (pigmented nevi=116; blue nevi=16; atypical nevi=5); Other: 18 (spitz nevi, Bowen's disease, sarcoid, nevus spilus, hemangioma, and others) |
| 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 |
| **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? | 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 | **Excluded participants**: 10 cases were "considered unfit for evaluation" due to poor quality image<br>**Reference interval**: "biopsy specimens...were obtained after the clinical and dermatoscopic photographs had been performed" |
| Was there an appropriate interval between index test and reference standard? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

## Notes

| | |
|---|---|
| **Notes** | - |

### *McGovern 1992*

#### Patient Selection

| **A. Risk of Bias** | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection:** Between 1 Nov 1989 and 31 Oct 1990<br>**Country:** USA |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| **B. Concerns regarding applicability** | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented lesions excised to rule out dysplasia, lentigo maligna or malignant melanoma<br>**Setting:** Secondary (general dermatology); Army Dermatology clinic - appears to be open access<br>**Prior testing:** No prior testing. Multiple reasons given for seeking dermatological consultation, including (in descending order): increasing size, "mole check", inflammation, color change, itch, follow-up, variegation, cosmetic, referral, irregular border, seen for other lesion, unknown, large size<br>**Setting for prior testing**: N/A<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 179; No. included: Not reported<br>**Sample size (lesions):** No. eligible: 237; No. included: 13 lesions excluded and 32 lesions unaccounted for.<br>**Participant characteristics:** Mean age: 44 (SD 18); Range: 3 months to 86 years; Male: 89 (49%)<br>**Lesion characteristics:** lesion site: head/neck: 71 (30%); trunk: 52 (23%); upper limbs/shoulder: 22 (9%); lower limbs/hip: 33 (14%); back = 58 (24%); genitalia = 1 (0.4%) |
| Are the included patients and chosen study setting appropriate? | Unclear |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

#### Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** ABCD; Assessed only 'BCD'; also referred to in paper as 3 point checklist; Glasgow/MacKie original seven-point checklist (Keefe 1990)<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Unclear<br>**Diagnostic threshold:** Described in detail; ABCD excluded Asymmetry' -one half does not match the other half)<br>**Diagnosis based on:** Single observer in clinic diagnoses used (n=NR)<br>**Observer qualifications:** Not reported likely dermatologists<br>**Experience in practice:** Not described<br>Any other details: Border irregularity- edges are ragged, notched, or blurred; Color irregularity- pigmentation is not uniform; shades of tan, brown and black are present with dashes of red, white, or blue; Diameter- greater than 6 mm, the size of a pencil eraser<br>7-point: Increasing size, Variegation, Inflammation, Irregular outline,Greater than 1cm diameter, Itch, Bleeding One point awarded for each feature |

## Visual Inspection - in-person

| 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? | No |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | Unclear |
| **Could the conduct or interpretation of the index test have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Was the test applied and interpreted in a clinically applicable manner? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | Yes |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | Unclear |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Details: Shave excision=109; punch biopsy = 64; excision=47; snip biopsy=17<br>Disease positive: 16 lesions; Disease negative: 221<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 6; Lentigo maligna 6; BCC: 4;<br>Dysplastic naevus 28; Sebhorrheic keratosis: 32; Benign naevus: 110; Lentigo 12; blue naevus 9; actinic keratosis 6; dermatofibroma 6; atypical naevus 4; other 14 |
| 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 |
| **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 | **Excluded participants**: Missing data for the different algorithms; approximately 32 lesions unaccounted for (13 excluded due to lesion size of 8mm or less). ABCD evaluated=192/224 lesions; 3 point evaluated =192/224 lesions; 7 point evaluated =205/224 lesions<br>**Time interval to reference test**: not reported |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | Unclear |
| **Could the patient flow have introduced bias?** | High risk |

Notes

| Notes | - |
|---|---|

### *Menzies 2009*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** December 2005 to August 2006<br>**Country** Australia |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented lesions which, after routine naked eye examination by the GP, would have been biopsied or referred, i.e. a SPL (suspicious pigmented lesion). GPs were recruited from practices with at least 3 clinicians; excluded if they already used dermoscopy or SDDI in their routine practice.<br>**Setting:** Primary<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion<br>**Setting for prior testing:** Primary<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** NR<br>**Sample size (lesions):** No. included: 374<br>**Participant characteristics:** None reported<br>**Lesion characteristics** None reported |
| Are the included patients and chosen study setting appropriate? | Yes |
| 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?** | Unclear |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Other test data:** clinical diagnosis and placed in a sealed envelope before proceeding to dermoscopy examination<br>**Diagnostic threshold:** Not reported; initial diagnosis recorded along with confidence of diagnosis (scale 1 to 10; 1 not at all confident and 10 extremely confident), certainty of melanoma (scale 0 to 100%; 0 definitely not melanoma and 100 definitely melanoma) and management (biopsy, referral).<br>**Diagnosis based on:** Single observer (n=63; 102 GPs initially recruited; 74 (72.5%) completed the educational intervention and online assessment; 63 GPs from 19 practices finally participated)<br>**Observer qualifications:** GP<br>**Experience in practice:** Not fully described; assumed to be low experience with pigmented lesions. GPs must have each excised or referred >=10 PSL in previous 12-month period; excluded if dermoscopy or SDDI already used in routine practice. During the pretrial period all GPs underwent a training programme in the use of dermoscopy.<br>**Dermoscopy:** evaluated in same study; no algorithm |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis plus other<br><br>*Histology (not further described):* described as to standard practice and not necessarily blinded to the GP's diagnosis; author confirmed that all melanoma had histological diagnosis and >50% of benign had histology or follow-up<br><br>Total excised or referred: 163. Immediate excision/referral: 110. Excision/referral after SDDI: 48. Excision/examination after patient self referral 5<br>Disease positive: 37; Disease negative: total of 126 benign or unknown were 'excised OR referred' so some would have had specialist dx only.<br><br>*Clinical FU plus histology of suspicious lesions:* Short term digital monitoring (SDDI) available as an option for lesions considered not to be melanoma but that were still considered suspicious; follow-up imaging occurred initially at 3 months with any morphological changes to result in biopsy or referral; some lesions continued SDDI for a further 3 months; Length of FU: 3-6 months<br>No. patients: Initially recommended for SDDI: 192; SDDI continued for further 3 months: 6; Underwent SDDI only (no excision): 146<br>Disease positive: 15 (SDDI then histologically confirmed); Disease negative: 176 benign (incl 1 missed *in situ* melanoma); 4 unknown<br><br>*Expert opinion:* GPs could refer for specialist opinion or lesions could undergo dermoscopy telemedicine (images reviewed by an expert in dermoscopy and SDDI). Dermoscopy telemedicine was blinded to the GP's diagnosis. Observe for change group, i.e. discharged after dermoscopy: 72 Plus a proportion of those in Excise/refer group will have had expert dx alone but details not given<br>Disease positive: 0; Disease negative: 71 benign; 1 unknown<br><br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 33; Melanoma (in situ): 1<br>BCC: 6<br><br>2 Bowen's disease; 323 benign; 9 unknown |
| Is the reference standards likely to correctly classify the target condition? | No |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | No |
| **Could the reference standard, its conduct, or its interpretation have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Expert opinion (with no histological confirmation) was not used as a reference standard | No |
| 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?** | High |

## Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | **Excluded participants**: nine lesions with unknown diagnoses, plus BCC and Bowen's excluded from some analyses<br>**Time interval to reference test**: Not reported; Histopathological and specialist examination occurred according to standard practice |
| Was there an appropriate interval between index test and reference standard? | Unclear |
| Did all patients receive the same reference standard? | No |
| 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? | Yes |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

Notes

| Notes | - |
|---|---|

### *Morales Callaghan 2008*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** 1 January 2005 - 31 December 2005<br>**Country** Spain |
| 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:** Randomly selected melanocytic lesions; melanocytic on both clinical and dermoscopic criteria<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Dermatoscopic suspicion in all cases<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** Location/site of lesion - Palms, soles, mucous membranes of face, under nails; non-melanocytic appearance<br>**Sample size (patients):** No. included: 166<br>**Sample size (lesions):** No. included: 200<br>**Participant characteristics:** Mean age 33.7y (SD 14.5), range 8 to 84yrs; Male gender: 64 (38.6%); Fitzpatrick phototype II (44%); type III (41.5%)<br>**Lesion characteristics:** Macular component=181 (90.5%), Papular component=125 (62.5%) Both = 106 (53%), either one or other = 94 (47%). Asymmetrical 144 (72%). Irregular borders 154 (77%). 4 colours in 40 (20%), 3 colours in 96 (48%), 2 colours in 57 (28.5%), 1 colour in 1 (0.5%). History of bleeding 7 (3.5%). changes reported by patient 154 (77%). Lesion site: trunk 155 (77.5%), including the back in 106 (53%). Lesion size: mean long axis diameter 7.9mm (SD 8.6)mm, mean short axis diameter 5.1 (SD 5). |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Clinical examination and/or case notes<br>**Other test data**: Appears that dermoscopy was undertaken by same clinician(s) subsequent to clinical evaluation; clinical history was constructed following a standardised protoco and a presumptive clinical diagnosis recorded. Each lesion was then photographed and immediately afterwards examined using a manual<br>dermatoscope<br>**Diagnostic threshold:** Not reported; presumptive clinical diagnosis<br>**Diagnosis based on:** Consensus (n=2)<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Not clearly described; assumed to be High - “both dermatologists had experience in dermoscopy.”<br>**Dermoscopy:** evaluated in same study; Pattern analysis |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | Yes |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

### Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Details: Lesions described using terminology proposed by US National Insts of Health<br>Disease positive: 6/6 lesions; Disease negative: 194/194 lesions (assuming the 9 'Other' diagnosis lesions were not malignant), or 185/185 (removing the 9 'other' diagnosis lesions from dataset)<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 6 (3%)<br>Other: Atypical mole (104), Common mole (70), congenital nevus (6), Blue nevus (3), Spitz/Reed nevus (1), Spilus nevus (1), Others [unclear whether benign or malignant] (9) |
| 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 |
| **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? | 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 | **Exclusions**: none reported<br>**Time interval to reference test**: "Samples for histologic analysis were taken immediately after clinical and dermoscopic examination"<br>**Time interval between index test(s)**: Images taken at same time |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | Yes |
| Were all patients included in the analysis? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Low risk |

### Notes

| | |
|---|---|
| **Notes** | - |

## *Morton 1998a*

### Patient Selection

| **A. Risk of Bias** | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective<br>**Period of data collection** 1992 - 1994<br>**Country** Scotland |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| **B. Concerns regarding applicability** | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** All biopsies generated at Pigmented lesion clinic during time period.<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Not reported<br>**Setting for prior testing:** N/A<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 1999<br>**Sample size (lesions):** 763 lesions examined by one of two consultants<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis referred to as 'clinical diagnosis'; no dermoscopy used<br>**Diagnostic threshold:** Not reported NR; clinical diagnosis<br>**Diagnosis based on:** Single observer and average data presented; (n= 10 in total)<br>**Observer qualifications:** 2 consultant dermatologists,<br>**Experience in practice:** High (2 consultants each with >10 years experience in dermatology)<br>**Any other detail:** Data from same study for senior registrar and registrar presented in Morton 1998b and Morton 1998c |

### Visual Inspection - in-person

| **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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

### Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** - Histological diagnosis alone<br>**Target condition (Final diagnoses) (For full sample of 1999 biopsies)**<br>Melanoma (invasive): 102 (82 SSM, 11 nodular melanoma, 4 partially regressed, 2 acral lentiginous, 2 metastatic CM deposits, 1 desmoplastic melanoma); Melanoma (in situ): 24; Lentigo maligna: 2<br>Benign: 1871 benign (breakdown by lesion type not reported) |
| 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 |
| **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 | Excluded participants: None reported<br>Time interval to reference test: Not reported<br>Time interval between index test(s): 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Morton 1998b*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective<br>**Period of data collection** 1992 - 1994<br>**Country** Scotland |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** All biopsies generated at Pigmented lesion clinic during time period.<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Not reported<br>**Setting for prior testing:** N/A<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 1999<br>**Sample size (lesions):** 567 lesions examined by senior registrar<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None reported |
| Are the included patients and chosen study setting appropriate? | |
| Did the study avoid including participants with multiple lesions? | |
| **Are there concerns that the included patients and setting do not match the review question?** | |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis referred to as 'clinical diagnosis'; no dermoscopy used<br>**Diagnostic threshold:** Not reported NR; clinical diagnosis<br>**Diagnosis based on:** Single observer and average data presented; (n= 10 in total)<br>**Observer qualifications:** 2 senior registrars<br>**Experience in practice:** Moderate – 2 two senior registrars each with 3-5 years experience<br>**Any other detail:** Data from same study for consultants and for registrar presented in Morton 1998a and Morton 1998c |

## Visual Inspection - in-person

| A. Risk of Bias | |
|---|---|
| Were the index test results interpreted without knowledge of the results of the reference standard? | |
| If a threshold was used, was it pre-specified? | |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | |

| B. Concerns regarding applicability | |
|---|---|
| Was the test applied and interpreted in a clinically applicable manner? | |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | |
| Was the test interpretation carried out by an experienced examiner? | |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | |

## Visual inspection - image-based

| A. Risk of Bias | |
|---|---|
| Were the index test results interpreted without knowledge of the results of the reference standard? | |
| If a threshold was used, was it pre-specified? | |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | |

| B. Concerns regarding applicability | |
|---|---|
| Was the test applied and interpreted in a clinically applicable manner? | |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | |
| Was the test interpretation carried out by an experienced examiner? | |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | |

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** - Histological diagnosis alone<br>**Target condition (Final diagnoses) (For full sample of 1999 biopsies)**<br>Melanoma (invasive): 102 (82 SSM, 11 nodular melanoma, 4 partially regressed, 2 acral lentiginous, 2 metastatic CM deposits, 1 desmoplastic melanoma); Melanoma (in situ): 24; Lentigo maligna: 2<br>Benign: 1871 benign (breakdown by lesion type not reported) |
| Is the reference standards likely to correctly classify the target condition? | |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | |
| **Could the reference standard, its conduct, or its interpretation have introduced bias?** | |

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

## Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | Excluded participants: None reported<br>Time interval to reference test: Not reported<br>Time interval between index test(s): N/A |
| Was there an appropriate interval between index test and reference standard? | |
| Did all patients receive the same reference standard? | |
| Were all patients included in the analysis? | |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | |

## Notes

| Notes | - |
|---|---|

### *Morton 1998c*

## Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective<br>**Period of data collection** 1992 - 1994<br>**Country** Scotland |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** All biopsies generated at Pigmented lesion clinic during time period.<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Not reported<br>**Setting for prior testing:** N/A<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 1999<br>**Sample size (lesions):** 669 lesions examined by registrar<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None reported |
| Are the included patients and chosen study setting appropriate? | |
| Did the study avoid including participants with multiple lesions? | |
| **Are there concerns that the included patients and setting do not match the review question?** | |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis referred to as 'clinical diagnosis'; no dermoscopy used<br>**Diagnostic threshold:** Not reported NR; clinical diagnosis<br>**Diagnosis based on:** Single observer and average data presented; (n= 10 in total)<br>**Observer qualifications:** Registrars<br>**Experience in practice:** Low – 6 rotating registrars each with 1-2 years experience<br>**Any other detail:** Data from same study for consultants and for senior registrars presented in Morton 1998a and Morton 1998b |

## Visual Inspection - in-person

| A. Risk of Bias | |
|---|---|
| Were the index test results interpreted without knowledge of the results of the reference standard? | |
| If a threshold was used, was it pre-specified? | |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | |

| B. Concerns regarding applicability | |
|---|---|
| Was the test applied and interpreted in a clinically applicable manner? | |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | |
| Was the test interpretation carried out by an experienced examiner? | |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | |

## Visual inspection - image-based

| A. Risk of Bias | |
|---|---|
| Were the index test results interpreted without knowledge of the results of the reference standard? | |
| If a threshold was used, was it pre-specified? | |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | |

| B. Concerns regarding applicability | |
|---|---|
| Was the test applied and interpreted in a clinically applicable manner? | |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | |
| Was the test interpretation carried out by an experienced examiner? | |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | |

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** - Histological diagnosis alone<br>**Target condition (Final diagnoses) (For full sample of 1999 biopsies)**<br>Melanoma (invasive): 102 (82 SSM, 11 nodular melanoma, 4 partially regressed, 2 acral lentiginous, 2 metastatic CM deposits, 1 desmoplastic melanoma); Melanoma (in situ): 24; Lentigo maligna: 2<br>Benign: 1871 benign (breakdown by lesion type not reported) |
| Is the reference standards likely to correctly classify the target condition? | |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | |
| **Could the reference standard, its conduct, or its interpretation have introduced bias?** | |

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

## Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | Excluded participants: None reported<br>Time interval to reference test: Not reported<br>Time interval between index test(s): N/A |
| Was there an appropriate interval between index test and reference standard? | |
| Did all patients receive the same reference standard? | |
| Were all patients included in the analysis? | |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | |

## Notes

| Notes | - |
|---|---|

### *Pizzichetta 2004*

## Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** Jan 1996 to Dec 2001<br>**Country** Participants recruited from 5 participating centres (4 in Italy and 1 in USA) study conducted in Italy |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Unclear |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Clinical and/or dermoscopic hypomelanotic (extent of pigmentation <=30%) and amelanotic skin lesions seen and excised at the five participating centres<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Clinical and/or dermatoscopic suspicion<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** Poor quality or unavailable index test image (considered under Flow and TIming)<br>**Sample size (patients):** No. included: 151<br>**Sample size (lesions):** No. eligible: 174; No. included: 151<br>**Participant characteristics:** mean age 47 years (± 17.5 SD); male gender: 73 (48%)<br>**Lesion characteristics:** Lesion site - head/neck (5.3%); trunk (20.5%); upper limbs/shoulder (11.9$); lower limbs/hip (25.2%); back (21.2%); abdomen (11.3%); hand (3.3%); foot (1.3%). Melanoma thickness: ≤1mm 85.3% (n=29); >1mm 14.7% (n=15) |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | Yes |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** only the gender, age at diagnosis and the site of the skin lesion were known to the observer<br>**Other test data**: File contained clinical and dermoscopic images; unclear whether both observed at the same time.<br>**Diagnostic threshold:** investigated clinical features such as elevation, ulceration, shape, borders, colour<br>**Diagnosis based on:** Single observer (n=1)<br>**Observer qualifications:** Not reported likely dermatologist<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described<br>**Dermoscopy:** evaluated in same study; Pattern analysis |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **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<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 34 (39 in full sample); Melanoma (in situ): 5;<br>Other diagnoses reported only for full sample of 151 (only 108 with clinical images for VI evaluation):<br>55 (40 with clinical images) "amelanotic/ hypomelanotic non melanocytic lesions" (25 BCC, 4 SCC, 10 dermatofibroma, 8 Bowen's disease, 8 seborrhoeic keratosis)<br>52 (29 with clinical images) "amelanotic/ hypomelanotic benign melanocytic lesions" (24 compound naevi, 17 dermal naevi, 5 Spitz naevi, 4 congenital naevi and 2 combined naevi). |
| 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 |
| **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? | 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 | **Excluded participants:** 23 lesions excluded due to image quality; further 43 lesions were not available for evaluation by clinical images ("mainly benign melanocytic lesions").<br>**Time interval to reference test:** Not reported<br>**Time interval between index test(s):** not reported |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

Notes

| Notes | - |
|---|---|

### *Rao 1997*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** not reported<br>**Country** USA |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Patients with atypical melanocytic lesions or suspected early malignant melanoma<br>**Setting:** Private care<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Private care<br>**Exclusion criteria:** lesions over 13mm in diameter were excluded as they could not fit entirely within the standardized photographs<br>**Sample size (patients):** No. included: 63<br>**Sample size (lesions):** No. included: 72<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** Melaoma thickness - ≤1mm: 100% of MM (n=21) |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

Index Test

| | |
|---|---|
| Index tests | **Visual inspection** ABCD<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** Unclear<br>**Other test data**: Dermoscopic images also presented to observer but unclear whether both viewed at the same time or not; "Each color transparency was independently analyzed" by observers. The 1) clinical, 2) ”overall” dermoscopic, and 3) ABCD ”scored dermoscopic diagnoses of either MM or AMN were recorded for each lesion by the same observers. No indication of blinding between images<br>**Diagnostic threshold:** Clinical variables were defined as follows: Asymmetry (A): Both silhouette and colour distribution were considered. Border irregularity (B): This was judged by the unevenness of the perimeter. Color (C): Color variegation and number of colours were evaluated. Diameter (D): The largest in situ diameter in millimetres of each lesion was recorded<br>**Diagnosis based on:** Single observer (n=4)<br>**Observer qualifications:** Two experienced dermatologists, and two melanoma fellows<br>**Experience in practice:** Mixed experience (low and high experience combined)<br>**Experience with index test:** Not reported<br>**Dermoscopy:** evaluated in same study; ABCD and no algorithm |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | 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<br>Details: Each of the 72 melanocytic neoplasms was histopathologically diagnosed as with AMN or an early MM by a dermapathologist with special expertise in melanocytic neoplasms. Each lesion was completed excised and step sectioned.<br>Disease positive: 21 MMs; Disease negative: 51 AMN<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 21<br>51 atypical melanocytic nevus |
| 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 |
| **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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported<br>**Time interval between index test(s)**: not reported |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | Yes |
| Were all patients included in the analysis? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Low risk |

### Notes

| Notes | - |
|---|---|

## *Rosendahl 2011*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** 30-month period; dates NR<br>**Country** Australia |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Consecutive series of pigmented lesions submitted for histology from the primary care skin cancer practice of one author.<br>**Setting:** Primary care skin cancer practice<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Primary<br>**Exclusion criteria:** Poor image quality (considered under Flow and Timing); no other exclusion criteria reported<br>**Sample size (patients):** No. included: 389<br>**Sample size (lesions):** No. eligible: 466 pigmented lesions out of 1959 lesions excised or biopsied; No. included: 463<br>**Participant characteristics:** Mean age: 57y (SD 17). Male gender: 67.4%<br>**Lesion characteristics:** (53.1%) melanocytic. Lesion site: 17.7% head or face; Trunk: 52.1%; 27.6% extremities; 2.2% palms or soles. Melanoma thickness: ≤1mm: 1/29 melanoma (3.4%) |
| Are the included patients and chosen study setting appropriate? | Yes |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs overview and close up image presented<br>**Prior test data:** No further information used<br>**Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone.<br>**Diagnostic threshold:** Clinical diagnosis/subjective impression. Observers gave a diagnosis with level of confidence (from 0 for definitely benign to 100 for definitely malignant) after viewing the clinical images. (NB used authors threshold for detection of any skin cancer which includes lesions clinically considered to be MM, BCC pigmented epithelial carcinoma including SCC, keratoacanthoma, actinic keratosis and Bowen's disease as test positive; review only considered histologically confirmed MM, BCC or invasive SCC to be disease positive)<br>**Diagnosis based on:** Single observer (n=NR)<br>**Observer qualifications:** Expert dermatologist (based on author communication).<br>**Experience in practice:** Expert<br>**Experience with dermoscopy:** Expert |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | 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<br>Details: Excise or biopsy<br>Disease positive: 138; Disease negative: 325<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 9; Melanoma (in situ): 20; BCC: 72; cSCC: 5 (including 2 keratoacanthoma)<br>'Benign' diagnoses: 18 Bowen's disease and 14 actinic keratosis, 217 benign melanocytic plus additional 140 benign non melanocytic<br>*authors considered Bowen's disease, actinic keratosis and keratoacanthoma as malignant; all considered benign for review analysis |
| 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 |
| **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? | 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 | **Excluded participants**: Lesions were excluded due to poor image quality (n=3)<br>**Time interval to reference test**: Unclear; lesions 'routinely photographed' if scheduled for excision or biopsy but not further described<br>Time interval between index test(s): consecutive |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

### Notes

| Notes | - |
|---|---|

## *Scope 2008*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** after January 2003<br>**Country** Not reported |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Images of pigmented skin lesions selected from a database of standardised patient images provided by a New Zealand-based teledermatology company (MoleMap). Images were selected on the basis that (1) at least 8 clinically atypical nevi were apparent on the back; (2) most of the lesions on the back and all of the atypical nevi had close-up clinical digital images; (3) 1-year follow-up images (close-up clinical and dermoscopic images) were available to show that lesions considered to be benign were in fact biologically indolent by revealing no change; and (4) the image quality of both the overview and the close-up images were acceptable<br>**Setting:** New Zealand based teledermatology company; images were sent electronically to participants as a powerpoint file.<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Unspecified<br>**Exclusion criteria:** Poor quality index test image (considered under flow/timing); naevi on any body site except the back<br>**Sample size (patients):** No. eligible: 12; No. included: 12<br>**Sample size (lesions):** No. eligible: 145; No. included: 145<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None reported |
| Are the included patients and chosen study setting appropriate? | Unclear |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** Ugly duckling<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** No further information used<br>**Diagnostic threshold:** For each lesion that was deemed as different, the participants had to mark the lesion number on the form, identify it as either completely different or somewhat different from the other moles, give a short qualitative description of how the lesion differs, and report whether they would like to have a biopsy performed on the lesion<br>**Diagnosis based on:** Average (n=34)<br>**Observer qualifications:** Four subgroups in terms of clinical expertise: group 1, pigmented lesion experts (n = 8); group 2, dermatologists who were considered non-experts in pigmented lesion evaluation (n = 13); group 3, dermatology nurses (n = 5, including 1 dermatology medical photographer); and group 4, non-clinical medical staff (n = 8).<br>**Experience in practice:** Mixed experience (low and high experience combined)<br>**Other detail:** The study was sent electronically to participants as a powerpoint file (Microsoft Corp, Redmond, Washington) that contained the clinical image interface and a word document that contained questionnaire and response forms. The participants were not shown dermoscopic images. However, dermoscopic images of lesions (with a 1-year follow-up dermoscopic image) were available to the investigators to verify that lesions considered benign did not show dermoscopic features suggestive of malignancy, and the 1-year follow-up images confirmed that the lesions were in fact biologically indolent by revealing no change. |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | No |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | High 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 |
| Was the test interpretation carried out by an experienced examiner? | 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 plus follow up<br>**Details**: Unclear; all MMs were excised with histological confirmation and all benign had 1-year follow-up images (close-up clinical and dermoscopic images) to show that lesions considered to be benign were in fact biologically indolent by revealing no change, not clear whether any of the benign group were excised<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 5 'malignant melanoma'<br>Benign naevus: 140 |
| Is the reference standards likely to correctly classify the target condition? | Unclear |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | 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 | Unclear |
| 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 | **Excluded participants:** excluded if unacceptable image quality of both the overview and the close-up images<br>**Time interval to reference test:** not reported |
| Was there an appropriate interval between index test and reference standard? | Unclear |
| Did all patients receive the same reference standard? | No |
| 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? | Yes |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

### Notes

| Notes | - |
|---|---|

## *Soyer 1995*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Unclear<br>**Period of data collection** Not reported<br>**Country** Austria |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented skin lesion, difficult to diagnose on clinical grounds alone<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Clinical suspicion<br>**Setting for prior testing:** Secondary (general dermatology); referred by dermatologists or general physicians<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** NR<br>**Sample size (lesions):** No. included: 159<br>**Participant characteristics:** None reported<br>**Lesion characteristics** "23 melanomas with a Breslow index of </= 0.75mm, 13 melanomas with a Breslow index >/=0.76mm and </= 1.5mm, 12 melanomas with a Breslow index >/=1.51mm and </=3.5mm, 2 melanomas with a Breslow index of >/=3.5mm." |
| Are the included patients and chosen study setting appropriate? | Unclear |
| 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?** | Unclear |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Other test data**: Dermoscopy undertaken by same clinician(s) subsequent to clinical evaluation<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** n= 2 (1 or 2 per lesion)<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Not clearly described; assumed to be High; “Each lesion was examined clinically by .. one of the authors .. and a clinical diagnosis was recorded.” “After application of a drop of immersion oil, each lesion was examined dermoscopically …; the examination was performed by a dermatologist expert in dermoscopy and a dermoscopic diagnosis was recorded”<br>**Experience with index test:** Not described<br>**Other detail:** "Photographic documentation was performed using an incident light stereomicroscope (Wild M 650) equipped with a Minolta XG-M camera"<br>**Dermoscopy:** evaluated in same study; Pattern analysis |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 65 (41%); Disease negative: 94 (59%)<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 50; Melanoma (in situ): 15<br>BCC: pigmented basal cell carcinoma (3)<br>Sebhorrheic keratosis: 18; Clark's nevus of dysplastic nevus (61 cases); lentigo actinica lentigo (2), pigmented actinic keratosis (4), angioma (3), angiokeratoma (2). |
| 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 |
| **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? | 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Stanganelli 1998*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case control<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** Just states 1997<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | No |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Images of pigmented skin lesions selected from computerised files of the skin cancer clinic.<br>**Setting:** Training study; images selected from skin cancer clinic<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Unspecified<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** Not reported<br>**Sample size (lesions):** No. included: 30PSLs<br>**Participant characteristics:** None reported<br>**Lesion characteristics** None 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** No further information used<br>**Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone (images were randomised).<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Average; n= 20<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Not described; 30 dermatologists with “experience in ELM but (with) no formal training” attended a seminar on clinical and ELM diagnosis of PSL; 20 then participated in a test of their diagnostic accuracy. A second session on ELM was then held.<br>**Other detail:** The observers received 2hrs seminar of the principles of clinical diagnosis of NMLs~~S~~, BCC, MN and MM. The participants were then invited to undergo an anonymous test of their diagnostic accuracy.<br>**Dermoscopy:** evaluated in same study; no algorithm |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **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<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 10<br>BCC: 4<br>Mild//moderate dysplasia: 3; Sebhorrheic keratosis: 3; Benign naevus: Melanocytic nevi-7<br>Other: 1 hemangioma1 subungunal hemorraghe1 plantar intraepidermal hemorraghe |
| 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 |
| **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? | 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Stanganelli 2000*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective<br>**Period of data collection** 1994-1996<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Patients with pigmented skin lesions referred by dermatologists and general practitioners either for pre-surgical assessment or consultation<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** patients referred for pre-surgical assessment or consultation indicating they have had prior tests<br>**Setting for prior testing:** Primary some patients referred for consultation only; dermoscopy findings are reported back and management decision remains with referring clinician; Secondary (general dermatology)<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 1556<br>**Sample size (lesions):** No. eligible: 3372; No. included: 3372<br>**Participant characteristics:** Median age 30 years, range 10 to 94; Male: 522 (34%)<br>**Lesion characteristics:** None reported |
| Are the included patients and chosen study setting appropriate? | Yes |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** ABCD<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Other test data**: Dermoscopic and clinical images subsequently presented separately to observer subsequent to diagnosis using clinical images alone.<br>**Diagnostic threshold: NR**<br>**Diagnosis based on:** Single observer; n= 1<br>**Observer qualifications:** Not reported; described as one of the co-authors and study based in skin cancer clinic - likely dermatologist<br>**Experience in practice:** Not described<br>**Other detail:** A crude clinical image (magn X6 and X10) was recorded in the digital database<br>**Dermoscopy:** evaluated in same study (image based); Pattern analysis |

## Visual Inspection - in-person

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | Yes |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | Unclear |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability | |
|---|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis plus follow up; histology report of known surgical excisions (n=262) plus a cancer-registry based follow up of benign cases (n=3110)<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 55; BCC: 43<br>'Benign' diagnoses: 3274 |
| Is the reference standards likely to correctly classify the target condition? | No |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | Unclear |
| **Could the reference standard, its conduct, or its interpretation have introduced bias?** | High 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported<br>**Time interval between index test(s)**: not clearly reported just indicated that D-ELM was performed soon after clinical examination |
| Was there an appropriate interval between index test and reference standard? | Unclear |
| Did all patients receive the same reference standard? | No |
| Were all patients included in the analysis? | Yes |
| 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? | Yes |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

### Notes

| Notes | - |
|---|---|

## *Stanganelli 2005*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Unclear (likely case series)<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** NR<br>**Country** Italy<br>**Test set derived** A training set of 22 melanomas and 218 melanocytic nevi was randomised from the dataset. The test set was formed by the complement (the remaining 20 melanomas and 217 nevi). A further subset of images from the original dataset, consisting of 31 melanomas and 103 nevi, was used for the comparison between observers and CAD; derivation of the subset not reported. |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Unclear |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Melanocytic lesions from patients referred to the Skin Cancer Unit and undergoing clinical and dermoscopic evaluation; images were 'selected' from a larger image database. Potential overlap with Stanganelli 2000 (not possible to determine).<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Clinical and/or dermatoscopic suspicion<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 1556 referred / No. included: NR<br>**Sample size (lesions):** No. eligible: 3274 / No. included: 477 melanocytic lesions; 237 in test set and 134 in comparison between CAD and human operators<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** Melanoma thickness 61.2% <0.75mm |
| 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** General practitioners evaluated only clinical images; unclear for dermatologists<br>**Other test data**: dermatologists examined both clinical and dermoscopic images but unclear whether clinical diagnosis was made prior to presentation of dermoscopic images<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Average (n=6)<br>**Observer qualifications:** GP 3; Dermatologist 3<br>**Experience in practice:** Assumed Low for GPs; High for dermatologists - described as “dermatologists with experience in ELM (2 years)”<br>**Other detail:** Digital images included melanocytic lesions evaluated in ELM with a fixed x16 magnification<br>**Dermoscopy:** evaluated in same study; no algorithm |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | 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 plus cancer registry<br>All included lesions underwent histology but some were identified using a cancer-registry-based follow-up of benign diagnoses.<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 42 in full sample; 31 in CAD vs human observer interp and 20 in test set<br>'Benign' diagnoses: 435 melanocytic nevi; 103 in CAD-observer comp and 217 in test set |
| 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 |
| **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 | Unclear |
| 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported |
| Was there an appropriate interval between index test and reference standard? | Unclear |
| Did all patients receive the same reference standard? | No |
| Were all patients included in the analysis? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

### Notes

| Notes | - |
|---|---|

## *Steiner 1987*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Prospective<br>**Period of data collection** not specified<br>**Country** Austria |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Small (< 10 mm) pigmented skin lesions considered diagnostically equivocal in that there was no absolute agreement on the clinical diagnosis among investigating clinicians at a pigmented lesions clinic.<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** NR<br>**Sample size (lesions):** 318<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI):** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A<br>**Other test data**: Dermoscopy undertaken by same clinician(s) subsequent to clinical evaluation<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Consensus (3 observers) "All lesions were independently seen and diagnosed by the three investigators, and the diagnosis that appeared most probable to at least two of the three investigators was recorded as the clinical"; n= 3<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** High experience or 'Expert' "experienced dermatologists"<br>**Experience with index test:** - "experienced dermatologists" |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | Yes |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 73 melanomas, 20 BCCs; Disease negative: 225<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 49; Melanoma (in situ): 15; Lentigo maligna 9 (also includes lentigo maligna melanoma)<br>BCC: 20<br>Sebhorrheic keratosis: 20; Junctional naevi 39; Blue naevus 29; Dysplastic naevus 75; Lentigo simplex and nevoid lentigo 19; Angioma/angiokeratoma 15 |
| 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 |
| **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? | 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: assumed consecutive; following diagnosis, lesions subsequently excised<br>**Time interval between index test(s)**: consecutive |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | Yes |
| Were all patients included in the analysis? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Low risk |

Notes

| Notes | - |
|---|---|

*Thomas 1998*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case control; separate recruitment<br>**Data collection:** Retrospective<br>**Period of data collection** NR; appears to be post-1992<br>**Country** France |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | No |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Retrospective selection of all 460 cases of melanoma and a nonselected consecutive group of 680 nonmelanoma pigmented tumours<br>**Setting:** Secondary (general dermatology)<br>**Prior testing:** Selected for excision (no further detail) All excised<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** Not reported<br>**Sample size (lesions):** No. included: 1140<br>**Participant characteristics:** None reported<br>**Lesion characteristics: Other test data**: Dermoscopy undertaken by same clinician(s) subsequent to clinical evaluation |
| 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 | **Visual inspection (VI):** ABCDE<br>**Method of diagnosis:** In person diagnosis; dermatologist making referral for excision made the diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** Number of characteristics present (from>=1 to all 5)<br>**Diagnosis based on:** Single observer;n= NR<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Assumed to be High; described as 'trained' dermatologists<br>**Other detail:** Preliminary meeting held to precisely define each criterion, agree on the significance of each abnormality and define the appropriate way to fill in the study form. ABCDE: Criterion A was defined as geometrical asymmetry in two axes of the tumour, criterion B as irregular (unsharp or ill-defined or angular) borders, criterion C as presence of at least two different colours within the lesion (with the exception of the usual symmetrical darkening of the lesion in its center), criterion D as diameter equal or superior to 6mm. Criterion E, the only anamnestic (based on the patient's description of the natural history of the lesion) criterion was defined as enlargement of the surface (and not in height) of the lesion. |

## Visual Inspection - in-person

| 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? | No |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Was the test applied and interpreted in a clinically applicable manner? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | Yes |
| Was the test interpretation carried out by an experienced examiner? | Yes |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | Low concern |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 460; Disease negative: 680<br>**Target condition (Final diagnoses)**<br>Melanoma (in situ and invasive, or not reported): 460<br>BCC: BCC<br>Sebhorrheic keratosis: 19; 576 benign pigmented naevus; 55 dysplastic naevi; 4 blue naevi; 2 compound naevi with sutton inflammatory infiltrate; 2 spitz; 1 reeds naevi; 3 haemangiomas; 9 dermatofibromas; 1 accessory nipple |
| 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? | No |
| **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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Troyanova 2003*

### Patient Selection

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

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Images of pigmented skin lesions <=13mm in diameter selected for a dermoscopy training study<br>**Setting:** Training study<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Not reported<br>**Exclusion criteria:** None reported<br>**Sample size (patients):**<br>**Sample size (lesions):** No. included: 50 lesions<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** Melanoma thickness: ≤1mm: 100% |
| 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 | **Visual inspection (VI):** No algorithm<br>**Method of diagnosis:** Clinical photographs and dermoscopic images<br>**Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone.<br>**Prior test data:** No further information used<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Average; n= 32<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** High experience or 'Expert'<br>**Experience with index test:** Low experience / novice users; experienced in PSL field but not ELM<br>**Dermoscopy:** evaluated in same study; no algorithm |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | 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<br>Disease positive: 25; Disease negative: 25<br>**Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 25<br>'Benign' diagnoses: ~~50~~25 "not melanoma" |
| 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 |
| **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? | 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported<br>**Time interval between index test(s)**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

Notes

| Notes | - |
|---|---|

### *Unlu 2014*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Retrospective image selection / Prospective interpretation<br>**Period of data collection** January 2008-January 2010<br>**Country** Turkey |
| 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:** Melanocytic lesions excised at Ankara University Department of Dermatology Pigmented Lesion Clinic<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic) Ankara University Department of Dermatology Pigmented Lesion Clinic<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** Location/site of lesion facial, nail and volar acral lesions were excluded; non-melanocytic appearance<br>**Sample size (patients):** No. included: 115<br>**Sample size (lesions):** No. included: 115<br>**Participant characteristics:** Mean age: 38.72y (+/- 18.46 y). Male gender: n=56 (49%).<br>**Lesion characteristics:** Lesion site: 100% trunk and limbs. Melanoma thickness: 10 (41.7%) <0.75mm; 14 (58.3%) >=0.75mm |
| Are the included patients and chosen study setting appropriate? | No |
| Did the study avoid including participants with multiple lesions? | Yes |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** No algorithm; Appears to be original clinical diagnosis at time of lesion presentation<br>**Method of diagnosis:** In person diagnosis Appears to be diagnosis on presentation<br>**Prior test data:** N/A in-person diagnosis<br>**Other test data**: Dermoscopic images presented to different observers<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Unclear - for visual inspection appears to be single examiner at time of clinic diagnosis (n=NR); dermoscopic images "scored by three other experienced dermatoscopists" (n=3)<br>**Observer qualifications:** Not reported; assumed dermatologists - described as experienced dermatoscopists<br>**Experience in practice:** Unclear for clinic diagnosis; dermatoscopists described as "experienced"<br>**Experience with index test:** Described as "experienced"<br>**Dermoscopy:** evaluated in same study by 3 experienced dermoscopists; 3-point rule; 7-point checklist; ABCD; CASH algorithm (image based) |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 24; Disease negative: 91<br>**Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 24<br>'Benign' diagnoses: 91 melanocytic benign lesions; 37 (32.2%) dermal nevi; 15 (13%) clark's nevi; 14 (12.2%) compound nevi; 13 (11.3%) blue nevi; 6 (5.2%) spitz nevi; 4 (3.5%) congenital melanocytic nevi; 2 (1.7%) junctional nevi |
| 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 |
| **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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported<br>**Time interval between index test(s):** Appear to be consecutively applied |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

## Notes

| Notes | - |
|---|---|

### *Viglizzo 2004*

## Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Not reported<br>**Period of data collection** Not reported<br>**Country** Italy |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | Unclear risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Pigmented skin lesions examined at the Dermoscopy Service and undergoing excision; a modified version of Kenet's risk stratification approach for dermoscopy (Ascierto 2000) was used to select high and very high risk lesions for excision; medium and low risk lesions were excised based on cosmetic or functional reasons. (2x2 data have been extracted only for melanocytic subgroup).<br>**Setting:** Specialist unit (skin cancer/pigmented lesions clinic) Dermoscopy service at a university department (Department of Endocrinologic and metabolic disease)<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. eligible: 349 patients; No. included: not reported<br>**Sample size (lesions):** No. eligible: 520 lesions; No. included: 79 lesions excised included in the final analysis<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI).** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** Unclear<br>**Other test data**: Dermoscopy undertaken by same clinician(s) subsequent to clinical evaluation<br>**Diagnostic threshold:** Not reported; correct diagnosis of melanoma<br>**Diagnosis based on:** Single observer (n=NR; "All dermoscopic evaluations were performed by the same operators")<br>**Observer qualifications:** Not reported; "each lesion was .. diagnosed clinically and dermoscopically" at the Dermoscopy service<br>**Experience in practice:** Not described<br>**Experience with dermoscopy:** Not described; assumed High as diagnosis at 'Dermoscopy service'<br>**Dermoscopy:** evaluated in same study; no algorithm |

## Visual Inspection - in-person

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear risk |

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

### Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>**Target condition (Final diagnoses)** Melanoma (invasive): 11; Melanoma (in situ): 1<br>Melanocytic lesion: 67 |
| 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 |
| **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? | 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

## *Walter 2012*

### Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Randomised controlled trial (control group only included)<br>**Data collection:** Prospective<br>**Period of data collection** March 2008 to May 2010<br>**Country** UK |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | Yes |
| Did the study avoid inappropriate exclusions? | Yes |
| **Could the selection of patients have introduced bias?** | Low risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Adults with any suspicious pigmented skin lesion, i.e. any lesion presented by a patient, or opportunistically seen by a family doctor or practice nurse, that could not immediately be diagnosed as benign and about which the patient could not be reassured.<br>**Setting:** Primary 15 general practices in eastern England<br>**Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion<br>**Setting for prior testing:** Primary<br>**Exclusion criteria:** Those unable to give informed consent or considered inappropriate to include by their family doctor.<br>**Sample size (patients):** No. eligible: 1297; No. included: 1293<br>**Sample size (lesions):** No. eligible: 1580; No. included: 1583<br>**Participant characteristics:** Mean age: 44.6y (SD 16.8). Male: 465 (36%). Ethnicity: White 1214 (93.9%); Mixed 45 (3.5%); Missing: 34 (2.6%)<br>**Lesion characteristics.** Lesion thickness ≤1mm: in 'more than half' of MM |
| Are the included patients and chosen study setting appropriate? | Yes |
| Did the study avoid including participants with multiple lesions? | No |
| **Are there concerns that the included patients and setting do not match the review question?** | High |

## Index Test

| | |
|---|---|
| Index tests | **Visual inspection (VI)** Glasgow/MacKie revised seven-point checklist (MacKie 1990)<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Single observer (n=30)<br>**Observer qualifications:** 28 GPs and 2 nurse practitioners recruited as 'lead clinicians' (2 per practice); appears though they conducted all skin examinations. Excluded GPs with known dermatological expertise, e.g. current hospital practitioners, clinical assistants in dermatology, and GPs with a special interest in dermatology<br>**Experience in practice:** Mixed GP experience - median of 15 years experience (range 4 to 27yrs); assumed low experience with pigmented skin lesions- seven had undergone some training in dermatology, three had a short dermatology training post, three were on clinical attachment to an out-patient clinic, and one was unspecified |

## Visual Inspection - in-person

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | Yes |
| Was the test interpretation carried out by an experienced examiner? | No |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis plus FU and Expert opinion<br>*Histology (not further described)* 215 (histology result missing in further 4)<br>Disease positive: 35; Disease negative: 180<br>*Clinical FU plus histology of suspicious lesions:* 22 of the 411 referred patients were monitored (not further described); 566 of the 1162 not referred underwent expert review and were then re-assessed at 3-6 months<br>Disease positive: 1; Disease negative: 588<br>*Expert opinion.* Reviewed by two dermatology experts using the recorded clinical history and examination, a digital photograph, and MoleMate image where available.<br>Disease positive: 0; Disease negative: 725<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 30; Melanoma (in situ): 6; BCC: 10<br>'Benign' diagnoses: 1306 |
| Is the reference standards likely to correctly classify the target condition? | No |
| Were the reference standard results interpreted without knowledge of the results of the index tests? | No |
| **Could the reference standard, its conduct, or its interpretation have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Expert opinion (with no histological confirmation) was not used as a reference standard | No |
| 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?** | High |

## Flow and Timing

| A. Risk of Bias | |
|---|---|
| Flow and timing | **Excluded participants:** 417 withdrew from control group after randomisation - 10 did not attend for dermatology assessment; 19 excluded; 1 died; 4 missing histology (in referred group; included as benign?); plus 12 with unknown outcome (in non-referred group, assumed benign and included)<br>**Time interval to reference test**: suspicious lesions referred under 2 week wait system |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | No |
| 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? | Yes |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

Notes

| Notes | - |
|---|---|

### *Westerhoff 2000*

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case control (for lesion selection; study was an RCT of dermoscopy training for PCPs)<br>**Data collection:** Retrospective<br>**Period of data collection** Not reported<br>**Country:** Australia |
| Was a consecutive or random sample of patients enrolled? | Yes |
| Was a case-control design avoided? | No |
| Did the study avoid inappropriate exclusions? | Unclear |
| **Could the selection of patients have introduced bias?** | High risk |

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Clinically atypical pigmented skin lesions; 50 invasive melanomas and 50 nonmelanomas randomly selected from the Sydney Melanoma Unit pigmented skin lesions (PSL) image database.<br>**Setting:** Specialist unit (lesion selection)<br>**Prior testing:** Selected for excision or followed up<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic)<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** No. included: NR<br>**Sample size (lesions):** No. included: 100<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** median Breslow thickness 0.6mm |
| 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 | **Visual inspection (VI):** No algorithm<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** Unclear; all participants "were instructed not to look at the surface microscopic image until they had scored the clinical image"<br>**Diagnostic threshold:** Not reported<br>**Diagnosis based on:** Average (n=37; 74 practising primary care practitioners randomised to dermoscopy education intervention or not). (Diagnoses were recorded for both groups of GPs at baseline (pre-test) and after the training intervention had been administered to the intervention group (post-test), resulting in 8 sets of 2x2 data based on interpretation of the same set of 100 lesions; post-test data for the intervention group of GPs was used for the Visual Inspection analysis.)<br>**Observer qualifications:** GP<br>**Experience in practice:** Considered to be Low; Only practitioners who had had no formal training with surface microscopy and did not use a surface microscope in their clinical practice were included.<br>**Experience with dermoscopy:** Low experience / novice users (non-training arm); 'Trained' for the intervention arm<br>**Other detail:** Camera designed for close-up clinical photography (Elicar Macrolens, Japan)<br>**Dermoscopy:** evaluated in same study; Menzies criteria (Intervention arm underwent training in Menzies criteria) |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | 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 plus follow up<br>*Histology:* All the lesions except two had been excised after photography and subjected to histopathological examination. Disease positive: 50 / Disease negative: 48<br>*Clinical FU plus histology of suspicious lesions:* The two benign PSL that had not been excised were monitored over a longer period of time and had shown no morphological change.<br>Length of FU: NR; Disease positive: 0 / Disease negative: 2<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 50 / 'Benign' diagnoses: 50 |
| 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 |
| **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? | 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 | **Excluded participants:** none reported<br>**Time interval to reference test:** "All the lesions except two had been excised after photography" |
| Was there an appropriate interval between index test and reference standard? | Yes |
| Did all patients receive the same reference standard? | No |
| Were all patients included in the analysis? | Yes |
| 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? | Unclear |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | High risk |

## Notes

| Notes | - |
|---|---|

### *Winkelmann 2016*

## Patient Selection

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

| B. Concerns regarding applicability | |
|---|---|
| Patient characteristics and setting | **Inclusion criteria:** Images of pigmented skin lesions previously analysed by a digital classifier MSDSLA; method of selection of the 12 not reported<br>**Setting:** Dermoscopy conference<br>**Prior testing:** Not reported<br>**Setting for prior testing:** Unspecified<br>**Exclusion criteria:** None reported<br>**Sample size (patients):** Not reported<br>**Sample size (lesions):** No. included: 12<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** Clinical photographs<br>**Prior test data:** Unclear<br>**Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone.<br>**Diagnostic threshold:** Not reported - biopsy decision<br>**Diagnosis based on:** Average (n=70)<br>**Observer qualifications:** Dermatologist<br>**Experience in practice:** Not described; recruited “dermatologists at a dermoscopy conference”; no further details<br>**Other detail:** Authors report that practitioners with a particular interest in skin cancer or technology may have chosen to attend this conference and/or self-selected to take part in the study.<br>**Dermoscopy:** evaluated in same study; no algorithm |

## Visual Inspection - in-person

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

## Visual inspection - image-based

| 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? | Unclear |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **Could the conduct or interpretation of the index test have introduced bias?** | Unclear 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? | No |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **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<br>Disease positive: 5 / Disease negative: 7<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive): 3 / Melanoma (in situ): 2<br>Mild/moderate dysplasia: 7 low grade dysplastic nevi |
| 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 |
| **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? | 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported |
| 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? | Yes |
| 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? | Yes |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

Notes

| Notes | - |
|---|---|

***Zaumseil 1983***

Patient Selection

| A. Risk of Bias | |
|---|---|
| Patient Sampling | **Study design:** Case series<br>**Data collection:** Not reported<br>**Period of data collection** 1976-1981<br>**Country** Germany |
| Was a consecutive or random sample of patients enrolled? | Unclear |
| 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:** Skin lesions undergoing excision<br>**Setting:** Secondary (not further specified)<br>**Prior testing:** Selected for excision (no further detail)<br>**Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) Described as 'skin clinic'<br>**Exclusion criteria:** Disagreement between evaluators on tumour histological classification Those in which the histological diagnosis was 'unclear' were excluded melanoma metastases were excluded<br>**Sample size (patients):** Not reported<br>**Sample size (lesions):** No. included: 7063<br>**Participant characteristics:** None reported<br>**Lesion characteristics:** None 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 | **Visual inspection (VI)** No algorithm<br>**Method of diagnosis:** In person diagnosis<br>**Prior test data:** N/A in-person diagnosis<br>**Diagnostic threshold:** Primary diagnosis of melanoma (method of Kopf 1975 was cited)<br>**Diagnosis based on:** Single observer (n=NR)<br>**Observer qualifications:** Not reported<br>**Experience in practice:** Not described<br>**Experience with index test:** Not described |

## Visual Inspection - in-person

| 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 |
| For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others? | |
| **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? | Yes |
| Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? | No |
| Was the test interpretation carried out by an experienced examiner? | Unclear |
| **Are there concerns that the index test, its conduct, or interpretation differ from the review question?** | High |

## Visual inspection - image-based

| A. Risk of Bias |
|---|

| B. Concerns regarding applicability |
|---|

### Reference Standard

| A. Risk of Bias | |
|---|---|
| Target condition and reference standard(s) | **Reference standard** Histological diagnosis alone<br>Disease positive: 337 / Disease negative: 6726<br>**Target condition (Final diagnoses)**<br>Melanoma (invasive or in situ): 337<br>Other diagnoses only listed for the 89 FPs: 23 benign nevi; 13 BCC; 12 blue nevus; 11 angiomatosis; 10 sebhorrheic keratosis; 6 histiocytoma; 4 spitz nevus; 4 lentigo; 3 Bowen's disease; 1 acrospiroma; 1 keratinizing papilloma |
| 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 |
| **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? | 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 | **Excluded participants**: none reported<br>**Time interval to reference test**: not reported |
| 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? | Yes |
| 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? | |
| If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less? | |
| **Could the patient flow have introduced bias?** | Unclear risk |

### Notes

| Notes | - |
|---|---|

*Footnotes*

ABCD(E) - asymmetry, border, colour, differential structures (enlargement); AK - actinic keratosis; AMN – atypical melanocytic naevi; BCC - basal cell carcinoma; BD – Bowen's disease; BN – benign naevi; BPC – between person comparison (of tests); CAD – computer assisted diagnosis; CCS – case control study; CD – compact disc; CM – cutaneous melanoma; CMM - cutaneous malignant melanoma; CS – case series; CSCC – cutaneous squamous cell carcinoma; DF – dermatofibroma; Dx – diagnosis; ELM – epiluminescence microscopy; FPs - false positives; FU – follow-up ; GP – general practitioner; H&E – haematoxylin and eosin stain; LPLK – lichen planus-like keratosis; LS – lentigo simplex; MM – malignant melanoma; MiS – melanoma in situ (or lentigo maligna); MN – melanocytic naevi; MSDSLA - multispectral digital skin lesion analysis device; N/A – not applicable; NC – non comparative; NMLs – non melanocytic lesions; NR – not reported; P – prospective; PCPs – primary care providers; PLC – pigmented lesion clinic; PSL – pigmented skin lesion; R –retrospective; RCM – reflectance confocal microscopy; RCT – randomised controlled trial; SCC - squamous cell carcinoma; SD – standard deviation; SDDI - short-term sequential digital dermoscopy imaging; SK – seborrhoeic keratosis; SN – Spitz nevi; SSM – superficial spreading melanoma; SVS – support vector system; VI – visual inspection; WPC – within person comparison (of tests).

## Characteristics of excluded studies

### *Abbasi 2004*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*systematic review* |

*Aldridge 2011*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on test observer<br>*medical students and lay persons* |

*Aldridge 2011a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on test observer |

*Aldridge 2013*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*not test accuracy study* |

*Alendar 2009*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*only 7 reported verified histologically* |

*Argenziano 1999*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Only includes melanoma* |

*Argenziano 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Table V gives se/sp data for 108 lesions but can't derive the number of melanoma for this subset of the original 128*<br>Authors contacted 10/5/16; 24/6/16 |

*Argenziano 2012*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*no follow-up of test negatives* |

*Argenziano 2014*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data |

*Ascierto 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

*Badertscher 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data |

*Bafounta 2001*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*systematic review* |

### *Banky 2005*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>EXCLUDE on index test |

### *Basarab 1996*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*not all suspected of skin cancer*<br>EXCLUDE on 2x2 data |

### *Bauer 2000*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Does not provide 2x2 data for visual inspection alone* |

### *Bauer 2005*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*follow-up/monitoring study* |

### *Becker 1954*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### *Benelli 2000*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*only inter-rater reliability data given (n=25); authors have published much larger evaluations of 7FFM and ABCD* |

### *Blum 2004*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*comment paper* |

### *Blum 2004a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*letter*<br>EXCLUDE *Letter only; limited data presented - evaluates '3-colour' rule as developed By* MacKie 1992 *(excluded as assessment of individual lesion features only)* |

### *Blum 2004b*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Evaluates dermoscopy only* |

### *Bolognia 1990*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*no ref standard diagnosis for index test negatives* |

### *Bono 2001*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*aim of the study is to determine what features are present in amelanotic cutaneous melanoma* |

### *Borsari 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE if individual lesion characteristics |

### *Borve 2012*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*includes participants without skin lesions*<br>EXCLUDE on sample size<br>*<5 BCC* |

### *Brown 2000*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*systematic review* |

### *Brown 2009*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on test observer<br>*lay persons* |

### *Buhl 2012*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*folow up/monitoring*<br>EXCLUDE duplicate or related publication<br>*same patients as Haenssle 2010 #191* |

### *Burki 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### *Burr 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### *Burton 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*can only get 2x2 data for referral accuracy*<br>EXCLUDE on 2x2 data |

### *Carli 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*Only 39/1042 with ref test* |

### *Carli 2003a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on sample size |

### *Carli 2004*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on sample size<br>*<5 MM per arm*<br>EXCLUDE on 2x2 data |

### *Carli 2004a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>EXCLUDE but contact authors<br>*Author passed away; unable to make contact with co-authors* |

### *Carli 2004b*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*'Clinical diagnosis' - Dataset covers 1997-2001, but dermoscopy routinely introduced 1998; authors contacted but no response.* |

### *Carli 2005*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Only sensitivity data given (% with correct diagnosis); % of benign lesions incorrectly diagnosed was not reported*<br>EXCLUDE but contact authors |

### *Carlos-Ortega 2007*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Gives se/sp for visual inspection and dermoscopy in the English abstract. 68 patients/70 lesions were included but only 36 seem to have had visual inspection results and all underwent dermoscopy. Two observers performed each test blinded to each other. Table I gives 22 with BCC and 11 with melanoma overall (no. D+ not reported for those with VI results), but using either or both of these numbers with the se/sp provided does not give the same PPV and NPV as given by the authors*<br>EXCLUDE but contact authors<br>*data not clearly presented for 2x2; translator suggested alternative but still does not work out to what is in paper; tried contacting authors twice, no reply* |

### *Chen 2001*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*Systematic review comparing PCP accuracy with dermatologist accuracy.* |

### *Chen 2006*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*only given AUC* |

### Chiaravalloti 2014

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Includes melanoma only* |

### Ciudad-Blanco 2014

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Includes melanoma only* |

### Cooper 2002

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*Insufficient data for inclusion in melanoma review* |

### Cornell 2015

| | |
|---|---|
| Reason for exclusion | EXCLUDE on test observer |

### Cox 2008

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*Se and sp estimates for diagnosis of melanoma for both the seven-point checklist and the revised (10-point) checklist; reference standard not reported for any of the 381 TWR referrals for melanoma*<br>EXCLUDE but contact authors<br>*Author contacted 10/05/16; co-author contacted 24-6-16* |

### De Giorgi 2011

| | |
|---|---|
| Reason for exclusion | EXCLUDE Duplicate publication<br>Study appears to use same lesions as Carli 2003b (included study). Both studies have the same numbers of melanomas and benign nevi and have common co-authors (De Giorgi 2011 in particular). Although not explicit, the De Giorgi 2011 paper appears to have used the same lesions and study design but with different observers. The original Carli 2003 paper reported using 8 expert observers while the later paper recruited 8 dermatologists who had undergone a dermoscopy training course but who reported no experience in assessing pigmented skin lesions. |

### DeCoste 1993

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Not given the total number of D+/D- or total number of lesions included. Just given the sensitivity/specificity values* |

### Di Carlo 2014

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test. Videothermography not relevant for the review and there is no 2x2 data for dermoscopy EXCLUDE if derivation study. Only includes AK and BCC; no 2x2 for dermoscopy |

### Di Chiacchio 2010

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*Excluding nail bed melanoma*<br>EXCLUDE on 2x2 data<br>*There is insufficient data to extract for a 2x2 table* |

### *Dreiseitl 2009*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Does not evaluate visual inspection alone* |

### *Duff 2001*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Does not evaluate visual inspection alone* |

### *Edmondson 1999*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*It seems that the reference standard here is expert diagnosis. This is not a teledermatology paper* |

### *Emmons 2011*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*not test accuracy study; promoting primary prevention* |

### *Engelberg 1999*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on sample size<br>*only 1 confirmed melanoma and 3 BCC* |

### *English 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*no accuracy data given* |

### *English 2004*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*no accuracy data* |

### *Fabbrocini 2008*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*there isn't sufficient data provided for each index test to populate 2x2 table*<br>EXCLUDE but contact authors<br>*Requested cross tabulation of each clinician's diagnosis (e.g. at threshold of >=3 on 7 point checklist) against the histological diagnosis and/or a cross tabulation of the remote diagnosis against the Face to Face diagnoses. Author responded* 30-6-16 cannot access data needed |

### *Federman 1995*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Not test accuracy* |

### *Fikrle 2013*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*Follow up study <50% of study participants have their final diagnosis reached by histopathology.* |

### *Freeman 1963*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Only gives % correct for each lesion type*<br>EXCLUDE but contact authors<br>*Tables 2 and 3 appear to give % correct diagnoses per lesion type, but does not give data on numbers misclassified as melanoma, or other malignancy, i.e. FPs. Author responded; paper too old, cannot provide data* |

### *Friedman 1985*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### *Funt 1963*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>EXCLUDE on 2x2 data<br>*No 2x2 data* |

### *Gerbert 1996*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*No breakdown of final diagnoses for included lesions*<br>EXCLUDE on 2x2 data<br>*Only gives % correct for each lesion type; not sensitivity/specificity* |

### *Gerbert 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data |

### *Giannotti 2004*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*a review* |

### *Grana 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>EXCLUDE if individual lesion characteristics<br>*only looking at lesion border* |

### *Grob 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### *Guibert 2000*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*Not designed as an accuracy study only observational. Can't get 2x2 data >50% of study participants did not recieve histology as ref standard.* |

### *Gunduz 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on sample size<br>*case study* |

### *Gutierrez 2013*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*test to improve histopathology diagnosis* |

### *Hacioglu 2013*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*Does not provide sufficient data for detection of melanoma* |

### *Haenssle 2010*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*test used for monitoring and not initial diagnosis; no VI data* |

### *Haenssle 2010a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Does not report specificity*<br>EXCLUDE duplicate or related publication<br>*same patients as Haenssle 2010 #191* |

### *Hallock 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*'clinical diagnosis'; dermoscopy used for 3 of 4 years* |

### *Haniffa 2007*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*looks like approximately 20% of patients received a final diagnosis by histology. 179 biopsies were performed. Total sample was 881 lesions* |

### *Har-Shai 2001*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*'clinical diagnosis'* |

### *Heal 2008*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Sensitivities and PPVs are given so theoretically a 2x2 could be worked out but the numbers do not appear to work out*<br>*Author response; the 2x2 table the Cochrane researchers want to create is not possible for our results, because sensitivity and PPV are based on different sample sizes.* |

*Healsmith 1994*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*Benign lesions described as 'clinically diagnosed' rather than histology/follow-up* |

*Higgins 1992*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Includes only benign lesions*<br>EXCLUDE on sample size<br>*No melanomas*<br>EXCLUDE on 2x2 data<br>*No malignant cases* |

*Hoorens 2016*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>EXCLUDE on reference standard<br>*No info on numbers undergoing histology; and no follow-up reported for benign appearing lesions*<br>EXCLUDE on 2x2 data |

*Huang 1996*

| | |
|---|---|
| Reason for exclusion | EXCLUDE if individual lesion characteristics<br>*Border irregularity not overall dx*<br>EXCLUDE on 2x2 data |

*Jamora 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*no referene standrd for index test negatives* |

*Janda 2014*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on sample size<br>*only one case of melanoma, one case of BCC and one of SCC* |

*Jensen 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*comment paper* |

*Jolliffe 2001*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Provides data for clinical diagnosis (including dermoscopy for some cases)* |

### *Jonna 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*only included index test positives to get PPV, not worth author contact on this one* |

### *Kaddu 1997*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on sample size<br>*Sample size <5; not test accuracy* |

### *Keefe 1990*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*Only 28% (60/214) of non melanoma group had excision* |

### *Kelly 1986*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*Can't disaggregate the severely dysplastic/in situ MM*<br>EXCLUDE on sample size<br>*unclear whether >5 in situ melanoma* |

### *Koh 1990*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*screening study; no adequate reference standard* |

### *Kroemer 2011*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Provides data for clinical diagnosis (including dermoscopy for some cases)* |

### *Krol 1991*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*No follow up reported for those who were test negative* |

### *Kurvers 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Collective intelligence - majority rule and quorum rule applied to large number of test interpreter decisions*<br>EXCLUDE duplicate or related publication<br>*re-analyses data from 2 previously published studies to determine whether collective intelligence (i.e. majority rules or quorum rules across a large number of observers) improves test accuracy. We have excluded one of these studies as the number of melanomas is not provided ( Argenziano 2003 ) and included the other in dermoscopy review ( Zalaudek 2006 ).* |

### *Kvedar 1997*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Not all suspected of skin cancer* |

### *Lechner 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*Erratum* |

### *Lewis 1999*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Study appears to meet all eligibility criteria but disease prevalence not given alongside se/sp*<br>EXCLUDE but contact authors<br>Authors contacted 10/05/2016; email returned |

### *Lindelöf 1994*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*only malignant melanoma*<br>EXCLUDE on 2x2 data<br>*not enough information given to derive a 2x2 table. only given for a sample of 50 patients who had a strong suspicion of melanoma clinically. Do not know what happened to those with no suspicion clinically* |

### *Lorentzen 2000*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Does not provide data for visual inspection alone* |

### *Luttrell 2012*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on test observer<br>*Accuracy data only given for lay-persons not interested in this population of test observers* |

### *Machet 2005*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>***[Note this is a staging study]* |

### *MacKenzie-Wood 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*only malignant diagnosis* |

### *MacKie 1990*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### *Mackie 1991*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*letter* |

### Mackie 2002

| | |
|---|---|
| Reason for exclusion | EXCLUDE if individual lesion characteristics<br>*presence of 3 or more colours on dermoscopy* |

### Mahendran 2005

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Face to face is 'clinical diagnosis', i.e. visual inspection +/- use of dermoscopy* |

### Mahon 1997

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*a summary of a comparison of two screening checklists* |

### Malvehy 2014

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Does not report data for visual inspection alone* |

### Marghoob 1995

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*letter* |

### Marghoob 2007

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### Markowitz 2015

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*Does not report sufficient data for detection of melanoma* |

### McCarthy 1995

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*leaflet* |

### McMullan 1956

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data |

### Menzies 2008

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*evaluates dermoscopy alone* |

### Menzies 2011

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*surveillance study; data used to id factors predictive of lesion changes* |

### Menzies 2013

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*evaluates dermoscopy only* |

### *Moffatt 2006*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*'clinical diagnosis'* |

### *Mohammad 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*only includes BCC* |

### *Morrison 2001*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Study gives % correct diagnosis within each histology group and then gives the % 'correct' diagnosis of skin cancer as 22% for FP and 87% for dermatologist. But these statistics appear to have been reached by taking the mean of the % correct diagnoses across the malignant groups and do not equate to sensitivity. i.e. If you take the mean of the FP correct (%) for the 4 malignant groups you get: (40+22+25+0)/4 = 21.75% and then the same for the dermatologist correct (%) column: (95+77+75+100)/4=86.75%* |

### *Nachbar 1994*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Data for visual inspection alone influenced by use of dermoscopy in most cases* |

### *Nathansohn 2007*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Not test accuracy; follow-up study* |

### *Nilles 1994*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Does not provide data for visual inspection alone* |

### *Osborne 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*not clear what the ref standard is*<br>EXCLUDE on 2x2 data |

### *Osborne 1999*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Only patients with melanoma included* |

### *Parslew 1997*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Not all suspected of skin cancer* |

### *Pazzini 1996*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data |

### *Perednia 1992*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Not test accuracy* |

### *Perrinaud 2007*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Does not provide data for visual inspection alone* |

### *Piccolo 2000*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*No data can be extracted for visual inspection alone* |

### *Piccolo 2002*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>EXCLUDE on 2x2 data<br>*not enough data to populate 2x2 table. No breakdown of index test results and ref standard.* |

### *Pizzichetta 2001*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Observer agreement only* |

### *Provost 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Not test accuracy; only reports concordance* |

### *Quereux 2011*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*self-administered questions to patients attending a GP surgery before their appointment to determine whether they are at high risk of melanoma--which is meant to highlight to the GP which patient to examine during their consultation* |

### *Rallan 2006*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*No data can be extracted for visual inspection alone* |

### *Rampen 1988*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Only melanoma included* |

### *Reeck 1999*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*Only includes index test negatives; i.e. those considered benign by referring clinician*<br>EXCLUDE on target condition |

### *Riddell 1961*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*All malignant* |

### *Rigel 1993*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### *Robati 2014*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*no follow-up of patients not referred to dermatology clinics, who did not recieve histopathology* |

### *Robinson 2010*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*self examination* |

### *Rosado 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*Systematic Review* |

### *Rossi 2000*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*Unclear reference standard in disease negative* |

### *Roush 1986*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*only dysplastic nevus* |

### *Salvio 2011*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>EXCLUDE on sample size |

### *Schindewolf 1994*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*evaluates CAD not VI* |

### *Schmoeckel 1987*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study |

### *Schwartzberg 2005*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*Does not provide sufficient data for detection of melanoma*<br>INCLUDE based on full report |

### *Seidenari 2006*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*assessing best means of follow-in up patients with previous melanoma - total body exam versus only lesions >2cm. No melanoma identified* |

### *Seidenari 2006a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE if individual lesion characteristics<br>*looks like this study is only looking at asymmetry judgement* |

### *Shariff 2010*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard |

### *Sondak 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*comment paper* |

### *Soyer 2004*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Does not provide data for visual inspection alone* |

### *Stanganelli 1998a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*can't derive specificity; only gives 'exact diagnoses for MM and 2 benign categories and not number benign misdiagnosed as MM* |

### *Stanley 2003*

| | |
|---|---|
| Reason for exclusion | EXCLUDE if individual lesion characteristics<br>*fuzzy histogram is based on the lesion's colour, which is an individual lesion characteristic* |

### *Stathopoulos 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*only includes index test positive patients, i.e. no FN or TN results* |

### *Stratigos 2007*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>EXCLUDE on 2x2 data |

### *Tandjung 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*'Malignant' includes: AK, Bowen's, dysplastic nevus, lentigo maligna, SCC, BCC, MM, keratoacanthoma*<br>EXCLUDE on index test<br>*GPs sent images for telederm opinion; then free to send for biopsy or not; results shown are only for those that wer biopsied, according to TD advice* |

### *Terrill 2009*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Whole body skin examination after patients referred on for further assessment by a specialist*<br>EXCLUDE on 2x2 data |

### *Terushkin 2010*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Not test accuracy - reports final diagnoses of those excised over a number of time periods and benign-malignant ratio* |

### *Terushkin 2010a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Not test accuracy - reports final diagnoses of those excised over a number of time periods and benign-malignant ratio* |

### *Thomson 2005*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*letter* |

### *Torrey 1941*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*includes non-cutaneous lesions* |

### *Ulrich 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on target condition<br>*Does not provide sufficient data for evaluation of melanoma* |

### *van der Rhee 2010*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*<50% of disease negative have an adequate reference standard* |

### *van der Rhee 2011*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on sample size<br>*<5 cases* |

### *Vasili 2010*

| | |
|---|---|
| Reason for exclusion | EXCLUDE conference abstract |

### *Wagner 1985*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data |

### *Walter 2010*

| | |
|---|---|
| Reason for exclusion | EXCLUDE not a primary study<br>*clinical trial protocol* |

### *Walter 2013*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on reference standard<br>*Final diagnosis reached by histology or expert opinion; no FU of non-excised lesions reported in this paper. The Walter 2012 trial report does report follow-up for enough benign lesions for control arm (weighted 7PCL) data to be included. Authors contacted and confirmed calculations (02/03/16).* |

### *Warshaw 2009*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*Study presents diagnostic accuracy of teledermatology and clinic diagnosis in comparison to histopathology; in order to include in our review, data would need to be presented as a 2x2 contingency table, either per type of malignancy e.g. tele-dx classification of melanoma vs not melanoma against histological diagnosis of melanoma/not melanoma, or with malignant diagnoses grouped together, ie tele-dx of malignancy vs not malignant against same histological breakdown.*<br>EXCLUDE but contact authors<br>*Authors contacted: the 2x2 table the Cochrane researchers want to create is not possible for our results, because sensitivity and PPV are based on different sample sizes. This can be seen in Table 2 of the paper which actually adds up to 11870 skin lesions across, as for each histological diagnosis of interest the first lesion with such a histological diagnosis was considered per patient. Hence, a patient might appear several times across the columns. Table 1 adds up to 8585 skin lesions – the first skin lesion in the data set per patient with a clinical diagnosis.* |

### *Warshaw 2009a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*As per Warshaw 2009* |

### *Warshaw 2010*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data<br>*As per Warshaw 2009 ; this 2010 paper presents combined data for pigmented and nonpigmented lesions* |

### *Westbrook 2006*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data |

### *Whitaker-Worth 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>EXCLUDE on test observer<br>*mixed medical student/clinicians*<br>EXCLUDE on 2x2 data<br>*not test accuracy study* |

### *Whited 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on sample size |

### *Williams 1991*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on 2x2 data |

### *Winkelmann 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE duplicate or related publication |

### *Winkelmann 2015a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE duplicate or related publication |

### *Wolf 1998*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*clinical diagnosis study; Test clearly described - "concerning the clinical diagnosis, we were not able to ascertain from the clinical data sheet whether the referring physicians used additional diagnostics techniques such as dermoscopy"* |

### *Yoo 2015*

| | |
|---|---|
| Reason for exclusion | EXCLUDE conference abstract |

### *Youl 2007*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*'clinical diagnosis' - dermoscopy used in some but not all cases.*<br>EXCLUDE but contact authors<br>*Response from author "One of the main issues is that we just don't know to what extent dermoscopy was used in that study. We just asked where they used it in a general sense and not for each case. However for each case GPs and skin clinic doctors did indicate whether they conducted a whole- or part-body skin examination (or just lesion specific)* |

### *Youl 2007a*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on index test<br>*Evaluates clinical diagnosis (some lesions had dermoscopy)* |

### *Zaballos 2013*

| | |
|---|---|
| Reason for exclusion | EXCLUDE on study population<br>*They do not have enough benign cases to include as full report.* |

### *Zou 2001*

| | |
|---|---|
| **Reason for exclusion** | EXCLUDE not a primary study<br>*Study uses results from* Stolz 1994<br>EXCLUDE on 2x2 data<br>*Just showing ROC curves* |

*Footnotes*

AK - actinic keratosis; AUC – area under the curve; BCC - basal cell carcinoma; CAD – computer assisted diagnosis; D+/D- - disease positive/disease negative; Dx – diagnosis; 7FFM – seven features for melanoma; FPs – false positives; FN – false negative; FU – follow up; GP – general practitioner; PCP - primary care provider; PPV – positive predictive value; MM - – malignant melanoma; NPV – negative predictive value; SCC - squamous cell carcinoma; se/sp – sensitivity/specificity; TD – teledermatology; TN – true negative; TWR – two week rule; VI - visual inspection.

### Characteristics of studies awaiting classification

*Footnotes*

### Characteristics of ongoing studies

*Footnotes*

## Summary of results tables

### 1 Summary of findings table

| | | | | |
|---|---|---|---|---|
| **Question:** | **What is the diagnostic accuracy of visual inspection for the detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants in adults?** | | | |
| **Population:** | Adults with lesions suspicious for melanoma, including:<br>• those with limited prior testing (presenting in primary, community or private dermatology settings), and<br>• referred populations (presenting in secondary care or specialist skin cancer clinics). | | | |
| **Index test:** | Visual inspection with or without the use of any established algorithms or checklist to aid diagnosis, including:<br>• in-person evaluations (face-to-face diagnosis), and<br>• image-based evaluations (diagnosis based on assessment of a clinical image). | | | |
| **Target condition:** | Cutaneous invasive melanoma and atypical intraepidermal melanocytic variants | | | |
| **Reference standard:** | Histology with or without long term follow-up | | | |
| **Action:** | If accurate, positive results ensure melanoma lesions are not missed but are appropriately referred and excised and those with negative results can be safely reassured and discharged. | | | |
| | | **Number of studies** | **Total lesions** | **Total cases** |
| **Quantity of evidence** | | 49* | 34351 | 2499 |
| **Limitations** | | | | |
| **Risk of bias:** | Potential risk for patient selection from case control design (6), inappropriate exclusion criteria (7) or lack of detail (27/49). All index test interpretation was blinded to reference standard diagnosis. Index test thresholds not clearly pre-specified (22/28 in-person evaluations; 13/16 image-based). Low risk for reference standard (42/49); high concern from use of expert diagnosis (6). Blinding of reference standard to visual inspection diagnosis not reported in any study. High risk for participant flow due to differential verification (11), and exclusions following recruitment (15); timing of tests was not mentioned in 37. | | | |
| **Applicability of evidence to question:** | Participants restricted to those with melanocytic lesions only (10), or to those with histopathology results (37) and included multiple lesions per participant (14). No description of diagnostic thresholds (24 in-person; 13 image-based) or reporting of average or consensus diagnoses (7 in-person; 13 image-based). Clinical images interpreted blinded to clinical information (11/16). Little information given concerning the expertise of the histopathologist (40/49). | | | |
| **FINDINGS:** | | | | |

| Question: | What is the diagnostic accuracy of visual inspection for the detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants in adults? |
|---|---|

Thirty-seven studies (providing 39 datasets) reporting accuracy data for the primary target condition were separated *a priori* into in-person (n=28) and image-based (n=11) evaluations. Subsequent analysis confirmed differences in accuracy according to the different approaches to diagnosis (P<0.001). Attempts to analyse studies by degree of prior testing were hampered by a lack of relevant information provided in the study publications and by the inclusion of lesions selected for biopsy or excision. Of the 28 in-person evaluations, only 17 could be clearly placed on the clinical pathway, and 11 were considered not to have provided sufficient information to allow the pathway to be identified (coded 'unclear' on pathway). The findings presented are based on results for in-person evaluations that could be clearly placed on the clinical pathway.

| Test: | In-person visual inspection using any or no algorithm at any threshold |
|---|---|

| Data: | Number of datasets | Total lesions | Total melanomas |
|---|---|---|---|
| All in-person evaluations | 28 | 25604 | 1748 |
| Studies clearly placed on the clinical pathway | 17 | 14700 | 622 |

**Place on pathway: Participants with limited prior testing (all lesions)**

| Datasets (n) | Lesions (n) | Melanomas (n) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|
| 3 | 1339 | 55 | 92% (26, 100) | 80% (74, 85) |

| Numbers in a cohort of 1000 lesions** | TP | FP | FN | TN | PPV | NPV |
|---|---|---|---|---|---|---|
| At a prevalence of 4% | 37 [10; 40] | 195 [252; 147] | 3 [30; 0] | 765 [708; 813] | 16% [4; 21] | 100% [96; 100] |
| At a prevalence of 9% | 83 [24; 90] | 185 [239; 139] | 7 [66; 0] | 725 [671; 771] | 31% [9; 39] | 99% [91; 100] |
| At a prevalence of 16% | 148 [42; 160] | 171 [221; 129] | 12 [118; 0] | 669 [619; 711] | 46% [16; 55] | 98% [84; 100] |

**Place on pathway: Participants with limited prior testing (only lesions selected for excision)**

| Datasets (n) | Lesions (n) | Melanomas (n) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|
| 2 | 4228 | 160 | 90% (70, 97) | 81% (67, 90) |

| Numbers in a cohort of 1000 lesions** | TP | FP | FN | TN | PPV | NPV |
|---|---|---|---|---|---|---|
| At a prevalence of 4% | 36 [28; 39] | 180 [312; 96] | 4 [12; 1] | 780 [648; 864] | 17% [8; 29] | 99% [98; 100] |
| At a prevalence of 9% | 81 [63; 88] | 170 [296; 91] | 9 [27; 2] | 740 [614; 819] | 32% [18; 49] | 99% [96; 100] |
| At a prevalence of 16% | 144 [112; 156] | 157 [273; 84] | 16 [48; 4] | 683 [567; 756] | 48% [29; 65] | 98% [92; 99] |

**Place on pathway: Referred participants (all lesions)**

| Datasets (n) | Lesions (n) | Melanomas (n) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|
| 2 | 3494 | 61 | 75% (49, 90) | 99% (95, 100) |

| Numbers in a cohort of 1000 lesions** | TP | FP | FN | TN | PPV | NPV |
|---|---|---|---|---|---|---|
| At a prevalence of 4% | 30 [20; 36] | 13 [51; 4] | 10 [20; 4] | 947 [909; 956] | 69% [28; 90] | 99% [98; 100] |

| Question: | What is the diagnostic accuracy of visual inspection for the detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants in adults? | | | | | |
|---|---|---|---|---|---|---|
| At a prevalence of 9% | 67 [44; 81] | 13 [48; 4] | 23 [46; 9] | 897 [862; 906] | 84% [48; 96] | 98% [95; 99] |
| At a prevalence of 16% | 119 [78; 144] | 12 [45; 3] | 41 [82; 16] | 828 [795; 837] | 91% [64; 98] | 95% [91; 98] |
| **Referred participants (only lesions selected for excision)** | | | | | | |
| **Datasets (n)** | **Lesions (n)** | **Melanomas (n)** | **Sensitivity (95% CI)** | | **Specificity (95% CI)** | |
| 8 | 5331 | 258 | 77% (62, 87) | | 96% (90, 98) | |
| **Numbers in a cohort of 1000 lesions**** | **TP** | **FP** | **FN** | **TN** | **PPV** | **NPV** |
| At a prevalence of 4% | 31 [25; 35] | 41 [99; 16] | 9 [15; 5] | 919 [861; 944] | 43% [20; 68] | 99% [98; 99] |
| At a prevalence of 9% | 69 [56; 78] | 39 [94; 15] | 21 [34; 12] | 871 [816; 895] | 64% [37; 84] | 98% [96; 99] |
| At a prevalence of 16% | 123 [99; 139] | 36 [87; 14] | 37 [61; 21] | 804 [753; 826] | 77% [53; 91] | 96% [92; 98] |
| **Referred participants with equivocal lesions (only lesions selected for excision)** | | | | | | |
| **Datasets (n)** | **Lesions (n)** | **Melanomas (n)** | **Sensitivity (95% CI)** | | **Specificity (95% CI)** | |
| 2 | 930 | 88 | 85% (56, 96) | | 89% (79, 95) | |
| **Numbers in a cohort of 1000 lesions**** | **TP** | **FP** | **FN** | **TN** | **PPV** | **NPV** |
| At a prevalence of 4% | 34 [22; 38] | 101 [197; 48] | 6 [18; 2] | 859 [763; 912] | 25% [10; 44] | 99% [98; 100] |
| At a prevalence of 9% | 76 [50; 86] | 96 [187; 46] | 14 [40; 4] | 814 [723; 865] | 44% [21; 66] | 98% [95; 100] |
| At a prevalence of 16% | 136 [89; 154] | 88 [172; 42] | 24 [71; 6] | 752 [668; 798] | 61% [34; 79] | 97% [90; 99] |

*Footnotes*

*37 of the 49 included studies (reporting on 39 cohorts of lesions) provide data for the primary target condition (defined as detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants) and are the main focus of this 'Summary of findings' table; the summary of methodological quality is based on the full sample of 49 studies.

** Number of true positives (TP), false positives (FP), false negatives (FN), true negatives (TN) for a hypothetical cohort of 1000 lesions have been estimated at the median and interquartile ranges of prevalence (25th and 75th percentiles), at average sensitivity and specificity and using the lower and upper limits of the 95% confidence intervals, denoted in square brackets [lower limit; upper limit].

# Additional tables

## 1 Primary analyses for detection of invasive melanoma or melanocytic intraepidermal variants by position on the clinical pathway

| a. In-person evaluations (n=28) | | | | | | |
|---|---|---|---|---|---|---|
| **Position on pathway** | **Datasets** | **Lesions (cancers)** | **Sensitivity (95% CI)** | **Variance** | **Specificity (95% CI)** | **Variance** |

| a. In-person evaluations (n=28) | | | | | | |
|---|---|---|---|---|---|---|
| **Position on pathway** | **Datasets** | **Lesions (cancers)** | **Sensitivity (95% CI)** | **Variance** | **Specificity (95% CI)** | **Variance** |
| **Participants with limited prior testing (unselected on reference standard)** | | | | | | |
| **Clear** | **3** | **1339 (55)** | **92.4 (26.2, 99.8)** | **6.26** | **79.7 (73.7, 84.7)** | **0.07** |
| **Participants with limited prior testing (selected for excision)** | | | | | | |
| **Clear** | **2 (ind)** | **4228 (160)** | **90.1 (70.0, 97.3)** | **0.53** | **81.3 (67.5, 90.0)** | **0.25** |
| Unclear | 1 | 353 (38) | 78.9 (62.7, 90.4) | - | 94.0 (90.7, 96.3) | - |
| Combined | 3 | 4581 (198) | 87.2 (73.2, 94.4) | 0.45 | 87.1 (74.6, 94.0) | 0.51 |
| **Referred participants (unselected on reference standard)** | | | | | | |
| **Clear** | **2** | **3494 (61)** | **74.6 (48.9,90.0)** | **0.14** | **98.6 (94.7,99.6)** | **0.77** |
| **Referred participants (selected for excision)** | | | | | | |
| **Clear** | **8** | **5331 (258)** | **76.7 (61.7, 87.1)** | **0.78** | **95.7 (89.7, 98.3)** | **1.73** |
| Unclear | 9 | 9611 (1015) | 82.8 (74.4, 88.9) | 0.34 | 89.2 (71.1, 96.5) | 3.21 |
| Combined | 17 | 14942 (1273) | 79.7 (71.7, 85.8) | 0.59 | 93.0 (85.4, 96.8) | 2.59 |
| **Referred participants with equivocal lesions (selected for excision)** | | | | | | |
| **Clear** | **2 (ind)** | **930 (88)** | **84.7 (55.5, 96.1)** | **0.93** | **89.5 (79.5, 95.0)** | **0.27** |
| Unclear | 1 | 318 (73) | 61.4 (49.0, 72.9) | - | 87.3 (82.5, 91.2) | - |
| Combined | 3 | 1248 (161) | 76.4 (48.4, 91.8) | 1.03 | 88.8 (81.8, 93.3) | 0.21 |
| **b. Image-based evaluations (n=11)** | | | | | | |
| **Position on pathway** | **Datasets** | **Lesions (cancers)** | **Sensitivity (95% CI)** | **Variance** | **Specificity (95% CI)** | **Variance** |
| **Participants with limited prior testing (selected for excision)** | | | | | | |
| **Clear** | **1** | **50 (9)** | **22.2 (2.8, 60.0)** | **-** | **70.7 (54.4, 83.9)** | **-** |
| Unclear | 1 | 463 (29) | 20.7 (8.0, 39.7) | - | 96.8 (94.6, 98.2) | - |
| Combined | 2 | 513 (38) | 21.4 (10.0, 40.1) | 0 | 90.9 (60.7, 98.1) | 1.50 |
| **Referred participants (unselected on reference standard)** | | | | | | |
| **Clear** | **1** | **134 (31)** | **74.2 (55.4, 88.1)** | **-** | **82.5 (73.8, 89.3)** | **1** |
| **Referred participants (selected for excision)** | | | | | | |

| a. In-person evaluations (n=28) | | | | | | |
|---|---|---|---|---|---|---|
| Position on pathway | Datasets | Lesions (cancers) | Sensitivity (95% CI) | Variance | Specificity (95% CI) | Variance |
| Unclear | 6 | 293 (96) | 60.3 (49.2, 70.5) | 0.02 | 77.0 (63.9, 86.4) | 0.40 |
| Referred participants with equivocal lesions (selected for excision) | | | | | | |
| Unclear | 2 | 303 (98) | 61.9 (46.7, 75.0) | 0.10 | 81.8 (75.2, 87.0) | 0.01 |

*Footnotes*

ind - sensitivity and specificity estimated 'independently' in separate models due to sparse data.

## 2 Secondary analyses for primary target condition by covariate

| Subgroup | Datasets | Lesions (cancers) | Diagnostic odds ratio (DOR) (95% CI) | Relative DOR (95% CI) | P value (DOR) | P value [1] (Hierarchical summary receiver-operator curves (HSROC) models) |
|---|---|---|---|---|---|---|
| **Differences in-person and image based** | | | | | | |
| In-person | 28 | 25604 (1748) | 37.5 (21.7, 64.7) | 8.54 (2.89, 25.3) | <0.001 | 0.001 |
| Image-based | 11 | 1243 (263) | 4.38 (1.79, 10.8) | | | |
| **Analyses based on in-person evaluations only (n=28):** | | | | | | |
| **Study setting** | | | | | | |
| Primary/Community/Private | 6 | 5920 (253) | 27.6 (6.95, 109) | | | |
| Secondary | 10 | 10419 (1019) | 39.0 (13.8, 110) | | | |
| Specialist clinic | 12 | 9265 (476) | 44.4 (17.2, 115) | Secondary/specialist vs. primary[2]: 1.51 (0.32, 7.09) | 0.59 | 0.62 |
| **Use of a diagnostic algorithm** | | | | | | |
| No algorithm used | 21 | 19330 (1076) | 37.3 (18.0, 77.3) | | | |
| Any algorithm used | 7 | 6274 (672) | 38.5 (11.3, 132) | 1.03 (0.25, 4.34) | 0.96 | 0.55 |
| **Type of reference standard used** | | | | | | |
| Histology alone | 22 | 20783 (1627) | 39.1 (19.7, 77.8) | | | |
| Histology plus any other | 6 | 4821 (121) | 29.7 (6.60, 134) | 0.76 (0.14, 4.02) | 0.74 | 0.68 |
| **Prevalence** | | | | | | |
| Prevalence ≤0.1 | 16 | 21907 (811) | 63.7 (28.6, 142) | | | |
| Prevalence>0.1 | 12 | 3697 (937) | 19.6 (8.39, 45.8) | 0.31 (0.09, 1.00) | 0.05 | 0.06 |

*Footnotes*

1 Likelihood ratio test assessing differences in both accuracy and threshold

2 Secondary vs Primary 1.41 (0.25, 7.93), P=0.68; Specialist vs Primary 1.61 (0.30, 8.63), P=0.56; Specialist vs Secondary 1.14 (0.28, 4.68), P=0.85.

## 3 Visual inspection for detection of melanoma and intraepidermal melanocytic variants - by algorithm

| Test (threshold) | Datasets | Lesions (melanomas) | Pooled Sensitivity (95% CI) | Pooled Specificity (95% CI) | Diagnostic odds ratio (DOR) (95% CI) |
|---|---|---|---|---|---|
| **In-person evaluations** | | | | | |
| No algorithm | 21 | 19330 (1076) | 0.78 (0.68, 0.85) | 0.93 (0.88, 0.96) | 46.2 (21.9, 97.5) |
| (A)BCD(E)* | 6 *** | 5501 (654) | 0.83 (0.75, 0.88) | 0.88 (0.64, 0.97) | 36.6 (7.94, 168) |
| 7point checklist at >=2 | 1 | 205 (12) | 0.92 (0.62, 1.00) | 0.65 (0.58, 0.72) | 22.8 (2.08, 176) |
| 7point checklist at >=3 | 1 | 205 (12) | 0.42 (0.15, 0.72) | 0.93 (0.89, 0.96) | 11.8 (3.22, 43.3) |
| 7point checklist at >=4 | 1 | 205 (12) | 0.25 (0.07, 0.57) | 0.98 (0.96, 1.00) | 31.8 (4.71, 215) |
| 7point checklist (revised) at >=3 | 1 | 773 (18) | 0.94 (0.73, 1.00) | 0.80 (0.77, 0.83) | |
| new Collas at >=1 | 1 | 353 (38) | 0.76 (0.60, 0.89) | 0.50 (0.44, 0.56) | 3.24 (1.49, 7.07) |
| **Image-based evaluations** | | | | | |
| No algorithm | 9 | 1090 (217) | 0.58 (0.43, 0.71) | 0.84 (0.76, 0.90) | 7.47 (4.12, 13.5) |
| ABCD(E)** | 2 | 153 (46) | 0.53 (0.37, 0.70) | 0.71 (0.45, 0.88) | 2.87 (0.93, 8.79) |

*Footnotes*

* Combines data from studies using ABCD with threshold not reported (n=2), ABCDE with at least 2 characteristics present (n=3) and BCD with at least 2 characteristics present (n=1)

** Combines data from studies using ABCD with at least 2 characteristics present (n=1) and ABCDE with at least 2 characteristics present (n=1)

*** Due to non-convergence, the bivariate models were fitted assuming zero correlation between the logit sensitivity and logit specificity and removing the random effects term for specificity when estimating sensitivity and the random effects term for sensitivity when estimating specificity.

## 4 Secondary analyses for detection of melanoma and intraepidermal melanocytic variants by observer

| Subgroup | Datasets | Lesions (melanomas) | Diagnostic odds ratio (DOR) (95% CI) | Relative DOR (RDOR) (95% CI) | P value (for RDOR) | P value* (Hierarchical summary receiver-operator curves (HSROC) models) |
|---|---|---|---|---|---|---|
| **In-person evaluations** | | | | | | |
| Expert consultant | 9 | 3547 | 29.0 (11.0, 76.2) | 1 | | 0.36 |
| Consultant | 13 | 16858 | 38.4 (16.9, 87.6) | 1.32 (0.37, 4.71) | 0.65 | |
| Resident/registrar | 2 | 1339 | 12.9 (1.99, 84.0) | 0.45 (0.05, 3.67) | 0.44 | |
| Mixed (secondary care) | 2 | 2704 | 48.0 (4.54, 507) | 1.65 (0.13, 21.4) | 0.69 | |
| GP | 3 | 1236 | 211 (24.9, 1788) | 7.28 (0.69, 76.3) | 0.09 | |
| **Image-based evaluations** | | | | | | |
| Expert consultant | 6 | 974 | 20.5 (4.82, 86.9) | 1 | | 0.22 |
| Consultant | 4 | 200 | 3.76 (1.15, 12.3) | 0.18 (0.04, 0.90) | 0.04 | |
| Mixed (secondary care) | 1 | 200 | 10.9 (2.02, 59.2) | 0.53 (0.07, 3.97) | 0.50 | |
| Mixed (secondary/primary care) | 1 | 40 | 11.5 (0.94, 142) | 0.56 (0.04, 7.51) | 0.63 | |
| Mixed (primary care) | 2 | 184 | 6.60 (1.73, 25.2) | 0.32 (0.07, 1.40) | 0.11 | |

*Footnotes*

* Likelihood ratio test assessing differences in both accuracy and threshold

## 5 Results for studies reporting data for more than one observer

| Study<br>Algorithm (diagnostic approach)<br>Dis/Non-dis*; prevalence | Observer qualification | Sensitivity (95% CIs) | Specificity (95% CIs) | Observer qualification | Sensitivity (95% CIs) | Specificity (95% CIs) | Observer qualification | Sensitivity (95% CIs) | Specificity (95% CIs) |
|---|---|---|---|---|---|---|---|---|---|
| **Target condition: Invasive melanoma and/or atypical intraepidermal melanocytic variants** | | | | | | | | | |
| Benelli 2001<br>ABCDE (i-b)<br>12 / 38; 24% | | | | Dermatologist (n=65) | 50% (21, 79) | 50% (33, 67) | Expert dermatologists (n=1) | 58% (28, 85) | 53% (36, 69) |
| Morton 1998<br>No algorithm (in-p)<br>Different lesions per obs | Registrar (n=6)<br>69 / 694; 9% | 79% (59, 92) | 98% (97, 99) | Senior registrar (n=2)<br>31 / 536; 5% | 90% (74, 98) | 97% (96, 99) | Expert dermatologists (n=2)<br>28 / 641; 4% | 91% (82, 97) | 99% (97, 99) |
| Stanganelli 2005<br>No algorithm (i-b)<br>31 / 103; 23% | GP (n=3) | 81% (63, 93) | 73% (63, 81) | | | | Experienced dermatologists (n=3) | 74% (55, 88) | 83% (74, 89) |
| **Target condition: Invasive melanoma alone** | | | | | | | | | |
| Lorentzen 1999<br>No algorithm (i-b)<br>49 / 183; 21% | | | | Non-expert dermatology residents (n=5) | 61% (46, 75) | 88% (82, 92) | Experienced dermatologists (n=4) | 78% (63, 88) | 89% (84, 93) |
| Rao 1997<br>ABCD (i-b)<br>21 / 51; 29% | | | | Melanoma Fellow 1 (n=1) | 90% (70, 99) | 80% (67, 90) | Dermatologist 1 (n=1) | 76% (53, 92) | 82% (69, 92) |
| | | | | Melanoma Fellow 2 (n=1) | 86% (64, 97) | 75% (60, 86) | Dermatologist 2 (n=1) | 86% (64, 97) | 75% (60, 86) |
| Scope 2008<br>Ugly Duckling (i-b)<br>5 / 140; 3% | Dermatology nurse + medical photographer (n=5) | 60% (15, 95) | 96% (91, 98) | General dermatologists (n=13) | 80% (28, 99) | 86% (79, 91) | Expert dermatologists (n=8) | 80% (28, 99) | 95% (90, 98) |

| Study<br>Algorithm (diagnostic approach)<br>Dis/Non-dis*; prevalence | Observer qualification | Sensitivity (95% CIs) | Specificity (95% CIs) | Observer qualification | Sensitivity (95% CIs) | Specificity (95% CIs) | Observer qualification | Sensitivity (95% CIs) | Specificity (95% CIs) |
|---|---|---|---|---|---|---|---|---|---|
| **Target condition: Invasive melanoma and/or atypical intraepidermal melanocytic variants** | | | | | | | | | |
| Westerhoff 2000<br>No algorithm (i-b)<br>50 / 50; 50% | GP pre-dermoscopy training (n=37) | 54% (39, 68) | 53% (38, 67) | GP post-dermoscopy training<br>(n=37) | 62% (47, 75) | 54% (39, 68) | | | |

*Footnotes*

in-p - in-person; i-b - image-based; obs - observer; GP - general practitioner

*Number of diseased / number of non-diseased (prevalence of disease), for each definition of the target condition

## 6 Secondary analyses for alternative definitions of the target condition

| Subgroup | Datasets | Participants (cases) | Diagnostic odds ratio (DOR) (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | Relative DOR (RDOR) (95% CI) | P value (RDOR) | P value* (Hierarchical summary receiver-operator curves (HSROC) models) |
|---|---|---|---|---|---|---|---|---|
| **Differences in-person and image-based** | | | | | | | | |
| **Detection of invasive melanoma alone** | | | | | | | | |
| In-person | 7 | 6857 (208) | 62.4 (17.6, 222) | 86% (68, 94) | 91% (81, 96) | 4.21 (0.62, 28.6) | 0.13 | 0.27 |
| Image-based | 5 | 599 (150) | 14.8 (3.56, 61.9) | 76% (50, 91) | 83% (62, 93) | | | |
| **Detection of any skin lesion requiring excision** | | | | | | | | |
| In-person | 7 | 8091 (2187) | 20.5 (7.11, 59.3) | 81% (68, 90) | 81% (56, 93) | 1.70 (0.24, 12.3) | 0.55 | 0.87 |
| Image-based | 3 | 547 (138) | 11.9 (2.22, 65.3) | 75% (49, 90) | 79% (38, 96) | | | |

*Footnotes*

* Likelihood ratio test assessing differences in both accuracy and threshold

## 7 Results for studies reporting data for more than one definition of the target condition

| | Detection of invasive melanoma | | | Detection of invasive melanoma or atypical intraepidermal melanocytic variants | | | Detection of any lesion requiring excision | | |
|---|---|---|---|---|---|---|---|---|---|
| Study author | Dis/ non-dis (prev)* | Sensitivity (95% CIs) | Specificity (95% CIs) | Dis/ non-dis (prev)* | Sensitivity (95% CIs) | Specificity (95% CIs) | Dis/ non-dis (prev)* | Sensitivity (95% CIs) | Specificity (95% CIs) |
| In-person | | | | | | | | | |
| Ek 2005 | - | - | - | 23 / 2559; 1% | 48% (27, 69) | 99% (99, 99) | 1754 /828; 68% | 98% (97, 98) | 13% (11, 15) |
| McGovern 1992 | 6 / 186; 3% | 100% (54, 100) | 89% (83, 93) | 11 / 181; 6% | 73% (39, 94) | 88% (83, 93) | 15 / 177; 8% | 73% (45, 92) | 88% (82, 93) |
| Stanganelli 2000 | - | - | - | 55 / 3317; 2% | 67% (53, 79) | 99% (99, 100) | 98 / 3274; 3% | 71% (61, 80) | 99% (99, 99) |
| Steiner 1987 | - | - | - | 73 / 245; 23% | 59% (47, 70) | 87% (83, 91) | 93 / 225; 29% | 67% (56, 76) | 86% (81, 90) |
| Walter 2012 | 16 / 757; 2% | 94% (70, 100) | 80% (77, 83) | 18 / 755; 2% | 94% (73, 100) | 80% (77, 83) | 22 / 751; 3% | 82% (60, 95) | 80% (77, 83) |
| Image-based | | | | | | | | | |
| Carli 2002b | - | - | - | 10 / 43; 19% | 80% (44, 97) | 84% (69, 93) | 20 / 34; 37% | 80% (56, 94) | 74% (56, 87) |
| Rosendahl 2011 | - | - | - | 29 / 434; 6% | 21% (08, 40) | 97% (95, 98) | 104 / 359; 22% | 76% (67, 84) | 85% (81, 88) |
| Stanganelli 1998 | - | - | - | 10 / 20; 33% | 40% (12, 74) | 75% (51, 91) | 14 / 16; 47% | 64% (35, 87) | 75% (48, 93) |

*Footnotes*

*Number of diseased / number of non-diseased (prevalence of disease), for each definition of the target condition

# References to studies

## Included studies


### *Argenziano 2006*

* Argenziano G, Puig S, Zalaudek I, Sera F, Corona R, Alsina M, et al. Dermoscopy improves accuracy of primary care physicians to triage lesions suggestive of skin cancer. Journal of Clinical Oncology 2006;24(12):1877-82. [Other: ER4:17940973; PubMed: 16622262]

### *Barzegari 2005*

* Barzegari M, Ghaninezhad H, Mansoori P, Taheri A, Naraghi ZS, Asgari M. Computer-aided dermoscopy for diagnosis of melanoma. BMC Dermatology 2005;5:8. [Other: ER4:15465860; PubMed: 16000171]

### *Benelli 1999*

* Benelli C, Roscetti E, Pozzo VD, Gasparini G, Cavicchini S. The dermoscopic versus the clinical diagnosis of melanoma. European Journal of Dermatology 1999;9(6):470-6. [Other: ER4:18375029; PubMed: 10491506]

### *Benelli 2001*

* Benelli C, Roscetti E, Dal Pozzo V. Reproducibility of the clinical criteria (ABCDE rule) and dermatoscopic features (7FFM) for the diagnosis of malignant melanoma. European Journal of Dermatology 2001;11(3):234-9. [Other: ER4:18375028; PubMed: 11358731]

### *Bono 1996*

* Bono A, Tomatis S, Bartoli C, Cascinelli N, Clemente C, Cupeta C, et al. The invisible colours of melanoma. A telespectrophotometric diagnostic approach on pigmented skin lesions. European Journal of Cancer 1996;32(4):727-9. [DOI: http://dx.doi.org/10.1016/0959-8049(95)00649-4; Other: ER4:20569437; PubMed: 8695280]

### *Bono 2002*

* Bono A, Bartoli C, Cascinelli N, Lualdi M, Maurichi A, Moglia D, et al. Melanoma detection. A prospective study comparing diagnosis with the naked eye, dermatoscopy and telespectrophotometry. Dermatology 2002;205(4):362-6. [Other: ER4:15465870; PubMed: 12444332]

### *Bono 2002b*

* Bono A, Bartoli C, Baldi M, Tomatis S, Bifulco C, Santinami M. Clinical and dermatoscopic diagnosis of small pigmented skin lesions. European Journal of Dermatology 2002;12(6):573-6. [Other: ER4:18375034; PubMed: 12459531]

### *Bono 2006*

* Bono A, Tolomio E, Trincone S, Bartoli C, Tomatis S, Carbone A, et al. Micro-melanoma detection: a clinical study on 206 consecutive cases of pigmented skin lesions with a diameter < or = 3 mm. British Journal of Dermatology 2006; 155(3):570-3. [Other: ER4:15465872; PubMed: 16911283]

### *Bourne 2012*

* Bourne P, Rosendahl C, Keir J, Cameron A. BLINCK-A diagnostic algorithm for skin cancer diagnosis combining clinical features with dermatoscopy findings. Dermatology Practical & Conceptual 2012;2(2):202a12. [Other: ER4:17941081; PubMed: 23785600]

### *Carli 2002*

* Carli P, De Giorgi V, Argenziano G, Palli D, Giannotti B. Pre-operative diagnosis of pigmented skin lesions: in vivo dermoscopy performs better than dermoscopy on photographic images. Journal of the European Academy of Dermatology & Venereology 2002;16(4):339-46. [Other: ER4:15465882; PubMed: 12224689]

### *Carli 2002b*

* Carli P, de Giorgi V, Salvini C, Mannone F, Chiarugi A. The gold standard for photographing pigmented skin lesions for diagnostic purposes: contact versus distant imaging. Skin Research & Technology 2002;8(4):255-9. [Other: ER4:15465888; PubMed: 12423545]

### *Carli 2003b*

* Carli P, De Giorgi Vincenzo, Chiarugi A, Nardini P, Mannone F, Stante M, et al. Effect of lesion size on the diagnostic performance of dermoscopy in melanoma detection. Dermatology 2003;206(4):292-6. [Other: ER4:15465883; PubMed: 12771468]

### *Chang 2013*

* Chang WY, Huang A, Yang CY, Lee CH, Chen YC, Wu TY, et al. Computer-aided diagnosis of skin lesions using conventional digital photography: a reliability and feasibility study. PloS One 2013;8(11):e76212. [Other: ER4:15465893; PubMed: 24223698]

### *Collas 1999*

* Collas H, Delbarre M, De Preville P-A, Courville Ph, Neveu C, Dompmartin A, et al. Evaluation of the diagnosis of pigmented tumors of the skin and factors leading to a decision to excise [Evaluation du diagnostic des tumeurs pigmentées de la peau et des éléments conduisant à une décision d'exérèse]. Annales de dermatologie et de vénéréologie 1999;126(6-7):494-500. [Other: ER4:21450600; PubMed: 10495858]

### *Cristofolini 1994*

* Cristofolini M, Zumiani G, Bauer P, Cristofolini P, Boi S, Micciolo R. Dermatoscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. Melanoma Research 1994;4(6):391-4. [Other: ER4:15465898; PubMed: 7703719]

### *Cristofolini 1997*

* Cristofolini M, Bauer P, Boi S, Cristofolini P, Micciolo R, Sicher MC. Diagnosis of cutaneous melanoma: Accuracy of a computerized image analysis system (Skin View). Skin Research and Technology 1997;3(1):23-27. [Other: ER4:17941039; PubMed: 27333169]

### *de Giorgi 2012*

* de Giorgi V, Savarese I, Rossari S, Gori A, Grazzini M, Crocetti E, et al. Features of small melanocytic lesions: does small mean benign? A clinical-dermoscopic study. Melanoma Research 2012;22(3):252-6. [Other: ER4:18375042; PubMed: 22430838]

### *Dolianitis 2005*

* Dolianitis C, Kelly J, Wolfe R, Simpson P. Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesions. Archives of Dermatology 2005;141(8):1008-14. [Other: ER4:15465906; PubMed: 16103330]

### *Dummer 1993*

* Dummer W, Doehnel KA, Remy W. Videomicroscopy in differential diagnosis of skin tumors and secondary prevention of malignant melanoma. Hautarzt 1993;44(12):772-6. [Other: ER4:18375044; PubMed: 8113040]

### *Ek 2005*

* Ek EW, Giorlando F, Su SY, Dieu T. Clinical diagnosis of skin tumours: how good are we? ANZ Journal of Surgery 2005; 75(6):415-20. [DOI: 10.1111/j.1445-2197.2005.03394.x; Other: ER4:20569451; PubMed: 15943729]

### *Gachon 2005*

* Gachon J, Beaulieu P, Sei JF, Gouvernet J, Claudel JP, Lemaitre M, et al. First prospective study of the recognition process of melanoma in dermatological practice. Archives of Dermatology 2005;141(4):434-8. [Other: ER4:15465924; PubMed: 15837860]

### *Green 1991*

* Green A, Martin N, McKenzie G, Pfitzner J, Quintarelli F, Thomas BW, et al. Computer image analysis of pigmented skin lesions. Melanoma Research 1991;1(4):231-6. [Other: ER4:17941055; PubMed: 1823631]

### *Green 1994*

* Green A, Martin N, Pfitzner J, O'Rourke M, Knight N. Computer image analysis in the diagnosis of melanoma. Journal of the American Academy of Dermatology 1994;31(6):958-64. [Other: ER4:15465938; PubMed: 7962777]

### *Grimaldi 2009*

* Grimaldi L, Silvestri A, Brandi C, Nisi G, Brafa A, Calabro M, et al. Digital epiluminescence dermoscopy for pigmented cutaneous lesions, primary care physicians, and telediagnosis: a useful tool? Journal of Plastic, Reconstructive & Aesthetic Surgery: JPRAS 2009;62(8):1054-8. [Other: ER4:15465940; PubMed: 18547883]

### *Kopf 1975*

* Kopf AW, Mintzis M, Bart RS. Diagnostic accuracy in malignant melanoma. Archives of Dermatology 1975;111(10):1291-2. [DOI: 10.1001/archderm.1975.01630220055001; Other: ER4:21450617; PubMed: 1190800]

### *Krahn 1998*

* Krahn G, Gottlober P, Sander C, Peter RU. Dermatoscopy and high frequency sonography: two useful non-invasive methods to increase preoperative diagnostic accuracy in pigmented skin lesions. Pigment Cell Research 1998; 11(3):151-4. [Other: ER4:15465981; PubMed: 9730322]

### *Langley 2001*

* Langley RG, Rajadhyaksha M, Dwyer PJ, Sober AJ, Flotte TJ, Anderson RR. Confocal scanning laser microscopy of benign and malignant melanocytic skin lesions in vivo. Journal of the American Academy of Dermatology 2001; 45(3):365-376. [DOI: 10.1067/mjd.2001.117395; Other: ER4:20569473; PubMed: 11511832]

### *Lorentzen 1999*

* Lorentzen H, Weismann K, Petersen CS, Larsen FG, Secher L, Skodt V. Clinical and dermatoscopic diagnosis of malignant melanoma. Assessed by expert and non-expert groups. Acta Dermato-Venereologica 1999;79(4):301-4. [Other: ER4:17941062; PubMed: 10429989]

### *McGovern 1992*

* McGovern TW, Litaker MS. Clinical predictors of malignant pigmented lesions. A comparison of the Glasgow seven-point checklist and the American Cancer Society's ABCDs of pigmented lesions. Journal of Dermatologic Surgery & Oncology 1992;18(1):22-6. [Other: ER4:18375119; PubMed: 1740563]

### *Menzies 2009*

* Menzies SW, Emery J, Staples M, Davies S, McAvoy B, Fletcher J, et al. Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial. British Journal of Dermatology 2009;161(6):1270-7. [DOI: 10.1111/j.1365-2133.2009.09374.x; Other: ER4:15466005; PubMed: 19747359]

### *Morales Callaghan 2008*

* Morales Callaghan AM, Castrodeza-Sanz J, Martinez-Garcia G, Peral-Martinez I, Miranda-Romero A. Correlation between clinical, dermatoscopic, and histopathologic variables in atypical melanocytic nevi. Actas Dermo-Sifiliograficas 2008;99(5):380-9. [Other: ER4:17941068; PubMed: 18501170]

### *Morton 1998a*

* Morton CA, MacKie RM. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. British Journal of Dermatology 1998;138(2):283-7. [Other: ER4:20569481; PubMed: 9602875]

### *Morton 1998b*

Morton CA, MacKie RM. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. British Journal of Dermatology 1998;138(2):283-7. [ PubMed: 9602875]

### *Morton 1998c*

Morton CA, MacKie RM. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. British Journal of Dermatology 1998;138(2):283-7. [Other: ER4:20569481; PubMed: 9602875]

### *Pizzichetta 2004*

* Pizzichetta MA, Talamini R, Stanganelli I, Puddu P, Bono R, Argenziano G, et al. Amelanotic/hypomelanotic

melanoma: clinical and dermoscopic features. British Journal of Dermatology 2004;150(6):1117-24. [Other: ER4:15466066; PubMed: 15214897]

### *Rao 1997*

* Rao BK, Marghoob AA, Stolz W, Kopf AW, Slade J, Wasti Q, et al. Can early malignant melanoma be differentiated from atypical melanocytic nevi by in vivo techniques? Part I. Clinical and dermoscopic characteristics. Skin Research and Technology 1997;3(1):8-14. [Other: ER4:17941048; PubMed: 27333167]

### *Rosendahl 2011*

* Rosendahl C, Tschandl P, Cameron A, Kittler H. Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. Journal of the American Academy of Dermatology 2011;64(6):1068-73. [Other: ER4:15466083; PubMed: 21440329]

### *Scope 2008*

* Scope A, Dusza SW, Halpern AC, Rabinovitz H, Braun RP, Zalaudek I, et al. The "ugly duckling" sign: agreement between observers. Archives of Dermatology 2008;144(1):58-64. [Other: ER4:15465911; PubMed: 18209169]

### *Soyer 1995*

* Soyer HP, Smolle J, Leitinger G, Rieger E, Kerl H. Diagnostic reliability of dermoscopic criteria for detecting malignant melanoma. Dermatology 1995;190(1):25-30. [Other: ER4:18375054; PubMed: 7894091]

### *Stanganelli 1998*

* Stanganelli I, Serafini M, Cainelli T, Cristofolini M, Baldassari L, Staffa M, et al. Accuracy of epiluminescence microscopy among practical dermatologists: a study from the Emilia-Romagna region of Italy. Tumori 1998; 84(6):701-5. [Other: ER4:18375055; PubMed: 10080681]

### *Stanganelli 2000*

* Stanganelli I, Serafini M, Bucch L. A cancer-registry-assisted evaluation of the accuracy of digital epiluminescence microscopy associated with clinical examination of pigmented skin lesions. Dermatology 2000;200(1):11-6. [Other: ER4:15466129; PubMed: 10681607]

### *Stanganelli 2005*

* Stanganelli I, Brucale A, Calori L, Gori R, Lovato A, Magi S, et al. Computer-aided diagnosis of melanocytic lesions. Anticancer Research 2005;25(6C):4577-82. [Other: ER4:15466126; PubMed: 16334145]

### *Steiner 1987*

* Steiner A, Pehamberger H, Wolff K. In vivo epiluminescence microscopy of pigmented skin lesions. II. Diagnosis of small pigmented skin lesions and early detection of malignant melanoma. Journal of the American Academy of Dermatology 1987;17(4):584-91. [Other: ER4:17940992; PubMed: 3668003]

### *Thomas 1998*

* Thomas L, Tranchand P, Berard F, Secchi T, Colin C, Moulin G. Semiological value of ABCDE criteria in the diagnosis of cutaneous pigmented tumors. Dermatology 1998;197(1):11-7. [Other: ER4:15466141; PubMed: 9693179]

### *Troyanova 2003*

* Troyanova P. A beneficial effect of a short-term formal training course in epiluminescence microscopy on the diagnostic performance of dermatologists about cutaneous malignant melanoma. Skin Research & Technology 2003; 9(3):269-73. [Other: ER4:17941004; PubMed: 12877690]

### *Unlu 2014*

* Unlu E, Akay BN, Erdem C. Comparison of dermatoscopic diagnostic algorithms based on calculation: The ABCD rule of dermatoscopy, the seven-point checklist, the three-point checklist and the CASH algorithm in dermatoscopic evaluation of melanocytic lesions. Journal of Dermatology 2014;41(7):598-603. [Other: ER4:15466145; PubMed: 24807635]

### *Viglizzo 2004*

* Viglizzo G, Rongioletti F. Clinical, dermoscopic and pathologic correlation of pigmentary lesions observed in a dermoscopy service in the year 2003 [Correlazione clinico-dermoscopico-patologica delle lesioni cutanee pigmentate osservate in un servizio di dermoscopia nell'anno 2003]. Giornale Italiano di Dermatologia e Venereologia 2004; 139(4):339-344. [Other: ER4:18375099]

### *Walter 2012*

* Walter FM, Morris HC, Humphrys E, Hall PN, Prevost AT, Burrows N, et al. Effect of adding a diagnostic aid to best practice to manage suspicious pigmented lesions in primary care: randomised controlled trial. BMJ 2012;345:e4110. [Other: ER4:15466154; PubMed: 22763392]

### *Westerhoff 2000*

* Westerhoff K, McCarthy WH, Menzies SW. Increase in the sensitivity for melanoma diagnosis by primary care physicians using skin surface microscopy. British Journal of Dermatology 2000;143(5):1016-20. [Other: ER4:15466164;

PubMed: 11069512]

### *Winkelmann 2016*

* Winkelmann RR, Farberg AS, Tucker N, White R, Rigel DS. Enhancement of international dermatologists' pigmented skin lesion biopsy decisions following dermoscopy with subsequent integration of multispectral digital skin lesion analysis. Journal of Clinical and Aesthetic Dermatology 2016;9(7):53-55. [Other: ER4:25701735; PubMed: 27672411]

### *Zaumseil 1983*

* Zaumseil RP, Fiedler H, Gstöttner R. Clinico-diagnostic accuracy in malignant melanoma of the skin [Klinisch-diagnostische Treffsicherheit beim malignen Melanom der Haut]. Dermatologische Monatschrift 1983;169(2):101-105. [Other: ER4:21450660; PubMed: 6840366]

## Excluded studies

### *Abbasi 2004*

Abbasi NR, Shaw HM, Rigel DS, Friedman RJ, McCarthy WH, Osman I, et al. Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. JAMA 2004;292(22):2771-6.

### *Aldridge 2011*

Aldridge RB, Glodzik D, Ballerini L, Fisher RB, Rees JL. Utility of non-rule-based visual matching as a strategy to allow novices to achieve skin lesion diagnosis. Acta Dermato-Venereologica 2011;91(3):279-83.

### *Aldridge 2011a*

Aldridge RB, Zanotto M, Ballerini L, Fisher RB, Rees JL. Novice identification of melanoma: not quite as straightforward as the ABCDs. Acta Dermato-Venereologica 2011;91(2):125-30.

### *Aldridge 2013*

Aldridge RB, Naysmith L, Ooi ET, Murray CS, Rees JL. The importance of a full clinical examination: assessment of index lesions referred to a skin cancer clinic without a total body skin examination would miss one in three melanomas. Acta Dermato-Venereologica 2013;93(6):689-92.

### *Alendar 2009*

Alendar F, Drljevic I, Drljevic K, Alendar T. Early detection of melanoma skin cancer. Bosnian Journal of Basic Medical Sciences 2009;9(1):77-80.

### *Argenziano 1999*

Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Delfino M. Clinical and dermatoscopic criteria for the preoperative evaluation of cutaneous melanoma thickness. Journal of the American Academy of Dermatology 1999;40(1):61-8.

### *Argenziano 2003*

Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Sera F, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. Journal of the American Academy of Dermatology 2003;48(5):679-93.

### *Argenziano 2012*

Argenziano G, Zalaudek I, Hofmann-Wellenhof R, Bakos RM, Bergman W, Blum A, et al. Total body skin examination for skin cancer screening in patients with focused symptoms. Journal of the American Academy of Dermatology 2012; 66(2):212-9.

### *Argenziano 2014*

Argenziano G, Moscarella E, Annetta A, Battarra V C, Brunetti B, Buligan C, et al. Melanoma detection in Italian pigmented lesion clinics. Giornale Italiano di Dermatologia e Venereologia 2014;149(2):161-6.

### *Ascierto 2003*

Ascierto PA, Palmieri G, Botti G, Satriano RA, Stanganelli I, Bono R, et al. Early diagnosis of malignant melanoma: Proposal of a working formulation for the management of cutaneous pigmented lesions from the Melanoma Cooperative Group. International Journal of Oncology 2003;22(6):1209-15.

### *Badertscher 2015*

Badertscher N, Tandjung R, Senn O, Kofmehl R, Held U, Rosemann T, et al. A multifaceted intervention: no increase in general practitioners' competence to diagnose skin cancer (minSKIN) - randomized controlled trial. Journal of the European Academy of Dermatology & Venereology 2015;29(8):1493-9.

### *Bafounta 2001*

Bafounta ML, Beauchet A, Aegerter P, Saiag P. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. Archives of Dermatology 2001;137(10):1343-50.

### *Banky 2005*

Banky JP, Kelly JW, English DR, Yeatman JM, Dowling JP. Incidence of new and changed nevi and melanomas detected using baseline images and dermoscopy in patients at high risk for melanoma. Archives of Dermatology 2005; 141(8):998-1006.

### *Basarab 1996*

Basarab T, Munn SE, Jones R. Diagnostic accuracy and appropriateness of general practitioner referrals to a dermatology out-patient clinic. British Journal of Dermatology 1996;135(1):70-3.

### *Bauer 2000*

Bauer P, Cristofolini P, Boi S, Burroni M, Dell'Eva G, Micciolo R, et al. Digital epiluminescence microscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. A statistical comparison between visual and computer inspection. Melanoma Research 2000;10(4):345-9.

### *Bauer 2005*

Bauer J, Blum A, Strohhacker U, Garbe C. Surveillance of patients at high risk for cutaneous malignant melanoma using digital dermoscopy. British Journal of Dermatology 2005;152(1):87-92.

### *Becker 1954*

Becker S. PItfalls in the diagnosis and treatment of melanoma. A.M.A. Archives of Dermatology and Syphilology 1954; 69(1):11-30.

### *Benelli 2000*

Benelli C, Roscetti E, Dal Pozzo V. The dermoscopic (7FFM) versus the clinical (ABCDE) diagnosis of small diameter melanoma. European Journal of Dermatology 2000;10(4):282-7.

### *Blum 2004*

Blum A. Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology. British Journal of Dermatology 2004;151(2):511-2.

### *Blum 2004a*

Blum A, Clemens J, Argenziano G. Three-colour test in dermoscopy: a re-evaluation. British Journal of Dermatology 2004; 150(5):1040.

### *Blum 2004b*

Blum A, Hofmann-Wellenhof R, Luedtke H, Ellwanger U, Steins A, Roehm S, et al. Value of the clinical history for different users of dermoscopy compared with results of digital image analysis. Journal of the European Academy of Dermatology & Venereology 2004;18(6):665-9.

### *Bolognia 1990*

Bolognia JL, Berwick M, Fine JA. Complete follow-up and evaluation of a skin cancer screening in Connecticut. Journal of the American Academy of Dermatology 1990;23(6 Pt 1):1098-106.

### *Bono 2001*

Bono A, Maurichi A, Moglia D, Camerini T, Tragni G, Lualdi M, et al. Clinical and dermatoscopic diagnosis of early amelanotic melanoma. Melanoma Research 2001;11(5):491-4.

### *Borsari 2015*

Borsari S, Longo C, Piana S, Moscarella E, Lallas A, Alfano R, et al. When the 'Ugly Duckling' Loses Brothers, It Becomes the 'Only Son of a Widowed Mother'. Dermatology 2015;231(3):222-3.

### *Borve 2012*

Borve A, Holst A, Gente-Lidholm A, Molina-Martinez R, Paoli J. Use of the mobile phone multimedia messaging service for teledermatology. Journal of Telemedicine & Telecare 2012;18(5):292-6.

### *Brown 2000*

Brown N. Exploration of diagnostic techniques for malignant melanoma: an integrative review. Clinical Excellence for Nurse Practitioners 2000;4(5):263-71.

### *Brown 2009*

Brown NH, Robertson KM, Bisset YC, Rees JL. Using a structured image database, how well can novices assign skin lesion images to the correct diagnostic grouping? Journal of Investigative Dermatology 2009;129(10):2509-12.

### *Buhl 2012*

Buhl T, Hansen-Hagge C, Korpas B, Kaune KM, Haas E, Rosenberger A, et al. Integrating static and dynamic features of melanoma: the DynaMel algorithm. Journal of the American Academy of Dermatology 2012;66(1):27-36.

### *Burki 2015*

Burki TK. Total body exam or lesion detection screening for skin cancer? Lancet Oncology 2015;16(16):e590.

### *Burr 2015*

Burr S. The assessment, history taking and differential diagnosis of pigmented skin lesions. Dermatological Nursing 2015; 14(4):(5p).

### *Burton 1998*

Burton RC, Howe C, Adamson L, Reid AL, Hersey P, Watson A, et al. General practitioner screening for melanoma: sensitivity, specificity, and effect of training. Journal of Medical Screening 1998;5(3):156-61.

### *Carli 2003*

Carli P, De Giorgi V, Giannotti B, Seidenari S, Pellacani G, Peris K, et al. Skin cancer day in Italy: method of referral to open access clinics and tumor prevalence in the examined population. European Journal of Dermatology 2003;13(1):76-9.

### *Carli 2003a*

Carli P, Mannone F, De Giorgi V, Nardini P, Chiarugi A, Giannotti B. The problem of false-positive diagnosis in melanoma screening: the impact of dermoscopy. Melanoma Research 2003;13(2):179-82.

### *Carli 2004*

Carli P, de Giorgi V, Chiarugi A, Nardini P, Weinstock MA, Crocetti E, et al. Addition of dermoscopy to conventional naked-eye examination in melanoma screening: a randomized study. Journal of the American Academy of Dermatology 2004; 50(5):683-9.

### *Carli 2004a*

Carli P, De Giorgi V, Crocetti E, Mannone F, Massi D, Chiarugi A, et al. Improvement of malignant/benign ratio in excised melanocytic lesions in the 'dermoscopy era': a retrospective study 1997-2001. British Journal of Dermatology 2004; 150(4):687-92.

### *Carli 2004b*

Carli P, Nardini P, Crocetti E, De Giorgi V, Giannotti B. Frequency and characteristics of melanomas missed at a pigmented lesion clinic: a registry-based study. Melanoma Research 2004;14(5):403-7.

### *Carli 2005*

Carli P, Chiarugi A, De Giorgi V. Examination of lesions (including dermoscopy) without contact with the patient is associated with improper management in about 30% of equivocal melanomas. Dermatologic Surgery 2005;31(2):169-72.

### *Carlos-Ortega 2007*

Carlos-Ortega B, Sanchez-Alva ME, Ysita-Morales A, Angeles-Garay U. Correlation among simple observation and dermoscopy in the study of pigmented lesions of the skin. Revista Medica del Instituto Mexicano del Seguro Social 2007; 45(6):541-8.

### *Chen 2001*

Chen SC, Bravata DM, Weil E, Olkin I. A comparison of dermatologists' and primary care physicians' accuracy in diagnosing melanoma (Structured abstract). Archives of Dermatology 2001;137(12):1627-34.

### *Chen 2006*

Chen SC, Pennie ML, Kolm P, Warshaw EM, Weisberg EL, Brown KM, et al. Diagnosing and managing cutaneous pigmented lesions: primary care physicians versus dermatologists. Journal of General Internal Medicine 2006;21(7):678-82.

### *Chiaravalloti 2014*

Chiaravalloti AJ, Laduca JR. Melanoma screening by means of complete skin exams for all patients in a dermatology practice reduces the thickness of primary melanomas at diagnosis. Journal of Clinical & Aesthetic Dermatology 2014; 7(8):18-22.

### *Ciudad-Blanco 2014*

Ciudad-Blanco C, Aviles-Izquierdo JA, Lazaro-Ochaita P, Suarez-Fernandez R. Dermoscopic findings for the early detection of melanoma: an analysis of 200 cases. Actas Dermo-Sifiliograficas 2014;105(7):683-93.

### *Cooper 2002*

Cooper SM, Wojnarowska F. The accuracy of clinical diagnosis of suspected premalignant and malignant skin lesions in renal transplant recipients. Clinical and Experimental Dermatology 2002;27(6):436-8.

### *Cornell 2015*

Cornell E, Robertson K, McIntosh RD, Rees JL. Viewing Exemplars of Melanomas and Benign Mimics of Melanoma Modestly Improves Diagnostic Skills in Comparison with the ABCD Method and Other Image-based Methods for Lay Identification of Melanoma. Acta Dermato-Venereologica 2015;95(6):681-5.

### *Cox 2008*

Cox NH, Madan V, Sanders T. The U.K. skin cancer 'two-week rule' proforma: assessment of potential modifications to

improve referral accuracy. British Journal of Dermatology 2008;158(6):1293-8.

### *DeCoste 1993*

DeCoste SD, Stern RS. Diagnosis and treatment of nevomelanocytic lesions of the skin: A community-based study. Archives of Dermatology 1993;129(1):57-62.

### *De Giorgi 2011*

* De Giorgi V, Grazzini M, Rossari S, Gori A, Alfaioli B, Papi F, et al. Adding dermatoscopy to naked eye examination of equivocal melanocytic skin lesions: effect on intention to excise by general dermatologists. Clinical & Experimental Dermatology 2011;36(3):255-9. [Other: ER4:15465901]

### *Di Carlo 2014*

Di Carlo A, Elia F, Desiderio F, Catricala C, Solivetti FM, Laino L. Can video thermography improve differential diagnosis and therapy between basal cell carcinoma and actinic keratosis? Dermatologic Therapy 2014;27(5):290-7.

### *Di Chiacchio 2010*

Di Chiacchio N, Hirata SH, Enokihara MY, Michalany NS, Fabbrocini G, Tosti A. Dermatologists' accuracy in early diagnosis of melanoma of the nail matrix. Archives of Dermatology 2010;146(4):382-7.

### *Dreiseitl 2009*

Dreiseitl S, Binder M, Hable K, Kittler H. Computer versus human diagnosis of melanoma: evaluation of the feasibility of an automated diagnostic system in a prospective clinical trial. Melanoma Research 2009;19(3):180-4.

### *Duff 2001*

Duff CG, Melsom D, Rigby HS, Kenealy JM, Townsend PL. A 6 year prospective analysis of the diagnosis of malignant melanoma in a pigmented-lesion clinic: even the experts miss malignant melanomas, but not often. British Journal of Plastic Surgery 2001;54(4):317-21.

### *Edmondson 1999*

Edmondson PC, Curley RK, Marsden RA, Robinson D, Allaway SL, Willson CD. Screening for malignant melanoma using instant photography. Journal of Medical Screening 1999;6(1):42-6.

### *Emmons 2011*

Emmons KM, Geller AC, PuleoE, Savadatti SS, Hu SW, Gorham S, et al. Skin cancer education and early detection at the beach: a randomized trial of dermatologist examination and biometric feedback. Journal of the American Academy of Dermatology 2011;64(2):282-9.

### *Engelberg 1999*

Engelberg D, Gallagher RP, Rivers JK. Follow-up and evaluation of skin cancer screening in British Columbia. Journal of the American Academy of Dermatology 1999;41(1):37-42.

### *English 2003*

English DR, Burton RC, del Mar CB, Donovan RJ, Ireland PD, Emery G. Evaluation of aid to diagnosis of pigmented skin lesions in general practice: controlled trial randomised by practice. BMJ 2003;327(7411):375.

### *English 2004*

English DR, Del Mar C, Burton RC. Factors influencing the number needed to excise: excision rates of pigmented lesions by general practitioners. Medical Journal of Australia 2004;180(1):16-9.

### *Fabbrocini 2008*

Fabbrocini G, Balato A, Rescigno O, Mariano M, Scalvenzi M, Brunetti B. Telediagnosis and face-to-face diagnosis reliability for melanocytic and non-melanocytic 'pink' lesions. Journal of the European Academy of Dermatology & Venereology 2008; 22(2):229-34.

### *Federman 1995*

Federman D, Hogan D, Taylor JR, Caralis P, Kirsner RS. A comparison of diagnosis, evaluation, and treatment of patients with dermatologic disorders. Journal of the American Academy of Dermatology 1995;32(5, Part 1):726-9.

### *Fikrle 2013*

Fikrle T, Pizinger K, Szakos H, Panznerova P, Divisova B, Pavel S. Digital dermatoscopic follow-up of 1027 melanocytic lesions in 121 patients at risk of malignant melanoma. Journal of the European Academy of Dermatology & Venereology 2013;27(2):180-6.

### *Freeman 1963*

Freeman RG, Knox JM. Clinical accuracy in diagnosis of skin tumors. Geriatrics 1963;18:546-51.

### *Friedman 1985*

Friedman RJ, Rigel DS. The clinical features of malignant melanoma. Dermatologic Clinics 1985;3(2):271-83.

### *Funt 1963*

Funt TR. Early Recognition of Cutaneous Malignant Melanoma in Adults. Journal of the Florida Medical Association 1963; 50:280-2.

### *Gerbert 1996*

Gerbert B, Maurer T, Berger T, Pantilat S, McPhee SJ, Wolff M, et al. Primary care physicians as gatekeepers in managed care: Primary care physicians' and dermatologists' skills at secondary prevention of skin cancer. Archives of Dermatology 1996;132(9):1030-8.

### *Gerbert 1998*

Gerbert B, Bronstone A, Wolff M, Maurer T, Berger T, Pantilat S, et al. Improving primary care residents' proficiency in the diagnosis of skin cancer. Journal of General Internal Medicine 1998;13(2):91-7.

### *Giannotti 2004*

Giannotti B, Carli P. Improvement of early diagnosis of melanoma in a mediterranean population: The experience of the Florence melanoma clinic [Novita in tema di diagnosi precoce del melanoma cutaneo: L'esperienza del gruppo Fiorentino]. Giornale Italiano di Dermatologia e Venereologia 2004;139(2):89-96.

### *Grana 2003*

Grana C, Pellacani G, Cucchiara R, Seidenari S. A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions. IEEE Transactions on Medical Imaging 2003;22(8):959-64.

### *Grob 1998*

Grob JJ, Bonerandi JJ. The 'ugly duckling' sign: identification of the common characteristics of nevi in an individual as a basis for melanoma screening. Archives of Dermatology 1998;134(1):103-4.

### *Guibert 2000*

Guibert P, Mollat F, Ligen M, Dreno B. Melanoma screening: report of a survey in occupational medicine. Archives of Dermatology 2000;136(2):199-202.

### *Gunduz 2003*

Gunduz K, Koltan S, Sahin MT, E Filiz E. Analysis of melanocytic naevi by dermoscopy during pregnancy. Journal of the European Academy of Dermatology & Venereology 2003;17(3):349-51.

### *Gutierrez 2013*

Gutierrez R, Rueda A, Romero E. Learning Semantic Histopathological Representation for Basal Cell Carcinoma Classification. Proc. SPIE 8676, Medical Imaging 2013: Digital Pathology, 86760U (29 March 2013). [DOI: 10.1117/12.2007117]

### *Hacioglu 2013*

Hacioglu S, Saricaoglu H, Baskan EB, Uner SI, Aydogan K, Tunali S. The value of spectrophotometric intracutaneous analysis in the noninvasive diagnosis of nonmelanoma skin cancers. Clinical & Experimental Dermatology 2013;38(5):464-9.

### *Haenssle 2010*

Haenssle HA, Korpas B, Hansen-Hagge C, Buhl T, Kaune KM, Johnsen S, et al. Selection of patients for long-term surveillance with digital dermoscopy by assessment of melanoma risk factors. Archives of Dermatology 2010;146(3):257-64.

### *Haenssle 2010a*

Haenssle HA, Korpas B, Hansen-Hagge C, Buhl T, Kaune KM, Rosenberger A, et al. Seven-point checklist for dermatoscopy: performance during 10 years of prospective surveillance of patients at increased melanoma risk. Journal of the American Academy of Dermatology 2010;62(5):785-93.

### *Hallock 1998*

Hallock GG, Lutz DA. Prospective study of the accuracy of the surgeon's diagnosis in 2000 excised skin tumors. Plast Reconstr Surg 1998;101(5):1255-61.

### *Haniffa 2007*

Haniffa MA, Lloyd JJ, Lawrence CM. The use of a spectrophotometric intracutaneous analysis device in the real-time diagnosis of melanoma in the setting of a melanoma screening clinic. British Journal of Dermatology 2007;156(6):1350-2.

### *Har-Shai 2001*

Har-Shai Y, Hai N, Taran A, Mayblum S, Barak A, Tzur E, et al. Sensitivity and positive predictive values of presurgical clinical diagnosis of excised benign and malignant skin tumors: a prospective study of 835 lesions in 778 patients. Plastic & Reconstructive Surgery 2001;108(7):1982-9.

### *Heal 2008*

Heal CF, Raasch BA, Buettner PG, Weedon D. Accuracy of clinical diagnosis of skin lesions. British Journal of Dermatology

2008;159(3):661-8.

### *Healsmith 1994*

Healsmith MF, Bourke JF, Osborne JE, Graham-Brown RA. An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma. British Journal of Dermatology 1994;130(1):48-50.

### *Higgins 1992*

Higgins EM, Hall P, Todd P, Murthi R, Du Vivier AW. The application of the seven-point check-list in the assessment of benign pigmented lesions. Clinical & Experimental Dermatology 1992;17(5):313-5.

### *Hoorens 2016*

Hoorens I, Vossaert K, Pil L, Boone B, De Schepper S, Ongenae K, et al. Total-Body Examination vs Lesion-Directed Skin Cancer Screening. JAMA Dermatology 2016;152(1):27-34.

### *Huang 1996*

Huang CL, Wasti Q, Marghoob AA, Kopf AW, De David M, Rao BK, et al. Border irregularity: Atypical moles versus melanoma. European Journal of Dermatology 1996;6(4):270-3.

### *Jamora 2003*

Jamora MJ, Wainwright BD, Meehan SA, Bystryn J-C. Improved identification of potentially dangerous pigmented skin lesions by computerized image analysis. Archives of Dermatology 2003;139(2):195-8.

### *Janda 2014*

Janda M, Loescher LJ, Banan P, Horsham C, Soyer HP. Lesion selection by melanoma high-risk consumers during skin self-examination using mobile teledermoscopy. JAMA Dermatology 2014;150(6):656-8.

### *Jensen 2015*

Jensen JD, Elewski BE. The ABCDEF rule: Combining the 'ABCDE rule' and the "ugly duckling sign" in an effort to improve patient self-screening examinations. Journal of Clinical and Aesthetic Dermatology 2015;8(2):15.

### *Jolliffe 2001*

Jolliffe VM, Harris DW, Whittaker SJ. Can we safely diagnose pigmented lesions from stored video images? A diagnostic comparison between clinical examination and stored video images of pigmented lesions removed for histology. Clinical & Experimental Dermatology 2001;26(1):84-7.

### *Jonna 1998*

Jonna BP, Delfino RJ, Newman WG, Tope WD. Positive Predictive Value for Presumptive Diagnoses of Skin Cancer and Compliance with Follow-Up among Patients Attending a Community Screening Program. Preventive Medicine 1998; 27(4):611-6.

### *Kaddu 1997*

Kaddu S, Soyer HP, Wolf IH, Rieger E, Kerl H. Reticular lentigo. Hautarzt 1997;48(3):181-5.

### *Keefe 1990*

Keefe M, Dick DC, Wakeel RA. A study of the value of the seven-point checklist in distinguishing benign pigmented lesions from melanoma. Clinical & Experimental Dermatology 1990;15(3):167-71.

### *Kelly 1986*

Kelly JW, Crutcher WA, Sagebiel RW. Clinical diagnosis of dysplastic melanocytic nevi. A clinicopathologic correlation. Journal of the American Academy of Dermatology 1986;14(6):1044-52.

### *Koh 1990*

Koh HK, Caruso A, Gage I, Geller AC, Prout MN, White H, et al. Evaluation of melanoma/skin cancer screening in Massachusetts. Preliminary results. Cancer 1990;65(2):375-9.

### *Kroemer 2011*

Kroemer S, Fruhauf J, Campbell TM, Massone C, Schwantzer G, Soyer HP, et al. Mobile teledermatology for skin tumour screening: diagnostic accuracy of clinical and dermoscopic image tele-evaluation using cellular phones. British Journal of Dermatology 2011;164(5):973-9.

### *Krol 1991*

Krol S, Keijser LM, van der Rhee HJ, Welvaart K. Screening for skin cancer in The Netherlands. Acta Dermato-Venereologica 1991;71(4):317-21.

### *Kurvers 2015*

Kurvers RH, Krause J, Argenziano G, Zalaudek I, Wolf M. Detection accuracy of collective intelligence assessments for skin cancer diagnosis. JAMA Dermatology 2015;151(12):1346-53.

### *Kvedar 1997*

Kvedar JC, Edwards RA, Menn ER, et al. THe substitution of digital images for dermatologic physical examination. Archives of Dermatology 1997;133(2):161-7.

### *Lechner 2015*

Lechner SC, Pereira LC, Reategui E, Gordon C, Byrne M, Hooper MW, et al. Erratum to: Acceptability of a Rinse Screening Test for Diagnosing Head and Neck Squamous Cell Carcinoma Among Black Americans. Journal of Racial and Ethnic Health Disparities 2015;2(NUMB 1):68.

### *Lewis 1999*

Lewis K, Gilmour E, Harrison PV, Patefield S, Dickinson Y, Manning D, et al. Digital teledermatology for skin tumours: a preliminary assessment using a receiver operating characteristics (ROC) analysis. Journal of Telemedicine & Telecare 1999; 5(Suppl 1):S57-8.

### *Lindelöf 1994*

Lindelöf B, Hedblad MA. Accuracy in the clinical diagnosis and pattern of malignant melanoma at a dermatological clinic. Journal of Dermatology 1994;21(7):461-4.

### *Lorentzen 2000*

Lorentzen H, Weismann K, Kenet RO, Secher L, Larsen FG. Comparison of dermatoscopic ABCD rule and risk stratification in the diagnosis of malignant melanoma. Acta Dermato-Venereologica 2000;80(2):122-6.

### *Luttrell 2012*

Luttrell MJ, McClenahan P, Hofmann-Wellenhof R, Fink-Puches R, Soyer HP. Laypersons' sensitivity for melanoma identification is higher with dermoscopy images than clinical photographs. British Journal of Dermatology 2012; 167(5):1037-41.

### *Machet 2005*

Machet L, Nemeth-Normand F, Giraudeau B, Perrinaud A, Tiguemounine J, Ayoub J, et al. Is ultrasound lymph node examination superior to clinical examination in melanoma follow-up? A monocentre cohort study of 373 patients. British Journal of Dermatology 2005;152(1):66-70.

### *MacKenzie-Wood 1998*

MacKenzie-Wood AR, Milton GW, de Launey JW. Melanoma: accuracy of clinical diagnosis. Australasian Journal of Dermatology 1998;39(1):31-3.

### *MacKie 1990*

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

### *Mackie 1991*

MacKie RM, Doherty VR. Seven-point checklist for melanoma. Clinical & Experimental Dermatology 1991;16(2):151-3.

### *Mackie 2002*

MacKie RM, Fleming C, McMahon AD, Jarrett P. The use of the dermatoscope to identify early melanoma using the three-colour test. British Journal of Dermatology 2002;146(3):481-4.

### *Mahendran 2005*

Mahendran R, Goodfield MJ, Sheehan-Dare RA. An evaluation of the role of a store-and-forward teledermatology system in skin cancer diagnosis and management. Clinical & Experimental Dermatology 2005;30(3):209-14.

### *Mahon 1997*

Mahon SM. A comparison of findings from two checklists for the early detection of skin cancer. Missouri Nurse 1997; 66(2):12.

### *Malvehy 2014*

Malvehy J, Hauschild A, Curiel-Lewandrowski C, Mohr P, Hofmann-Wellenhof R, Motley R, et al. Clinical performance of the Nevisense system in cutaneous melanoma detection: An international, multicentre, prospective and blinded clinical trial on efficacy and safety. British Journal of Dermatology 2014;171(5):1099-107.

### *Marghoob 1995*

Marghoob AA, Slade J, Kopf AW, Rigel DS, Friedman RJ, Perelman RO. The ABCDs of melanoma: why change? Journal of the American Academy of Dermatology 1995;32(4):682-4.

### *Marghoob 2007*

Marghoob AA, Korzenko AJ, Changchien L, Scope A, Braun RP, Rabinovitz H. The beauty and the beast sign in dermoscopy. Dermatologic Surgery 2007;33(11):1388-91.

### *Markowitz 2015*

Markowitz O, Schwartz M, Feldman E, Bienenfeld A, Bieber AK, Ellis J, et al. Evaluation of optical coherence tomography as a means of identifying earlier stage basal cell carcinomas while reducing the use of diagnostic biopsy. Journal of Clinical & Aesthetic Dermatology 2015;8(10):14-20.

### *McCarthy 1995*

McCarthy JT. ABCDs of Melanoma. Cutis 1995;56(6):313.

### *McMullan 1956*

McMullan FH, Hubener LF. Malignant melanoma: A statistical review of clinical and histological diagnoses. A.M.A. Archives of Dermatology 1956;74(6):618-9.

### *Menzies 2008*

Menzies SW, Kreusch J, Byth K, Pizzichetta MA, Marghoob A, Braun R, et al. Dermoscopic evaluation of amelanotic and hypomelanotic melanoma. Archives of Dermatology 2008;144(9):1120-7.

### *Menzies 2011*

Menzies SW, Stevenson ML, Altamura D, Byth K. Variables predicting change in benign melanocytic nevi undergoing short-term dermoscopic imaging. Archives of Dermatology 2011;147(6):655-9.

### *Menzies 2013*

Menzies SW, Moloney FJ, Byth K, Avramidis M, Argenziano G, Zalaudek I, et al. Dermoscopic evaluation of nodular melanoma. JAMA Dermatology 2013;149(6):699-709.

### *Moffatt 2006*

Moffatt CR, Green AC, Whiteman DC. Diagnostic accuracy in skin cancer clinics: the Australian experience. International Journal of Dermatology 2006;45(6):656-60.

### *Mohammad 2015*

Mohammad E-A, Mansour M, Parichehr K, Farideh D, Amirhossein R, Ahmad S-A. Assessment of clinical diagnostic accuracy compared with pathological diagnosis of basal cell carcinoma. Indian Dermatology Online Journal 2015; 6(4):258-62.

### *Morrison 2001*

Morrison A, O'Loughlin S, Powell FC. Suspected skin malignancy: a comparison of diagnoses of family practitioners and dermatologists in 493 patients. International Journal of Dermatology 2001;40(2):104-7.

### *Nachbar 1994*

* Nachbar F, Stolz W, Merkle T, Cognetta AB, Vogt T, Landthaler M, et al. The ABCD rule of dermatoscopy. High prospective value in the diagnosis of doubtful melanocytic skin lesions. Journal of the American Academy of Dermatology 1994;30(4):551-9. [Other: ER4:15466022]

### *Nathansohn 2007*

Nathansohn N, Orenstein A, Trau H, Liran A, Schachter J. Pigmented lesions clinic for early detection of melanoma: preliminary results. Israel Medical Association Journal: Imaj 2007;9(10):708-12.

### *Nilles 1994*

Nilles M, Boedeker RH, Schill WB. Surface microscopy of naevi and melanomas--clues to melanoma. British Journal of Dermatology 1994;130(3):349-55.

### *Osborne 1998*

Osborne JE, Bourke JF, Holder J, Colloby P, Graham-Brown RA. The effect of the introduction of a pigmented lesion clinic on the interval between referral by family practitioner and attendance at hospital. British Journal of Dermatology 1998; 138(3):418-21.

### *Osborne 1999*

Osborne JE, Bourke JF, Graham-Brown RA, Hutchinson PE. False negative clinical diagnoses of malignant melanoma. British Journal of Dermatology 1999;140(5):902-8.

### *Parslew 1997*

Parslew RA, Rhodes LE. Accuracy of diagnosis of benign skin lesions in hospital practice: a comparison of clinical and histological findings. Journal of the European Academy of Dermatology and Venereology 1997;9(2):137-41.

### *Pazzini 1996*

Pazzini C, Pozzi M, Betti R, Vergani R, Crosti C. Improvement of diagnostic accuracy in the clinical diagnosis of pigmented skin lesions by epiluminescence microscopy. Skin Cancer 1996;11(2):159-61.

### *Perednia 1992*

Perednia DA, Gaines JA, Rossum AC. Variability in physician assessment of lesions in cutaneous images and its

implications for skin screening and computer-assisted diagnosis. Archives of Dermatology 1992;128(3):357-64.

### *Perrinaud 2007*

Perrinaud A, Gaide O, French LE, Saurat JH, Marghoob AA, Braun RP. Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems. British Journal of Dermatology 2007;157(5):926-33.

### *Piccolo 2000*

Piccolo D, Smolle J, Argenziano G, Wolf IH, Braun R, Cerroni L, et al. Teledermoscopy--results of a multicentre study on 43 pigmented skin lesions. Journal of Telemedicine & Telecare 2000;6(3):132-7.

### *Piccolo 2002*

Piccolo D, Peris K, Chimenti S, Argenziano G, Soyer HP. Jumping into the future using teledermoscopy. SKINmed 2002; 1(1):20-4.

### *Pizzichetta 2001*

Pizzichetta MA, Talamini R, Piccolo D, Argenziano G, Pagnanelli G, Burgdorf T, et al. The ABCD rule of dermatoscopy does not apply to small melanocytic skin lesions. Archives of Dermatology 2001;137(10):1376-8.

### *Provost 1998*

Provost N, Kopf AW, Rabinovitz HS, Stolz W, DeDavid M, Wasti Q, et al. Comparison of conventional photographs and telephonically transmitted compressed digitized images of melanomas and dysplastic nevi. Dermatology 1998; 196(3):299-304.

### *Quereux 2011*

Quereux G, Lequeux Y, Cary M, Jumbou O, Nguyen JM, Dreno B. Feasibility and effectiveness of a melanoma targeted screening strategy. Melanoma Research 2011;21:e1-2.

### *Rallan 2006*

Rallan D, Dickson M, Bush NL, Harland CC, Mortimer P, Bamber JC. High-resolution ultrasound reflex transmission imaging and digital photography: potential tools for the quantitative assessment of pigmented lesions. Skin Research & Technology 2006;12(1):50-9.

### *Rampen 1988*

Rampen FH, Rumke P. Referral pattern and accuracy of clinical diagnosis of cutaneous melanoma. Acta Dermato-venereologica 1988;68(1):61-4.

### *Reeck 1999*

Reeck MC, Chuang T-Y, Eads TJ, Faust HB, Farmer ER, Hood AF. The diagnostic yield in submitting nevi for histologic examination. Journal of the American Academy of Dermatology 1999;40(4):567-71.

### *Riddell 1961*

Riddell Jr JM. A report of 300 patients with skin cancer. Texas State Journal of Medicine 1961;57:588-92.

### *Rigel 1993*

Rigel DS, Friedman RJ. The rationale of the ABCDs of early melanoma. Journal of the American Academy of Dermatology 1993;29(6):1060-1.

### *Robati 2014*

Robati RM, Toossi P, Karimi M, Ayatollahi A, Esmaeli M. Screening for Skin Cancer: A Pilot Study in Tehran, Iran. Indian Journal of Dermatology 2014;59(1):1-4.

### *Robinson 2010*

Robinson JK, Turrisi R, Mallett K, Stapleton J, Pion M. Comparing the efficacy of an in-person intervention with a skin self-examination workbook. Archives of Dermatology 2010;146(1):91-4.

### *Rosado 2003*

Rosado B, Menzies S, Harbauer A, Pehamberger H, Wolff K, Binder M, et al. Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. Archives of Dermatology 2003;139(3):361-7; discussion 366.

### *Rossi 2000*

Rossi CR, Vecchiato A, Bezze G, Mastrangelo G, Montesco MC, Mocellin S, et al. Early detection of melanoma: an educational campaign in Padova, Italy. Melanoma Research 2000;10(2):181-7.

### *Roush 1986*

Roush GC, Kirkwood JM, Ernstoff M, Somma SJ, Duray PH, Klaus SN, et al. Reproducibility and validity in the clinical diagnosis of the nonfamilial dysplastic nevus: work in progress. Recent Results in Cancer Research 1986;102:154-8.

### *Salvio 2011*

Salvio AG, Assumpcao JA, Segalla JG, Panfilo BL, Nicolini HR, Didone R. One year experience of a model for melanoma continuous prevention in the city of Jau (Sao Paulo), Brazil. Anais Brasileiros de Dermatologia 2011;86(4):669-74.

### *Schindewolf 1994*

Schindewolf T, Schiffner R, Stolz W, Albert R, Abmayr W, Harms H. Evaluation of different image acquisition techniques for a computer vision system in the diagnosis of malignant melanoma. Journal of the American Academy of Dermatology 1994; 31(1):33-41.

### *Schmoeckel 1987*

Schmoeckel C, Braun-Falco O. Diagnosis of early malignant melanoma: sensitivity and specificity of clinical and histological criteria. Pigment Cell 1987;8:96-106.

### *Schwartzberg 2005*

Schwartzberg JB, Elgart GW, Romanelli P, Ma F, Federman DG, Kirsner RS. Accuracy and predictors of basal cell carcinoma diagnosis. Dermatologic surgery 2005;31(5):534-7.

### *Seidenari 2006*

Seidenari S, Longo C, Giusti F, Pellacani G. Clinical selection of melanocytic lesions for dermoscopy decreases the identification of suspicious lesions in comparison with dermoscopy without clinical preselection. British Journal of Dermatology 2006;154(5):873-9.

### *Seidenari 2006a*

Seidenari S, Pellacani G, Grana C. Asymmetry in dermoscopic melanocytic lesion images: a computer description based on colour distribution. Acta Dermato-Venereologica 2006;86(2):123-8.

### *Shariff 2010*

Shariff Z, Roshan A, Williams AM, Platt AJ. 2-Week wait referrals in suspected skin cancer: does an instructional module for general practitioners improve diagnostic accuracy? Surgeon Journal of the Royal Colleges of Surgeons of Edinburgh & Ireland 2010;8(5):247-51.

### *Sondak 2015*

Sondak VK, Glass LF, Geller AC. Risk-stratified screening for detection of melanoma. JAMA 2015;313(6):616-7.

### *Soyer 2004*

Soyer HP, Argenziano G, Zalaudek I, Corona R, Sera F, Talamini R, et al. Three-point checklist of dermoscopy. A new screening method for early detection of melanoma. Dermatology 2004;208(1):27-31.

### *Stanganelli 1998a*

Stanganelli I, Bucchi L. Epiluminescence microscopy versus clinical evaluation of pigmented skin lesions: effects of Operator's training on reproducibility and accuracy. Dermatology and Venereology Society of the Canton of Ticino. Dermatology 1998;196(2):199-203.

### *Stanley 2003*

Stanley RJ, Moss RH, Van Stoecker W, Aggarwal C. A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images. Computerized Medical Imaging & Graphics 2003;27(5):387-96.

### *Stathopoulos 2015*

Stathopoulos P, Ghaly G, Sisodia B, Harrop C. Positive predictive value of clinical diagnosis of head and neck non-melanoma skin malignancies. How accurate are we? Oral and Maxillofacial Surgery 2015;19(NUMB 4):387-90.

### *Stratigos 2007*

Stratigos A, Nikolaou V, Kedicoglou S, Antoniou C, Stefanaki I, Haidemenos G, et al. Melanoma/skin cancer screening in a Mediterranean country: results of the Euromelanoma Screening Day Campaign in Greece. Journal of the European Academy of Dermatology & Venereology 2007;21(1):56-62.

### *Tandjung 2015*

Tandjung R, Badertscher N, Kleiner N, Wensing M, Rosemann T, Braun R P, et al. Feasibility and diagnostic accuracy of teledermatology in Swiss primary care: process analysis of a randomized controlled trial. Journal of Evaluation in Clinical Practice 2015;21(2):326-31.

### *Terrill 2009*

Terrill PJ, Fairbanks S, Bailey M. Is there just one lesion? The need for whole body skin examination in patients presenting with non-melanocytic skin cancer. ANZ Journal of Surgery 2009;79(10):707-12.

### *Terushkin 2010*

Terushkin V, Braga JC, Dusza SW, Scope A, Busam K, Marghoob AA, et al. Agreement on the clinical diagnosis and

management of cutaneous squamous neoplasms. Dermatologic Surgery 2010;36(10):1514-20.

### *Terushkin 2010a*

Terushkin V, Warycha M, Levy M, Kopf AW, Cohen DE, Polsky D. Analysis of the benign to malignant ratio of lesions biopsied by a general dermatologist before and after the adoption of dermoscopy. Archives of Dermatology 2010; 146(3):343-4.

### *Thomson 2005*

Thomson MA, Loffeld A, Marsden JR. More skin cancer detected from nonurgent referrals. British Journal of Dermatology 2005;153(2):453-4.

### *Torrey 1941*

Torrey FA, Levin EA. Comparison of the Clinical and the Pathologic Diagnoses of Malignant Conditions of the Skin. Archives of Dermatology 1941;43(3):532.

### *Ulrich 2015*

Ulrich M, von Braunmuehl T, Kurzen H, Dirschka T, Kellner C, Sattler E, et al. The sensitivity and specificity of optical coherence tomography for the assisted diagnosis of nonpigmented basal cell carcinoma: an observational study. British Journal of Dermatology 2015;173(2):428-35.

### *van der Rhee 2010*

van der Rhee JI, Bergman W, Kukutsch NA. The impact of dermoscopy on the management of pigmented lesions in everyday clinical practice of general dermatologists: a prospective study. British Journal of Dermatology 2010;162(3):563-7.

### *van der Rhee 2011*

van der Rhee JI, Bergman W, Kukutsch NA. Impact of dermoscopy on the management of high-risk patients from melanoma families: a prospective study. Acta Dermato-Venereologica 2011;91(4):428-31.

### *Vasili 2010*

Vasili E, Shkodrani E, Harja D, Labinoti L, Zoto A. Retrospective study of 70 patients with NMSC. Melanoma Research 2010; 20:e63.

### *Wagner 1985*

Wagner RF, Wagner D, Tomich JM, Wagner KD, Grande DJ. Residents' Corner: Diagnoses of Skin Disease: Dermatologists vs. Nondermatologists. Journal of Dermatologic Surgery and Oncology 1985;11(5):476-9.

### *Walter 2010*

Walter FM, Morris HC, Humphrys E, Hall PN, Kinmonth AL, Prevost AT, et al. Protocol for the MoleMate UK Trial: a randomised controlled trial of the MoleMate system in the management of pigmented skin lesions in primary care [ISRCTN 79932379]. BMC Family Practice 2010;11:36.

### *Walter 2013*

Walter FM, Prevost AT, Vasconcelos J, Hall PN, Burrows NP, Morris HC, et al. Using the 7-point checklist as a diagnostic aid for pigmented skin lesions in general practice: a diagnostic validation study. British Journal of General Practice 2013; 63(610):e345-53.

### *Warshaw 2009*

Warshaw EM, Lederle FA, Grill JP, Gravely AA, Bangerter AK, Fortier LA, et al. Accuracy of teledermatology for pigmented neoplasms.[Erratum appears in J Am Acad Dermatol. 2010 Feb;62(2):319]. Journal of the American Academy of Dermatology 2009;61(5):753-65.

### *Warshaw 2009a*

Warshaw EM, Lederle FA, Grill JP, Gravely AA, Bangerter AK, Fortier LA, et al. Accuracy of teledermatology for nonpigmented neoplasms. Journal of the American Academy of Dermatology 2009;60(4):579-88. [DOI: 10.1016/j.jaad.2008.11.892]

### *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(255):348-52.

### *Westbrook 2006*

Westbrook RH, Goyal N, Gawkrodger DJ. Diagnostic accuracy for skin cancer: comparison of general practitioner with dermatologist and dermatopathologist. Journal of Dermatological Treatment 2006;17(1):57-8.

### *Whitaker-Worth 1998*

Whitaker-Worth DL, Susser WS, Grant-Kels JM. Clinical dermatologic education and the diagnostic acumen of medical students and primary care residents. International Journal of Dermatology 1998;37(11):855-9.

### *Whited 1998*

Whited JD, Mills BJ, Hall RP, Drugge RJ, Grichnik JM, Simel DL. A pilot trial of digital imaging in skin cancer. Journal of Telemedicine & Telecare 1998;4(2):108-12.

### *Williams 1991*

Williams HC, Smith D, du Vivier A. Melanoma: Differences observed by general surgeons and dermatologists. International Journal of Dermatology 1991;30(4):257-61.

### *Winkelmann 2015*

Winkelmann RR, Hauschild A, Tucker N, White R, Rigel DS. The impact of multispectral digital skin lesion analysis on German dermatologist decisions to biopsy atypical pigmented lesions with clinical characteristics of melanoma. Journal of Clinical & Aesthetic Dermatology 2015;8(10):27-9.

### *Winkelmann 2015a*

Winkelmann RR, Yoo J, Tucker N, White R, Rigel DS. Impact of guidance provided by a multispectral digital skin lesion analysis device following dermoscopy on decisions to biopsy atypical melanocytic lesions. Journal of Clinical & Aesthetic Dermatology 2015;8(9):21-4.

### *Wolf 1998*

Wolf IH, Smolle J, Soyer HP, Kerl H. Sensitivity in the clinical diagnosis of malignant melanoma. Melanoma Research 1998; 8(5):425-9.

### *Yoo 2015*

Yoo J, Tucker N, White R, Rigel D. The impact of probability of melanoma information provided by a multispectral digital skin lesion analysis device (MSDSLA) on resident dermatologists' decisions to biopsy clinical atypical lesions. Journal of the American Academy of Dermatology 2015;1):AB177.

### *Youl 2007*

Youl PH, Baade PD, Janda M, Del Mar CB, Whiteman DC, Aitken JF. Diagnosing skin cancer in primary care: how do mainstream general practitioners compare with primary care skin cancer clinic doctors? Medical Journal of Australia 2007; 187(4):215-20.

### *Youl 2007a*

Youl PH, Raasch BA, Janda M, Aitken JF. The effect of an educational programme to improve the skills of general practitioners in diagnosing melanocytic/pigmented lesions. Clinical and Experimental Dermatology 2007;32(4):365-70.

### *Zaballos 2013*

Zaballos P, Banuls J, Cabo H, Llambrich A, Salsench E, Puig S, et al. The usefulness of dermoscopy for the recognition of basal cell carcinoma--seborrhoeic keratosis compound tumours. Australasian Journal of Dermatology 2013;54(3):208-12.

### *Zou 2001*

Zou KH. Comparison of correlated receiver operating characteristic curves derived from repeated diagnostic test data. Academic Radiology 2001;8(3):225-33.


### 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]

### *American Academy of Dermatology 2015*

American Academy of Dermatology Ad Hoc Task Force for the ABCDEs of Melanoma, Tsao H, Olazagasti JM, Cordoro KM, Brewer JD, Taylor SC, et al. Early detection of melanoma: reviewing the ABCDEs. Journal of the American Academy of Dermatology 2015;72(4):717-23. [PubMed: 25698455]

### *Annessi 2007*

Annessi G, Bono R, Sampogna F, Faraggiana T, Abeni D. Sensitivity, specificity, and diagnostic accuracy of three dermoscopic algorithmic methods in the diagnosis of doubtful melanocytic lesions: the importance of light brown structureless areas in differentiating atypical melanocytic nevi from thin melanomas. Journal of the American Academy of Dermatology 2007;56(5):759-67. [PubMed: 17316894]

### *Argenziano 1998*

Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M. Epiluminescence microscopy for the diagnosis of

doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Archives of Dermatology 1998;134(12):1563-70. [ PubMed: 9875194]

### *Argenziano 2001*

Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions--a valuable tool for early diagnosis of melanoma. Lancet Oncology 2001;2(7):443-9. [ PubMed: 11905739]

### *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 & Venereology 2014; 28(9):1170-8. [ PubMed: 23962170]

### *Ascierto 2000*

Ascierto PA, Palmieri G, Celentano E, Parasole R, Caraco C, Daponte A, et al. Sensitivity and specificity of epiluminescence microscopy: evaluation on a sample of 2731 excised cutaneous pigmented lesions. The Melanoma Cooperative Study. British Journal of Dermatology 2000;142(5):893-8. [ PubMed: 10809845]

### *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]

### *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):330-9. [ PubMed: 27663092]

### *Binder 1997*

Binder M, Schwarz M, Steiner A, Kittler H, Muellner M, Wolff K, et al. Epiluminescence microscopy of small pigmented skin lesions: short-term formal training improves the diagnostic performance of dermatologists. Journal of the American Academy of Dermatology 1997;36(2 Pt 1):197-202. [ PubMed: 9039168]

### *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 2017*

Cancer Research UK. Skin cancer statistics. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer#heading-One (accessed prior to 19 July 2017).

### *Carli 1994*

Carli P, De Giorgi V, Donati E, Pestelli E, Gianotti B. Epiluminescence microscopy reduces the risk of removing clinically atypical, but histologically common, melanocytic lesions [La Microscopia a Epiluminescenza (Elm) Riduce Il Rischio Di Asportare Lesioni Melanocitarie Clinicamente Sospette Ma Istologicamente Comuni]. Giornale Italiano di Dermatologia e Venereologia 1994;129(12):5990605.

### *Chao 2013*

Chao D, London Cancer North and East. London Cancer, Guidelines for Cutaneous Malignant Melanoma Management August 2014. 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]

### *Chu 2006*

Chu H, Cole SR. Bivariate meta-analysis for sensitivity and specificity with sparse data: a generalized linear mixed model approach (comment). Journal of Clinical Epidemiology 2006;59(12):1331-2. [ PubMed: 17098577]

### *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).

### *Corbo 2012*

Corbo MD, Wismer J. Agreement between dermatologists and primary care practitioners in the diagnosis of malignant melanoma: review of the literature. Journal of Cutaneous Medicine & Surgery 2012;16(5):306-10. [ PubMed: 22971304]

### *Cordoro 2013*

Cordoro KM, Gupta D, Frieden IJ, McCalmont T, Kashani-Sabet M. Pediatric melanoma: results of a large cohort study and proposal for modified ABCD detection criteria for children. Journal of the American Academy of Dermatology 2013; 68(6):913-25. [ PubMed: 23395590]

### *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 2018*

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 cutaneous melanoma in adults. Cochrane Database of Systematic Reviews (in press).

### *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]

### *Elstein 2002*

Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. BMJ 2002;324(7339):729-32. [ PubMed: 11909793]

### *Erdmann 2013*

Erdmann F, Lortet-Tieulent J, Schuz J, Zeeb H, Greinert R, Breitbart EW, 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]

### *Farina 2000*

Farina B, Bartoli C, Bono A, Colombo A, Lualdi M, Tragni G, et al. Multispectral imaging approach in the diagnosis of cutaneous melanoma: potentiality and limits. Physics in Medicine and Biology 2000;45(5):1243-54. [ PubMed: 10843103]

### *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]

### *Gandini 2005*

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. European Journal of Cancer 2005;41(1):28-44. [ PubMed: 15617989]

### *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]

### *Gereli 2010*

Gereli MC, Onsun N, Atilganoglu U, Demirkesen C. Comparison of two dermoscopic techniques in the diagnosis of clinically atypical pigmented skin lesions and melanoma: seven-point and three-point checklists. International Journal of Dermatology 2010;49(1):33-8. [ PubMed: 20465608]

### *Gershenwald 2017*

Gershenwald JE, Scolyer RA, Hess KR, Sondak VK, Long GV, Ross, MI, et al. Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA: A Cancer Journal for Clinicians 2017; 67(6):472-492. [ PubMed: 29028110]

### *Ginsberg 1993*

Ginsberg JS, Brill-Edwards PA, Demers C, Donovan D, Ranju A. D-dimer in patients with clinically suspected pulmonary embolism. Chest 1993;104(6):1679-84. [ PubMed: 8252941]

### *Girardi 2006*

Girardi S, Gaudy C, Gouvernet J, Teston J, Richard MA, Grob JJ. Superiority of a cognitive education with photographs over ABCD criteria in the education of the general population to the early detection of melanoma: a randomized study. International Journal of Cancer 2006;118(9):2276-80. [ PubMed: 16331608]

### *Goldsmith 2014*

Goldsmith SM. A unifying approach to the clinical diagnosis of melanoma including "D" for "Dark" in the ABCDE criteria. Dermatology Practical and Conceptual 2014;4(4):75-8. [ PubMed: 25396093]

### *Goodson 2010*

Goodson AG, Florell SR, Hyde M, Bowen GM, Grossman D. Comparative analysis of total body and dermatoscopic photographic monitoring of nevi in similar patient populations at risk for cutaneous melanoma. Dermatologic Surgery 2010; 36(7):1087-98. [ PubMed: 20653722]

### *Harrington 2017*

Harrington E, Clyne B, Wesseling N, Sandhu H, Armstrong L, Bennett H, et al. Diagnosing malignant melanoma in ambulatory care: a systematic review of clinical prediction rules. BMJ Open 2017;7(3):e014096. [ PubMed: 28264830]

### *Herschorn 2012*

Herschorn A. Dermoscopy for melanoma detection in family practice. Canadian Family Physician 2012;58(7):740-5. [ PubMed: 22859635]

### *Kasprzak 2015*

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

### *Kittler 1999*

Kittler H, Seltenheim M, Dawid M, Pehamberger H, Wolff K, Binder M. Morphologic changes of pigmented skin lesions: a useful extension of the ABCD rule for dermatoscopy. Journal of the American Academy of Dermatology 1999;40(4):558-62. [ PubMed: 10188673]

### *Kittler 2002*

Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncology 2002;3(3):159-65. [ PubMed: 11902502]

### *Kittler 2011*

Kittler H, Rosendahl C, Cameron A, Tschandl P. Dermatoscopy. An algorithmic method based on pattern analysis. Austria: Facultas.WUV, 2011. [Other: ISBN-10: 3708907175]

### *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]

### *Lachs 1992*

Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. Annals of Internal Medicine 1992;117(2):135-40. [ PubMed: 1605428]

### *Leeflang 2013*

Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. CMAJ 2013;185(11):E537-44. [ PubMed: 23798453]

### *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 [Review]. 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]

### *Liu 2005*

Liu W, Hill D, Gibbs AF, Tempany M, Howe C, Borland R, et al. What features do patients notice that help to distinguish between benign pigmented lesions and melanomas?: the ABCD(E) rule versus the seven-point checklist. Melanoma Research 2005;15(6):549-54. [ PubMed: 16314742]

### *Loescher 2011*

Loescher LJ, Harris JM Jr, Curiel-Lewandrowski C. A systematic review of advanced practice nurses' skin cancer assessment barriers,skin lesion recognition skills,and skin cancer training activities. Journal of the American Academy of Nurse Practitioners 2011;23(12):667-73. [ PubMed: 22145657]

### *MacKie 1985*

MacKie RM, English J, Aitchison TC, Fitzsimons CP, Wilson P. The number and distribution of benign pigmented moles (melanocytic naevi) in a healthy British population. British Journal of Dermatology 1985;113(2):167-74. [ PubMed: 4027184]

### *MacKie 1992*

MacKie RM, Fleming C, McMahon AD, Jarett P. The use of the dermoscope to identify early melanoma using the three-colour test. British Journal of Dermatology 2002;146(3):481-4. [ PubMed: 11952549]

### *Maley 2014*

Maley A, Rhodes AR. Cutaneous melanoma: preoperative tumor diameter in a general dermatology outpatient setting. Dermatologic Surgery 2014;40(4):446-54. [ PubMed: 24479783]

### *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]

### *Mascaro 1998*

Mascaro JM Jr, Mascaro JM. The dermatologist's position concerning nevi: a vision ranging from "the ugly duckling" to "little red riding hood". Archives of Dermatology 1998;134(11):1484-5. [ PubMed: 9828892]

### *Menzies 1996*

Menzies SW, Ingvar C, Crotty KA, McCarthy WH. Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. Archives of Dermatology 1996;132(10):1178-82. [ PubMed: 8859028]

### *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. [ PubMed: 22033277]

### *Moons 1997*

Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and Bayes' theorem in assessing diagnostic probabilities: a clinical example. Epidemiology 1997;8(1):12-7. [ PubMed: 9116087]

### *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]

### *Morton 1998*

Morton CA, MacKie RM. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. British Journal of Dermatology 1998;138(2):283-7. [ PubMed: 9602875]

### *Moynihan 1994*

Moynihan GD. The 3 Cs of melanoma: time for a change? Journal of the American Academy of Dermatology 1994; 30(3):510-1. [ PubMed: 8113477]

### *NICE 2015*

National Institute for Health and Care Excellence. Melanoma: assessment and management. www.nice.org.uk/guidance/ng14 (accessed prior to 19 July 2017).

### *NICE 2015c*

National Institute for Health and Clinical Excellence. Suspected cancer: recognition and referral. www.nice.org.uk/guidance/ng12 (accessed prior to 19 July 2017).

### *Norman 1989*

Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. Archives of Dermatology 1989;125(8):1063-8. [ PubMed: 2757402]

### *Norman 2009*

Norman G, Barraclough K, Dolovich L, Price D. Iterative diagnosis. BMJ 2009;339:b3490. [ PubMed: 19773326]

### *Pasquali 2018*

Pasquali S, Hadjinicolaou AV, Chiarion Sileni V, Rossi CR, Mocellin S. Systemic treatments for metastatic cutaneous melanoma. Cochrane Database of Systematic Reviews 2018, Issue 2. Art. No.: CD011123 DOI: 10.1002/14651858.CD011123.pub2.

### *Pehamberger 1993*

Pehamberger H, Binder M, Steiner A, Wolff K. In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. Journal of Investigative Dermatology 1993;100(3):356s-62s. [ PubMed: 8440924]

### *Rademaker 2010*

Rademaker M, Oakley A. Digital monitoring by whole body photography and sequential digital dermoscopy detects thinner melanomas. Journal of Primary Health Care 2010;2(4):268-72. [ PubMed: 21125066]

### *Reitsma 2005*

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology 2005;58(10):982-90. [ PubMed: 16168343]

### *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]

### *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]

### *Rutter 2001*

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in Medicine 2001;20(19):2865-84. [ PubMed: 11568945]

### *Salerni 2012*

Salerni G, Carrera C, Lovatto L, Marti-Laborda RM, Isern G, Palou J, et al. Characterization of 1152 lesions excised over 10 years using total-body photography and digital dermatoscopy in the surveillance of patients at high risk for melanoma. Journal of the American Academy of Dermatology 2012;67(5):836-45. [ PubMed: 22521205]

### *SAS 2012*

SAS 2012 [Computer program]. Version 9.3. Cary, NC, USA: SAS Institute Inc., 2012.

### *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 19 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).

### *Sober 1979*

Sober AJ, Fitzpatrick TB, Mihm MC, Wise TG, Pearson BJ, Clark WH, et al. Early recognition of cutaneous melanoma. JAMA 1979;242(25):2795-9. [ PubMed: 501893]

### *Stolz 1994*

Stolz W, Riemann A, Cognetta AB, Pillet L, Abmayer W, Holzel D, et al. ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma. European Journal of Dermatology 1994;4(7):521-7. [EMBASE: 24349113]

### *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]

### *Takwoingi 2010*

Takwoingi Y, Deeks J. MetaDAS: a SAS macro for meta-analysis of diagnostic accuracy studies. User Guide Version 1.3. 2010. www.methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/MetaDAS%20Readme%20v1.3%20May%202 (accessed prior to 17 July 2017).

### *Takwoingi 2015*

Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. Statistical Methods in Medical Research 2015;24:1-19. [DOI: 10.1177/0962280215592269]

### *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]

### *Usher-Smith 2016*

Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ 2016; 353:i3139. [DOI: 10.1136/bmj.i3139]

### *Vestergaard 2008*

Vestergaard ME, Macaskill P, Holt PE, Menzies SW. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. British Journal of Dermatology 2008; 159(3):669-76. [ PubMed: 18616769]

### *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]

### *Yagerman 2014*

Yagerman SE, Chen L, Jaimes N, Dusza SW, Halpern AC, Marghoob A. 'Do UC the melanoma?' Recognising the importance of different lesions displaying unevenness or having a history of change for early melanoma detection. Australasian Journal of Dermatology 2014;55(2):119-24. [ PubMed: 24548383]

### *Zalaudek 2006*

Zalaudek I, Argenziano G, Soyer HP, Corona R, Sera F, Blum A, et al. Dermoscopy Working Group. Three-point checklist of dermoscopy: an open Internet study. British Journal of Dermatology 2006;154(3):431-7. [ PubMed: 16445771]


### Other published versions of this review


### *Dinnes 2015*

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 2015a*

Dinnes J, Wong KY, Gulati A, Chuchu N, Leonardi-Bee J, Bayliss SE, et al. 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

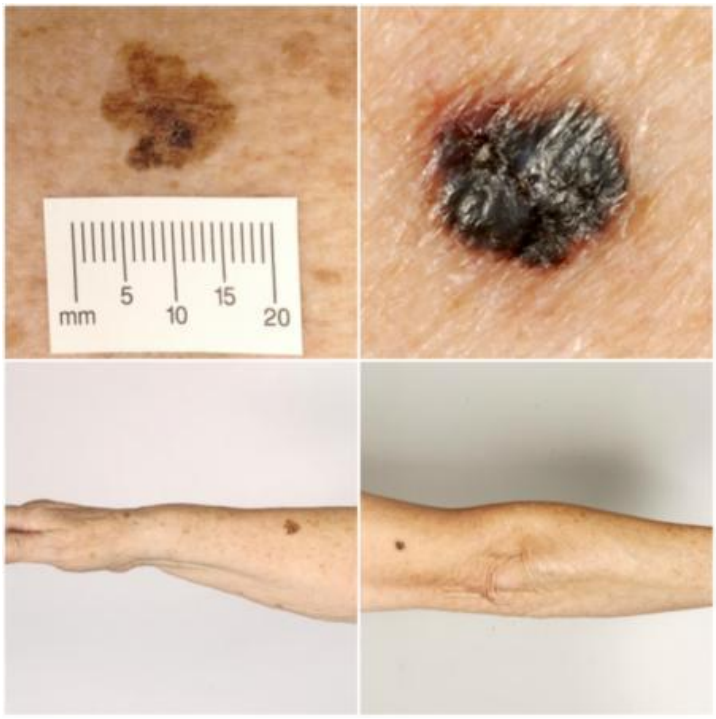
## Data and analyses

## Data tables by test

| Test | Studies | Participants |
|---|---|---|
| 1 Visual inspection - in-person (MM) | 7 | 6857 |
| 2 Visual inspection - image-based (MM) | 5 | 599 |
| 3 Visual inspection - in-person (MM+MiS) | 28 | 25604 |
| 4 Visual inspection - image-based (MM+MiS) | 11 | 1243 |
| 5 Visual inspection - in-person (Any) | 7 | 8091 |
| 6 Visual inspection - image-based (Any) | 3 | 547 |
| 7 MM2- VI - in-person - no algorithm | 21 | 19330 |
| 8 MM2- VI - in-person - no algorithm (alt thresholds) | 2 | 475 |
| 9 MM2- VI - in-person - (A)BCD(E) at NR or standard threshold | 6 | 5501 |
| 10 MM2-VI - in-person - ABCD at NR | 2 | 3548 |
| 11 MM2-VI - in-person - ABCDE at >=1 | 2 | 1541 |
| 12 MM2-VI - in-person - ABCDE at >=2 | 3 | 1761 |
| 13 MM2-VI - in-person - ABCDE at >=3 | 2 | 1541 |
| 14 MM2-VI - in-person - ABCDE at >=4 | 2 | 1541 |
| 15 MM2-VI - in-person - ABCDE at >=5 | 2 | 1541 |
| 16 MM2-VI - in-person - BCD at >=1 | 1 | 192 |
| 17 MM2-VI - in-person - BCD at >=2 | 1 | 192 |
| 18 MM2-VI - in-person - BCD at >=3 | 1 | 192 |
| 19 MM2-VI - in-person - 7point at >=2 | 1 | 205 |
| 20 MM2-VI - in-person - 7point at >=3 | 1 | 205 |
| 21 MM2-VI - in-person - 7point at >=4 | 1 | 205 |
| 22 MM2-VI - in-person - 7point(rev) at >=3 | 1 | 773 |
| 23 MM2-VI - in-person - Collas at >=1 | 1 | 353 |
| 24 MM2- VI - image-based - no algorithm | 9 | 1090 |
| 25 MM2- VI - image-based - no algorithm (alt threshold) | 0 | 0 |
| 26 MM2-VI - image-based - ABCD(E) at standard | 2 | 153 |
| 27 MM2-VI - image-based - ABCD at >=2 | 1 | 103 |
| 28 MM2-VI - image-based - ABCD at >=3 | 1 | 103 |
| 29 MM2-VI - image-based - ABCDE at >=2 | 1 | 50 |
| 30 MM2-VI - image-based - ABCDE at >=3 | 1 | 50 |
| 31 MM2- VI - in-person - experience NR | 12 | 16778 |
| 32 MM2- VI - in-person - experience High | 9 | 3547 |
| 33 MM2- VI - in-person - experience Moderate | 1 | 567 |
| 34 MM2- VI - in-person - experience Low | 4 | 2008 |
| 35 MM2- VI - in-person - experience Mixed | 2 | 2704 |
| 36 MM2- VI - image-based - experience NR | 5 | 663 |
| 37 MM2- VI - image-based - experience High | 5 | 540 |
| 38 MM2- VI - image-based - experience Low | 1 | 134 |
| 39 MM2- VI - image-based - experience Mixed | 2 | 90 |
| 40 VI - in-person - Expert consultant (MM+MiS) | 9 | 3547 |
| 41 VI - in-person - Consultant (MM+MiS) | 12 | 16778 |
| 42 VI - in-person - Resident/registrar (MM+MiS) | 2 | 1236 |
| 43 VI - in-person - Mixed qualifications (secondary care) (MM+MiS) | 2 | 2704 |
| 44 VI - in-person - GP (MM+MiS) | 3 | 1339 |
| 45 MM2- VI - image-based - Expert consultant | 4 | 700 |
| 46 MM2- VI - image-based - Consultant | 4 | 200 |
| 47 MM2- VI - image-based - Mixed qualifications (secondary care) | 1 | 200 |
| 48 MM2- VI - image-based - Mixed qualifications (secondary/primary care) | 1 | 40 |
| 49 MM2- VI - image-based - Mixed qualifications (primary care) | 2 | 184 |
| 50 MM2 - VI - image-based qual not reported | 0 | 0 |
| 51 MM2 - Selected on quality - pathway 2 or 3 | 5 | 5728 |
| 52 MM2 - Selected on quality - pathway 5 | 9 | 3556 |

# Figures

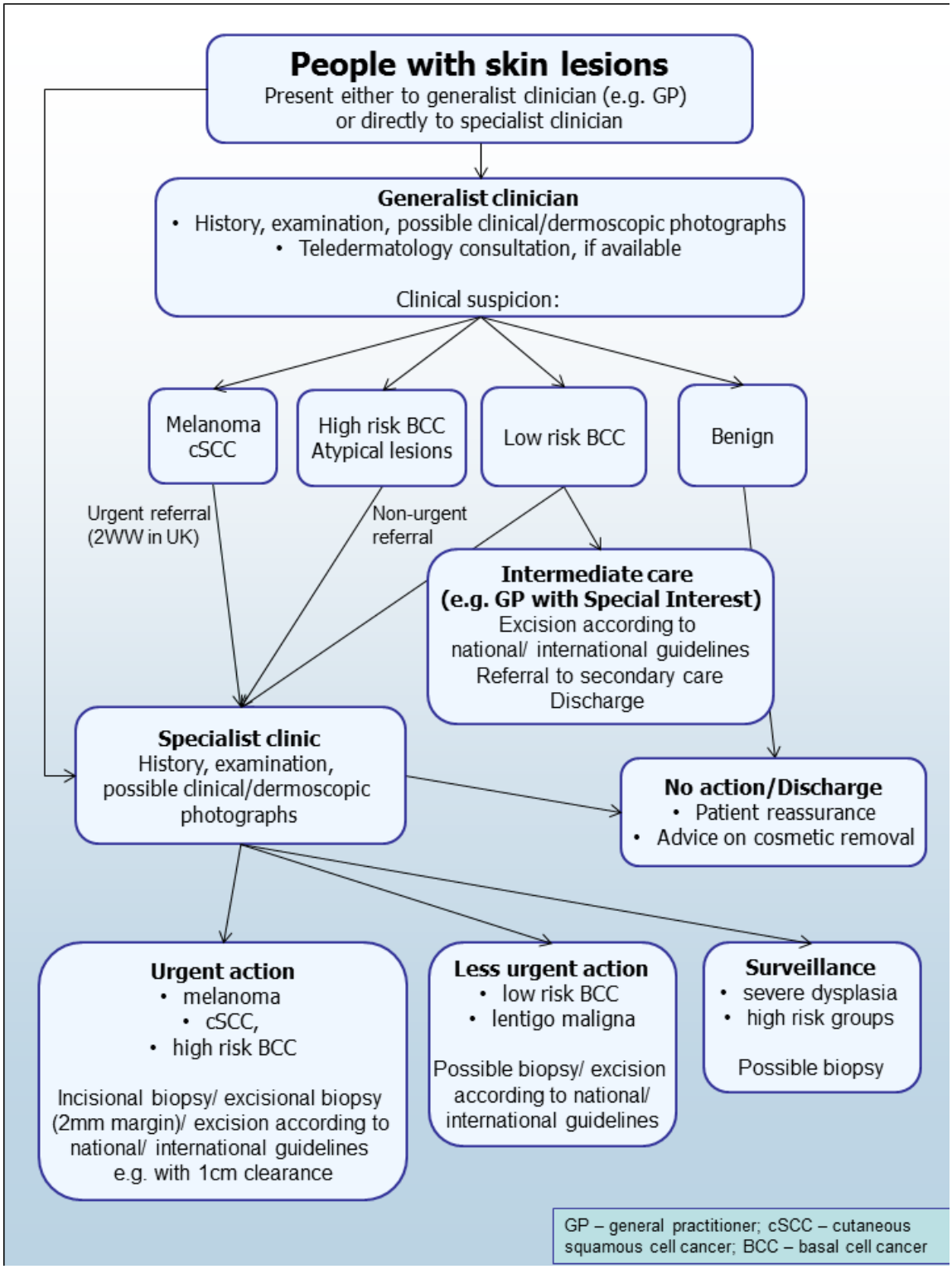
## Figure 1




*Caption*

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

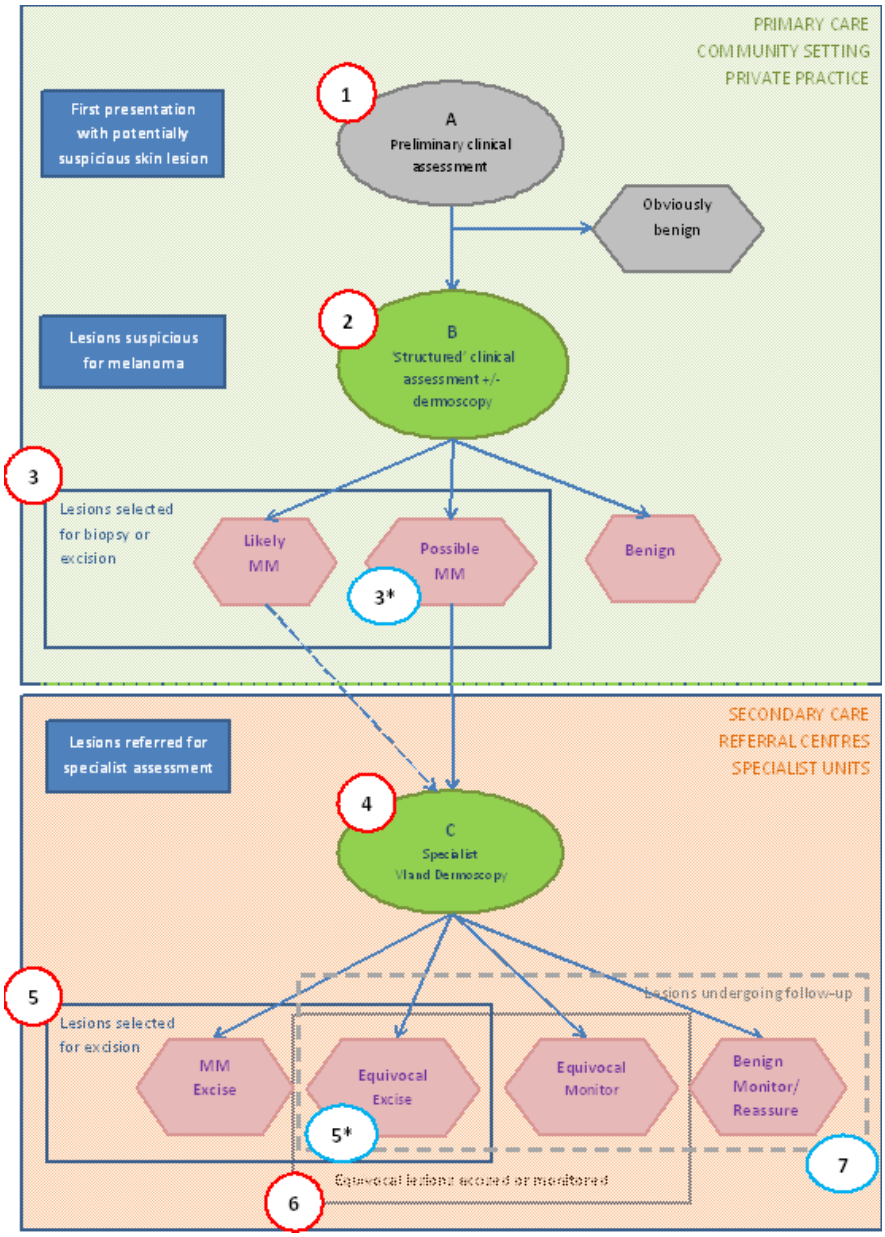
**Figure 2**




*Caption*

Current clinical pathway for people with skin lesions

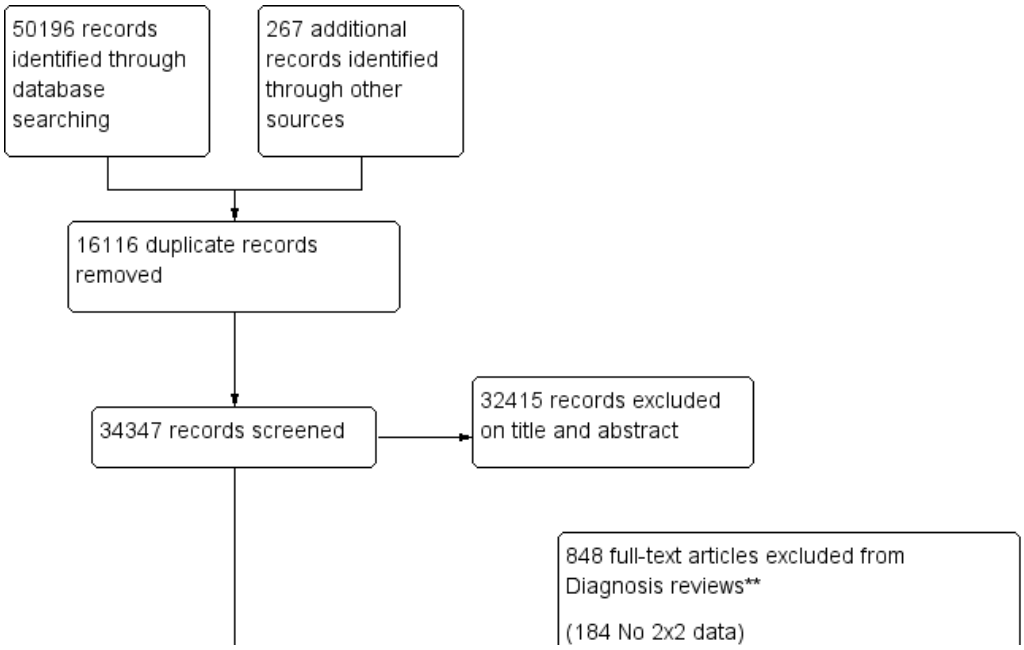
Figure 3

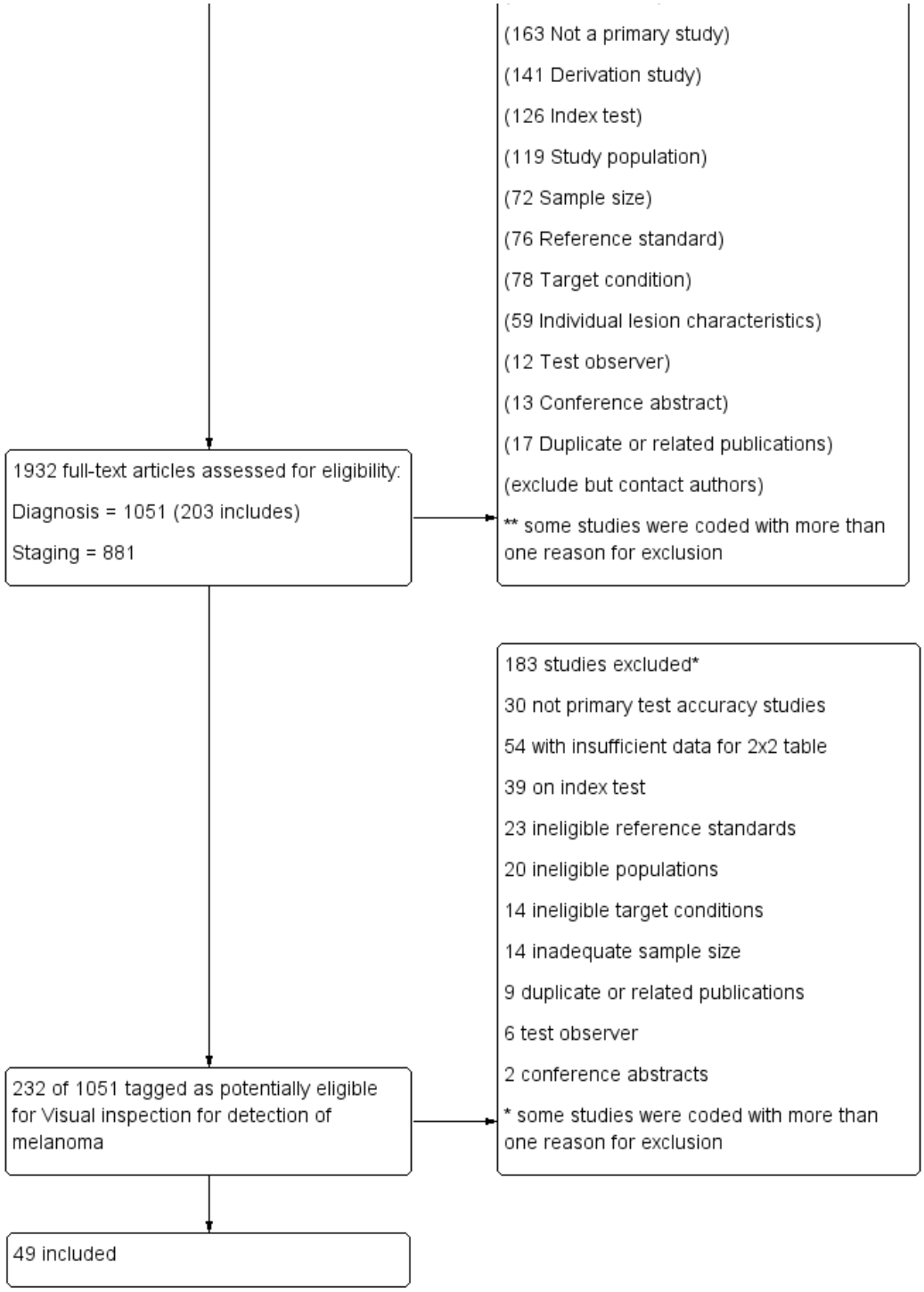



*Caption*

Clinical pathway

## Figure 4

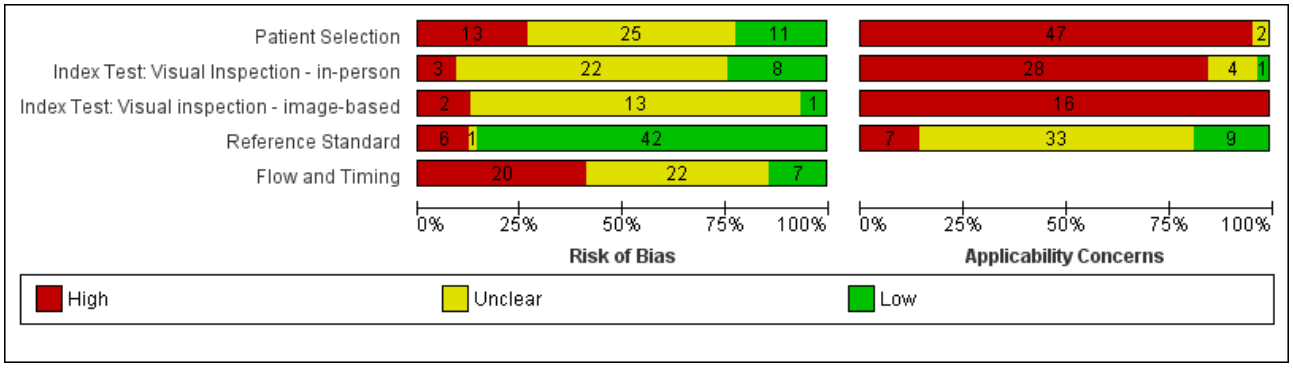





*Caption*

PRISMA flow diagram.

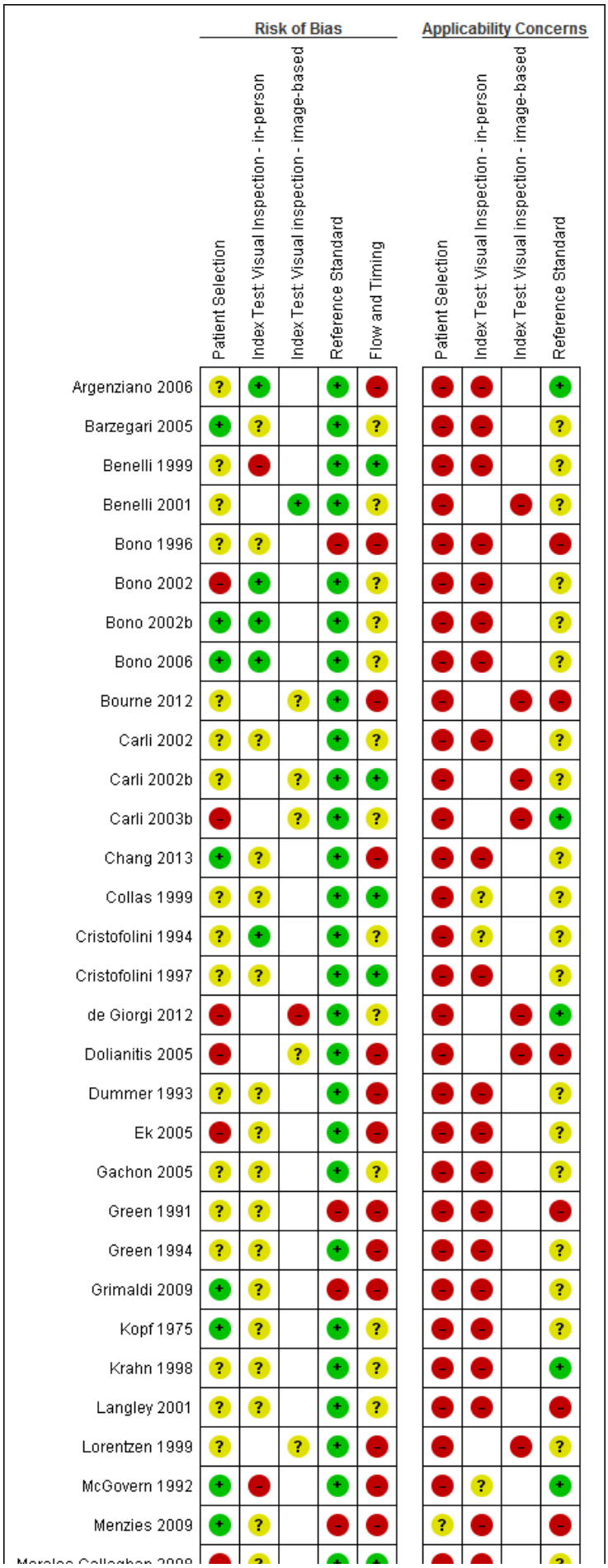
## Figure 5

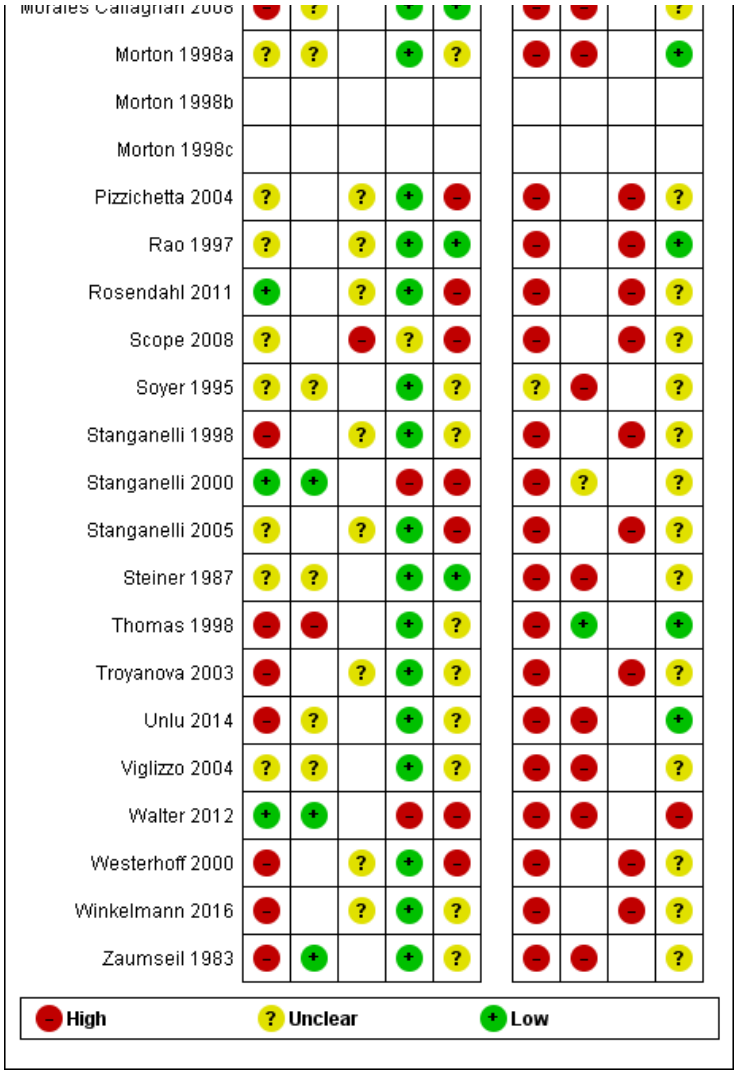



*Caption*

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

## Figure 6



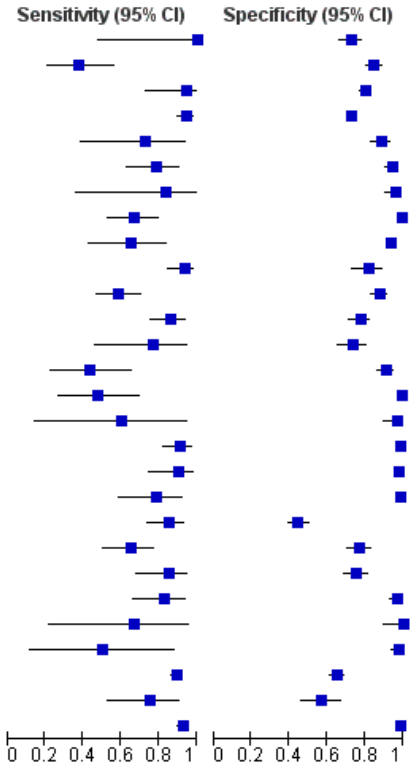



*Caption*

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

## Figure 7 (Analysis 7)

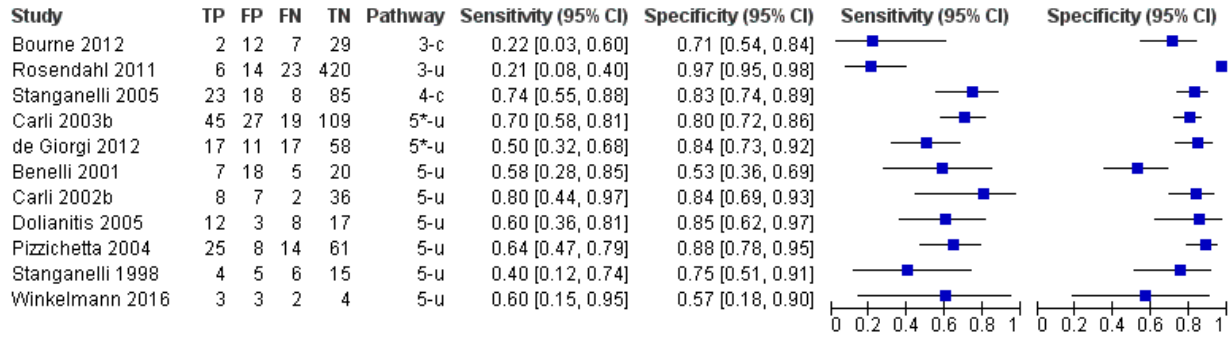
| Study | TP | FP | FN | TN | Pathway | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|
| Grimaldi 2009 | 5 | 63 | 0 | 167 | 2-c | 1.00 [0.48, 1.00] | 0.73 [0.66, 0.78] |
| Menzies 2009 | 12 | 46 | 20 | 253 | 2-c | 0.38 [0.21, 0.56] | 0.85 [0.80, 0.89] |
| Walter 2012 | 17 | 152 | 1 | 603 | 2-c | 0.94 [0.73, 1.00] | 0.80 [0.77, 0.83] |
| Gachon 2005 | 141 | 1058 | 8 | 2829 | 3-c | 0.95 [0.90, 0.98] | 0.73 [0.71, 0.74] |
| McGovern 1992 | 8 | 21 | 3 | 160 | 3-c | 0.73 [0.39, 0.94] | 0.88 [0.83, 0.93] |
| Collas 1999 | 30 | 19 | 8 | 296 | 3-u | 0.79 [0.63, 0.90] | 0.94 [0.91, 0.96] |
| Barzegari 2005 | 5 | 5 | 1 | 111 | 4-c | 0.83 [0.36, 1.00] | 0.96 [0.90, 0.99] |
| Stanganelli 2000 | 37 | 21 | 18 | 3296 | 4-c | 0.67 [0.53, 0.79] | 0.99 [0.99, 1.00] |
| Dummer 1993 | 15 | 49 | 8 | 699 | 5*-c | 0.65 [0.43, 0.84] | 0.93 [0.91, 0.95] |
| Soyer 1995 | 61 | 17 | 4 | 77 | 5*-c | 0.94 [0.85, 0.98] | 0.82 [0.73, 0.89] |
| Steiner 1987 | 43 | 31 | 30 | 214 | 5*-u | 0.59 [0.47, 0.70] | 0.87 [0.83, 0.91] |
| Bono 2002 | 57 | 56 | 9 | 191 | 5-c | 0.86 [0.76, 0.94] | 0.77 [0.72, 0.82] |
| Bono 2002b | 10 | 39 | 3 | 109 | 5-c | 0.77 [0.46, 0.95] | 0.74 [0.66, 0.81] |
| Bono 2006 | 10 | 16 | 13 | 167 | 5-c | 0.43 [0.23, 0.66] | 0.91 [0.86, 0.95] |
| Ek 2005 | 11 | 23 | 12 | 2536 | 5-c | 0.48 [0.27, 0.69] | 0.99 [0.99, 0.99] |
| Green 1991 | 3 | 2 | 2 | 63 | 5-c | 0.60 [0.15, 0.95] | 0.97 [0.89, 1.00] |
| Morton 1998a | 63 | 10 | 6 | 684 | 5-c | 0.91 [0.82, 0.97] | 0.99 [0.97, 0.99] |
| Morton 1998b | 28 | 14 | 3 | 522 | 5-c | 0.90 [0.74, 0.98] | 0.97 [0.96, 0.99] |
| Morton 1998c | 22 | 11 | 6 | 630 | 5-c | 0.79 [0.59, 0.92] | 0.98 [0.97, 0.99] |
| Benelli 1999 | 51 | 189 | 9 | 152 | 5-u | 0.85 [0.73, 0.93] | 0.45 [0.39, 0.50] |
| Carli 2002 | 35 | 46 | 19 | 156 | 5-u | 0.65 [0.51, 0.77] | 0.77 [0.71, 0.83] |
| Cristofolini 1994 | 28 | 46 | 5 | 141 | 5-u | 0.85 [0.68, 0.95] | 0.75 [0.69, 0.81] |
| Cristofolini 1997 | 29 | 4 | 6 | 137 | 5-u | 0.83 [0.66, 0.93] | 0.97 [0.93, 0.99] |
| Langley 2001 | 4 | 0 | 2 | 34 | 5-u | 0.67 [0.22, 0.96] | 1.00 [0.90, 1.00] |
| Morales Callaghan 2008 | 3 | 5 | 3 | 189 | 5-u | 0.50 [0.12, 0.88] | 0.97 [0.94, 0.99] |
| Thomas 1998 | 411 | 236 | 49 | 444 | 5-u | 0.89 [0.86, 0.92] | 0.65 [0.62, 0.69] |
| Unlu 2014 | 18 | 39 | 6 | 52 | 5-u | 0.75 [0.53, 0.90] | 0.57 [0.46, 0.67] |
| Zaumseil 1983 | 312 | 76 | 25 | 6650 | 5-u | 0.93 [0.89, 0.95] | 0.99 [0.99, 0.99] |



*Caption*

Forest plot of in-person evaluations of visual inspection for detection of invasive melanoma and melanocytic intra-epidermal variants by point on the clinical pathway where they are diagnosed

## Figure 8 (Analysis 8)



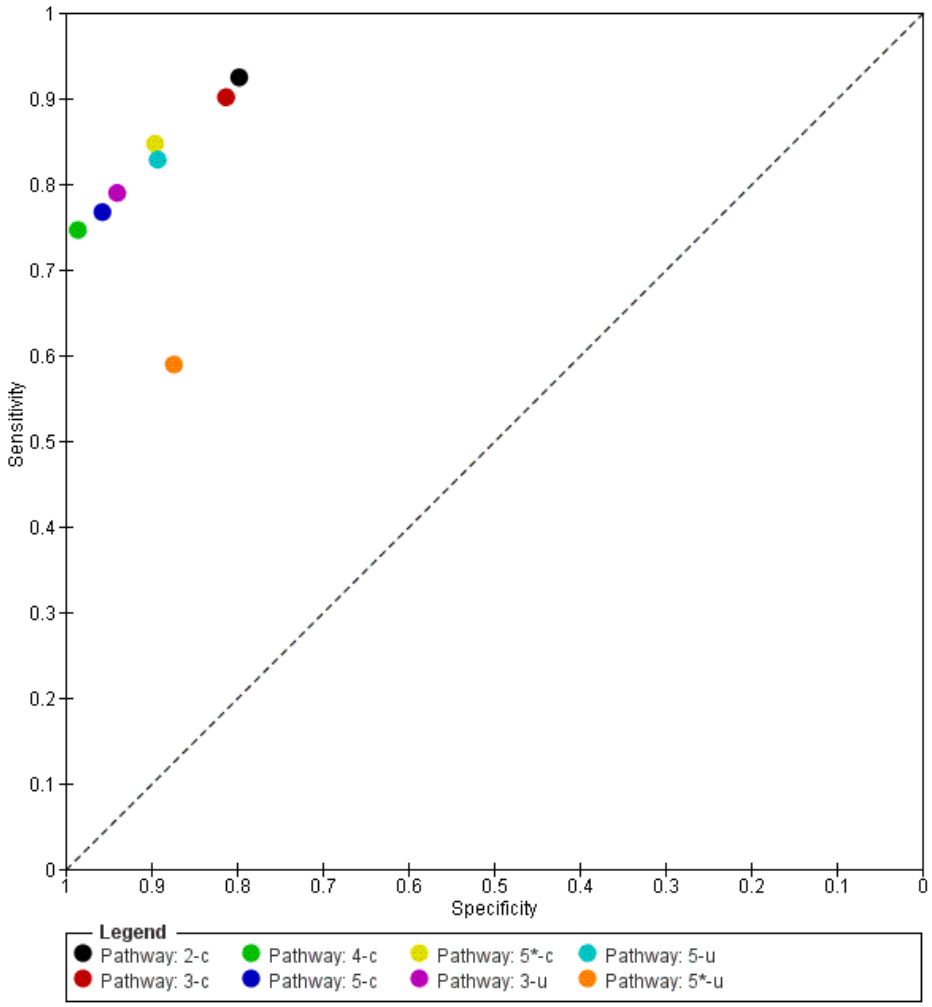
| Study | TP | FP | FN | TN | Pathway | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|
| Bourne 2012 | 2 | 12 | 7 | 29 | 3-c | 0.22 [0.03, 0.60] | 0.71 [0.54, 0.84] |
| Rosendahl 2011 | 6 | 14 | 23 | 420 | 3-u | 0.21 [0.08, 0.40] | 0.97 [0.95, 0.98] |
| Stanganelli 2005 | 23 | 18 | 8 | 85 | 4-c | 0.74 [0.55, 0.88] | 0.83 [0.74, 0.89] |
| Carli 2003b | 45 | 27 | 19 | 109 | 5*-u | 0.70 [0.58, 0.81] | 0.80 [0.72, 0.86] |
| de Giorgi 2012 | 17 | 11 | 17 | 58 | 5*-u | 0.50 [0.32, 0.68] | 0.84 [0.73, 0.92] |
| Benelli 2001 | 7 | 18 | 5 | 20 | 5-u | 0.58 [0.28, 0.85] | 0.53 [0.36, 0.69] |
| Carli 2002b | 8 | 7 | 2 | 36 | 5-u | 0.80 [0.44, 0.97] | 0.84 [0.69, 0.93] |
| Dolianitis 2005 | 12 | 3 | 8 | 17 | 5-u | 0.60 [0.36, 0.81] | 0.85 [0.62, 0.97] |
| Pizzichetta 2004 | 25 | 8 | 14 | 61 | 5-u | 0.64 [0.47, 0.79] | 0.88 [0.78, 0.95] |
| Stanganelli 1998 | 4 | 5 | 6 | 15 | 5-u | 0.40 [0.12, 0.74] | 0.75 [0.51, 0.91] |
| Winkelmann 2016 | 3 | 3 | 2 | 4 | 5-u | 0.60 [0.15, 0.95] | 0.57 [0.18, 0.90] |



*Caption*

Forest plot of image-based evaluations of visual inspection for detection of invasive melanoma and melanocytic intra-epidermal variants by point on the clinical pathway where they are diagnosed

## Figure 9 (Analysis 7)



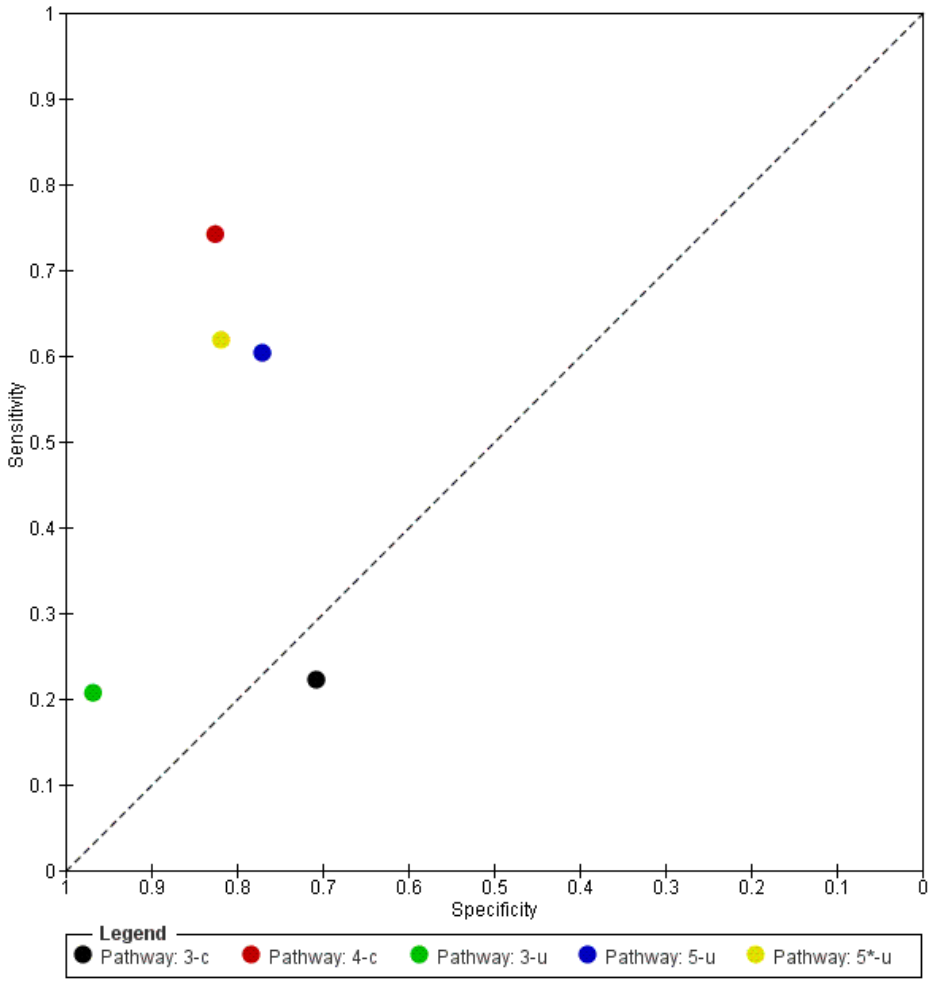


*Caption*

Summary estimates of accuracy of in-person visual inspection for the detection of invasive melanoma and melanocytic intra-epidermal variants by point on the clinical pathway where they are diagnosed

(confidence regions are not plotted due to small numbers of studies)

## Figure 10 (Analysis 8)

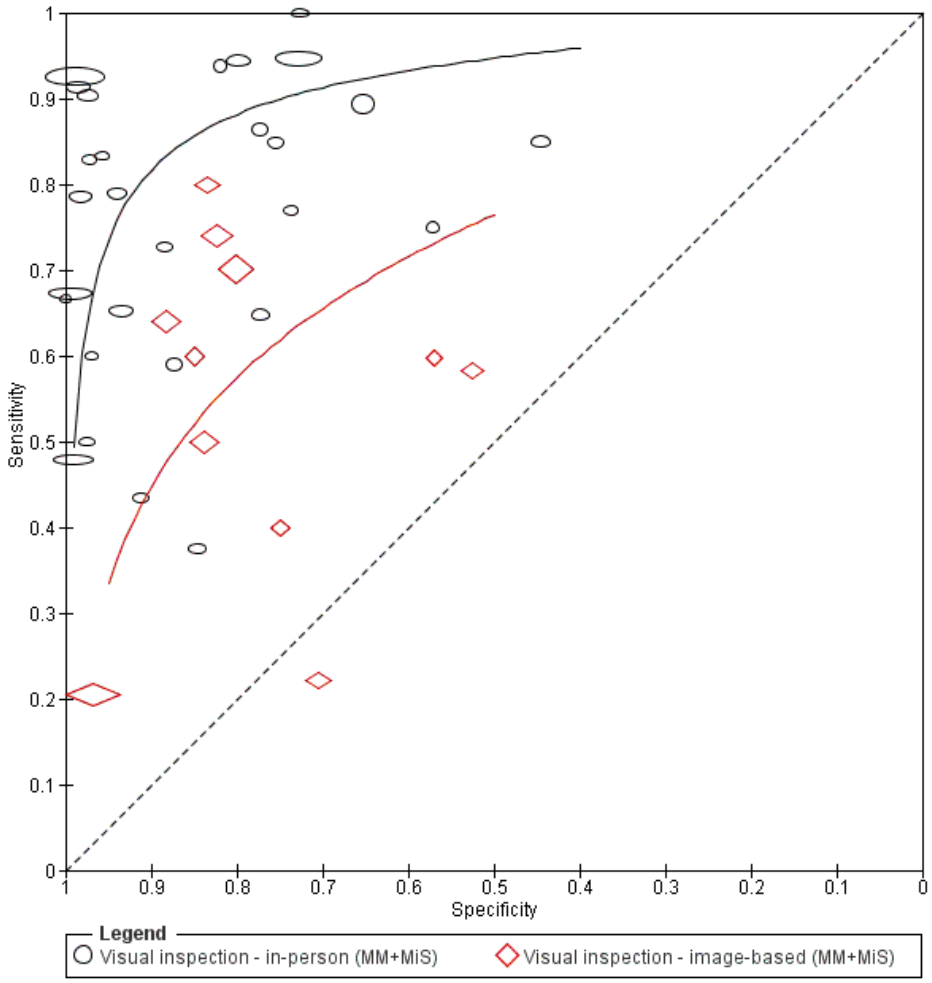



*Caption*

Summary estimates of accuracy of image-based visual inspection for the detection of invasive melanoma and melanocytic intra-epidermal variants by point on the clinical pathway where they are diagnosed

(confidence regions are not plotted due to small numbers of studies)

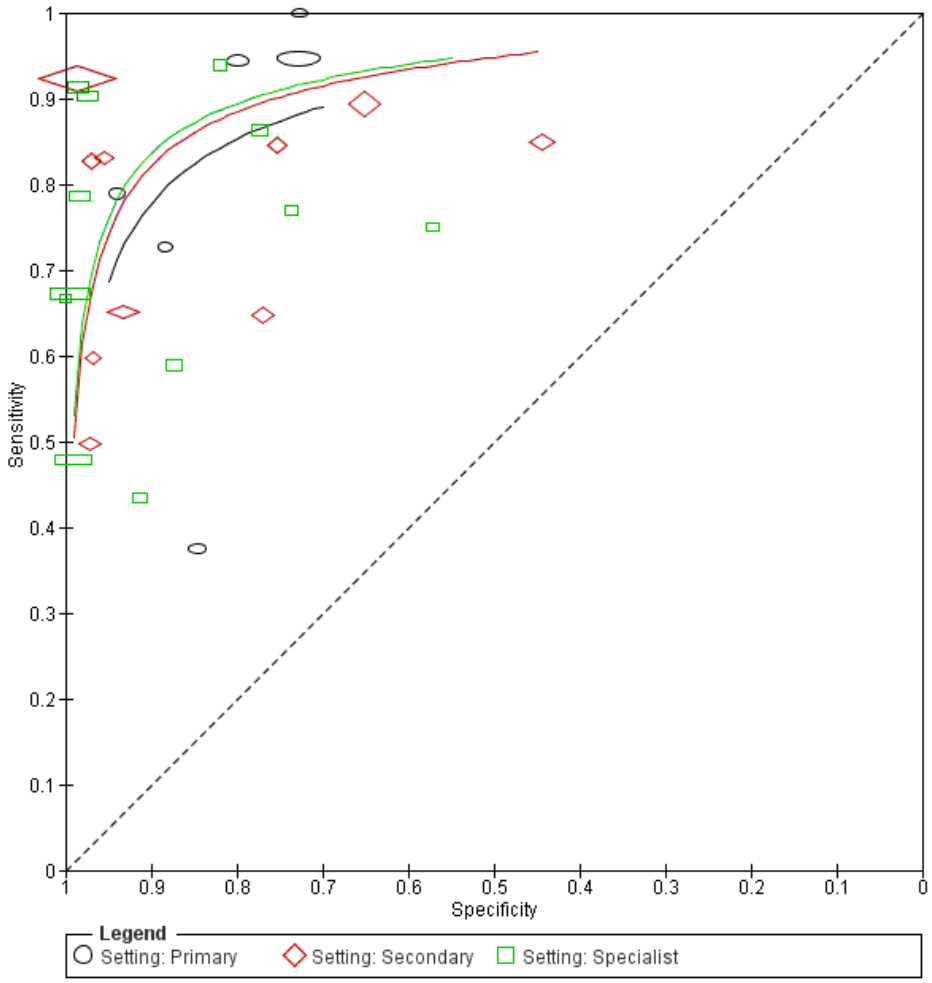
## Figure 11 (Analysis 2)




*Caption*

Summary ROC comparing in-person and image-based evaluations of visual inspection for detection of invasive melanoma or atypical intraepidermal melanocytic variants (MM+MiS).

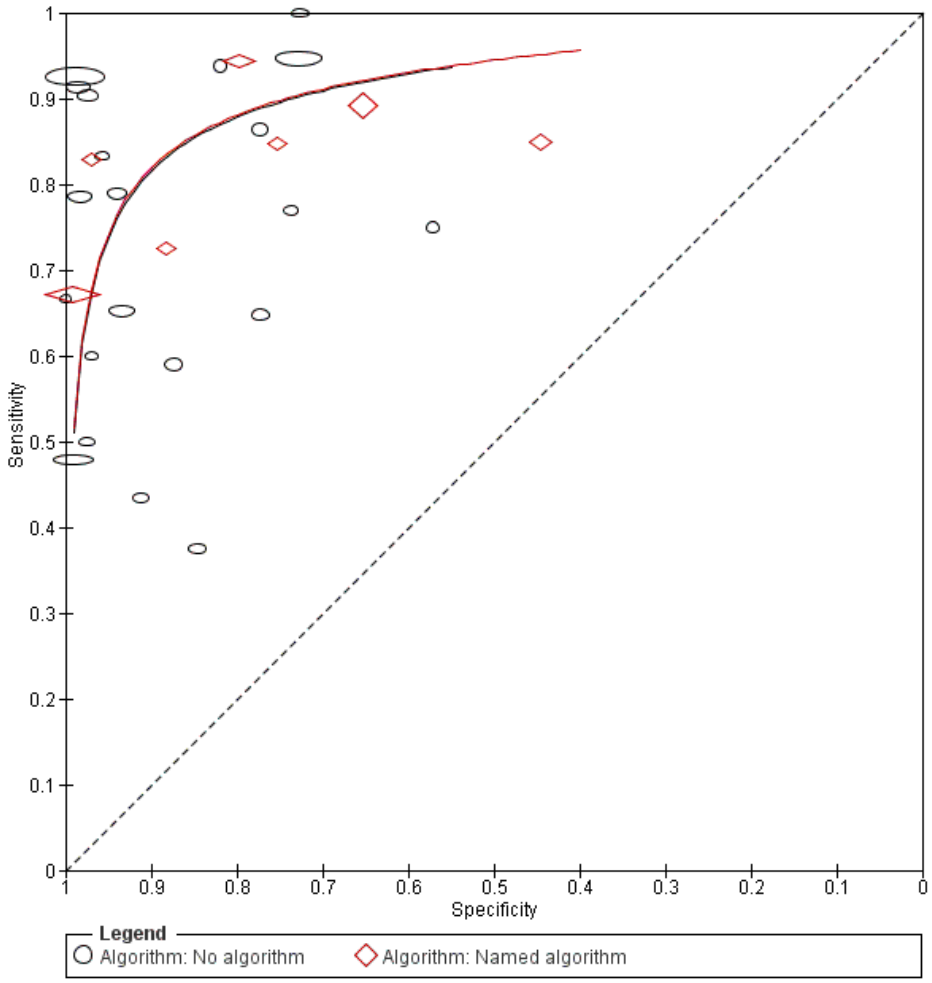
## Figure 12 (Analysis 5)




*Caption*

Summary ROC Plot of in-person visual inspection evaluations stratified by study setting for detection of invasive melanoma and atypical intraepidermal melanocytic variants (MM+MiS)

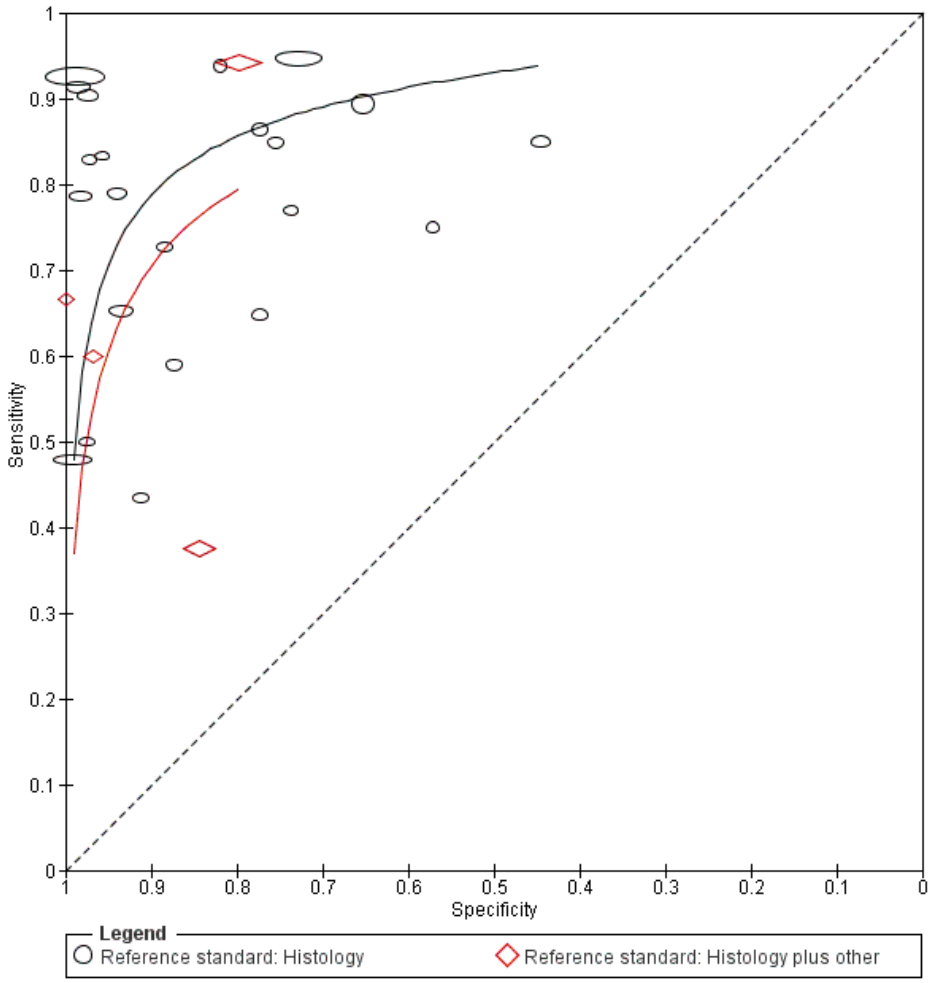
## Figure 13 (Analysis 6)




*Caption*

Summary ROC Plot of in-person visual inspection evaluations stratified by use of a published algorithm for detection of invasive melanoma and atypical intraepidermal melanocytic variants (MM+MiS)

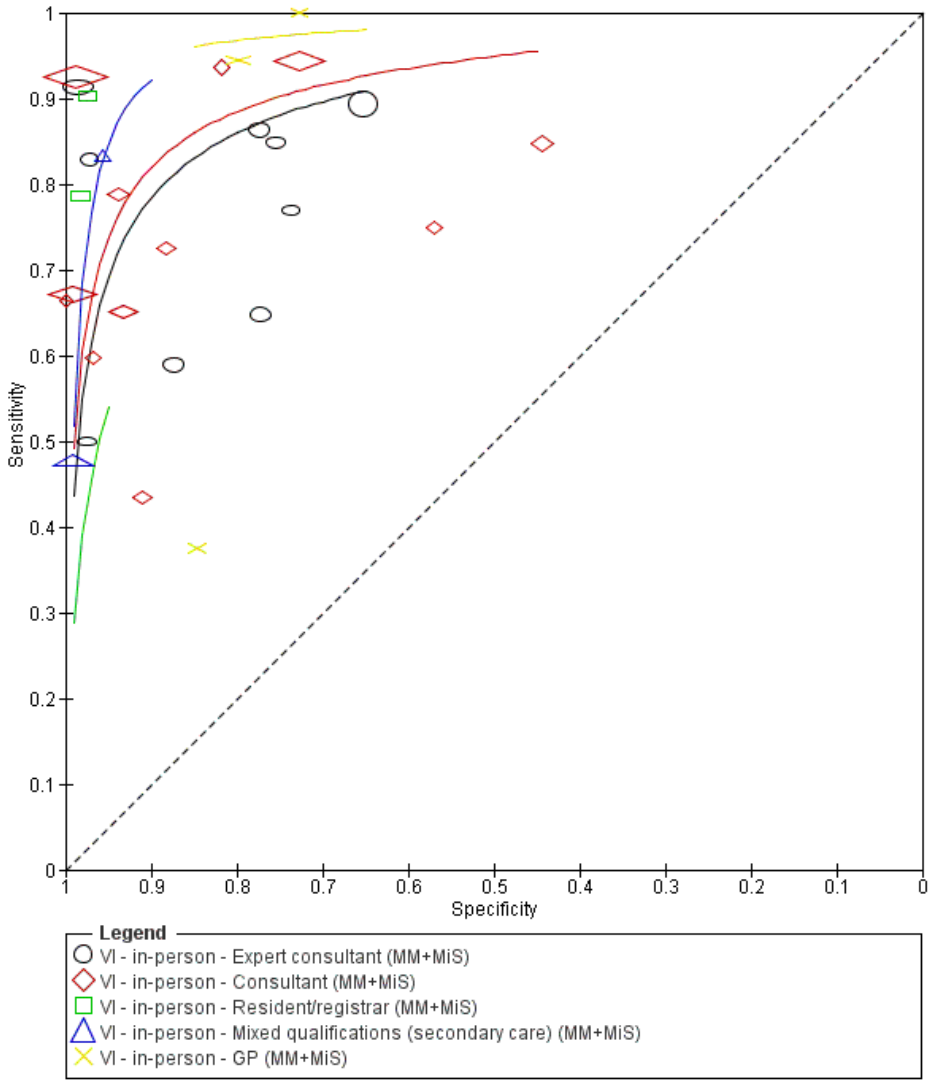
## Figure 14 (Analysis 3)




*Caption*

Summary ROC Plot of in-person visual inspection evaluations stratified by reference standard for detection of invasive melanoma and atypical intraepidermal melanocytic variants (MM+MiS)

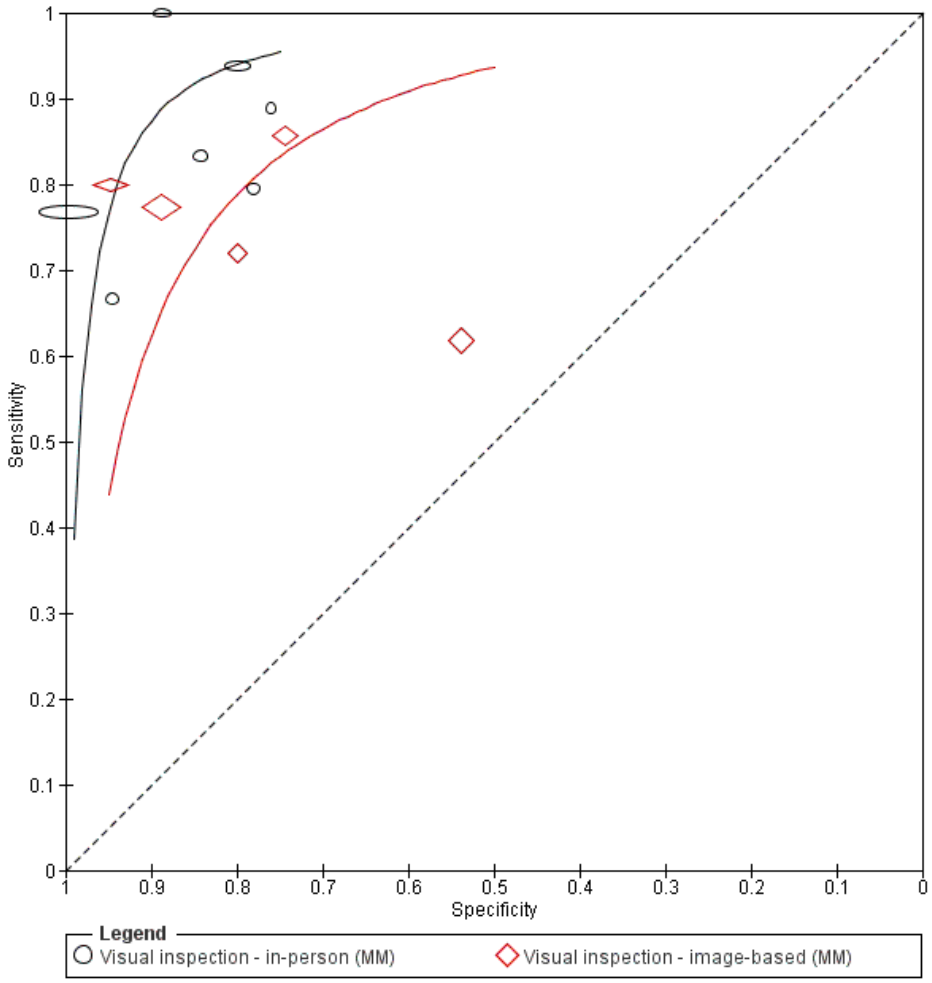
## Figure 15 (Analysis 13)




*Caption*

Summary ROC Plot of in-person visual inspection (VI) evaluations stratified by observer experience and qualifications for detection of invasive melanoma and atypical intraepidermal melanocytic variants (MM+MiS)

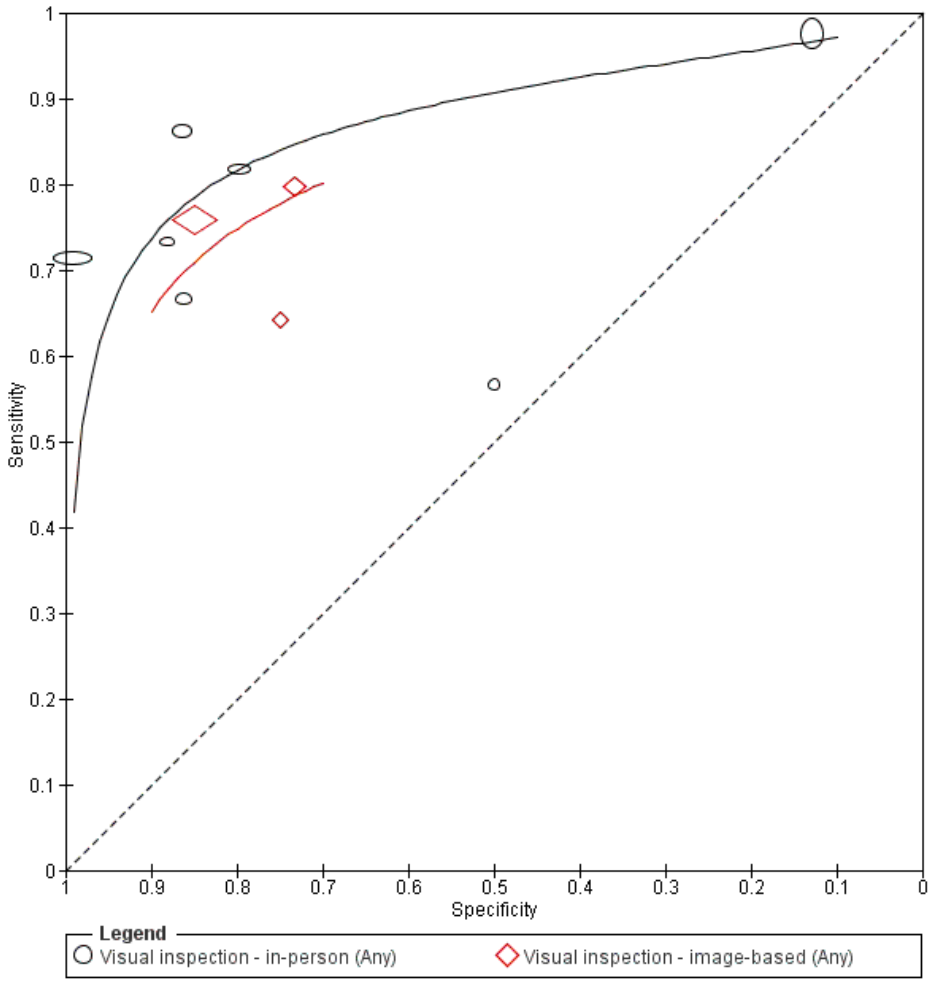
## Figure 16 (Analysis 9)




*Caption*

Summary ROC comparing in-person and image-based evaluations of visual inspection for detection of invasive melanoma alone (MM).

## Figure 17 (Analysis 10)




*Caption*

Summary ROC comparing in-person and image-based evaluations of visual inspection for detection of any skin lesion requiring excision (Any)

# 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 | |
|---|---|
| **Diagnosis of melanoma** | **Estimated number of studies** |
| 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 Table of acronyms and abbreviations used

| Acronym | Definition |
|---|---|
| 3PCL | three point checklist |
| 7FFM | seven features for melanoma |
| 7PCL | seven point checklist |
| ABCD(E) | asymmetry, border, colour, differential structures (enlargement) |
| AHM | amelanotic or hypomelanotic melanoma |
| AK | actinic keratosis |
| AMN | atypical melanocytic naevi |
| AUC | area under the curve |
| BCC | basal cell carcinoma |
| BD | Bowen's disease |
| BN | benign naevi |
| BNM | benign non-melanocytic |
| BPC | between person comparison (of tests) |
| CAD | computer assisted diagnosis |
| CCS | case control study |
| CD | compact disc |
| CM | cutaneous melanoma |
| CMM | cutaneous malignant melanoma |
| CS | case series |
| CSCC | cutaneous squamous cell carcinoma |
| D- | disease negative |
| D+ | disease positive |
| DF | dermatofibroma |
| Dx | diagnosis |
| ELM | epiluminescence microscopy |
| FN | false negative |
| FP | false positive |
| FU | Follow- up |
| GP | general practitioner |
| H&E | haematoxylin and eosin stain |
| LPLK | lichen planus- like keratosis |
| LS | lentigo simplex |
| MiS | melanoma in situ (or lentigo maligna) |
| MM | malignant melanoma |
| MN | melanocytic naevi |
| MSDSLA | multispectral digital skin lesion analysis device |

| Acronym | Definition |
|---|---|
| N/A | not applicable |
| NC | non comparative |
| NMLs | non melanocytic lesions |
| NPV | negative predictive value |
| NR | not reported |
| P | prospective |
| PCPs | primary care providers |
| PLC | pigmented lesion clinic |
| PPV | positive predictive value |
| PSL | pigmented skin lesion |
| R | retrospective |
| RCM | reflectance confocal microscopy |
| RCT | randomised controlled trial |
| SCC | squamous cell carcinoma |
| SD | standard deviation |
| SDDI | Short term sequential digital dermoscopy imaging |
| se | sensitivity |
| sp | specificity |
| SK | seborrhoeic keratosis |
| SN | Spitz nevi |
| SSM | superficial spreading melanoma |
| SVS | support vector system |
| TD | teledermatology |
| TN | true negative |
| TWR | two week rule |
| VI | visual inspection |
| WPC | within person comparison (of tests) |
| WPC-algs | within person comparison (of algorithms) |

## 4 Content of algorithms used to assist melanoma diagnosis by visual inspection alone

| **ABCD** (Friedman 1985; Rigel 1993; Pehamberger 1993)<br>**ABCDE** (Carli 1994; Cristofolini 1994; Thomas 1998; Benelli 1999; Benelli 2001; Abbasi 2004)<br>**BCD** (McGovern 1992) | **Seven-point checklist** ( MacKie 1985; MacKie 1990; Keefe 1990) | **Seven-point checklist (revised)** (MacKie 1990; Healsmith 1994) |
|---|---|---|
| A – asymmetry<br>• variable centripetal growth of melanocytes (Friedman 1985)<br>• “geometrical asymmetry in two axes of the tumour” (Thomas 1998; Benelli 1999; Benelli 2001)<br>• “one half does not match the other half” (McGovern 1992); not separately scored in study “because we believed that asymmetry and border | • sensory change (greater awareness of the lesion or mild itch);<br>• diameter of >=1 cm;<br>• growth of the lesion;<br>• an irregular edge;<br>• irregular pigment with different shades of brown and black in the lesion; | MacKie 1990, Mackie 1991, and Healsmith 1994 describe the revised criteria as:<br>Major signs<br>• Change in size<br>• Change in shape<br>• Change in colour<br>Minor signs<br>• Inflammation<br>Crusting or bleeding |

| | | |
|---|---|---|
| irregularity were linked”<br>B - irregular borders<br>• irregular shape with notching or scalloping of border (Friedman 1985)<br>• “edges are ragged, notched, or blurred (McGovern 1992)<br>• “irregular and notched” (Cristofolini 1994)<br>• “unsharp or ill-defined or angular” ( Thomas 1998)<br>• “ragged or indented” (Benelli 1999, Benelli 2001)<br>C - colour<br>• variable pigmentation, multiple colours; various of hues of brown, also black, blue, red and white (Friedman 1985 )<br>• “pigmentation is not uniform; shades of tan, brown and black are present with dashes of red, white, or blue” (McGovern 1992)<br>• “mottled-haphazard display” (Cristofolini 1994)<br>• “presence of at least two different colours within the lesion (with the exception of the usual symmetrical darkening of the lesion in its center)” (Thomas 1998; Benelli 2001)<br>• “multiple colours” (Abbasi 2004)<br>D - diameter equal or superior to 6mm<br>• all studies agree<br>E - evolution<br>• “changes in pigmentation” (Cristofolini 1994)<br>• “enlargement of the surface (and not in height) of the lesion; anamnestic criterion based on the patient’s description of the natural history of the lesion” (Thomas 1998)<br>• “elevation, enlargement or change in the color of the lesion” (Benelli 1999; Benelli 2001)<br>• “evolving (with respect to size, shape, shades of colour, surface features, or symptoms)” (Abbasi 2004)<br>McGovern 1992 describes 7 characteristics as: “increasing size, variegation, inflammation, irregular outline, greater than 1cm diameter, itch, bleeding”<br>These are expanded on in MacKie 1990, who describes the original (1985) criteria as:<br>• sensory change, often described as a greater awareness of the lesion but also as a mild itch;<br>• diameter of 1 cm or greater;<br>• growth of the lesion;<br>• an irregular edge;<br>• irregular pigment with different shades of brown and black in the lesion;<br>• inflammation (a reddish tinge within the lesion); and<br>• crusting, oozing, or bleeding. | • inflammation<br>• crusting, oozing, or bleeding.<br>Presence of 3 or more suggestive of melanoma | •<br>• Sensory change<br>• Diameter >=7 mm<br>“a patient with a pigmented lesion with any one of the major signs should be considered for referral and that the presence of any of the minor signs should be a further stimulus to referral.” (MacKie 1990) |

| | | |
|---|---|---|
| • >=3 criteria should prompt referral ( MacKie 1990) | | |
| | | |

## 5 Proposed sources of heterogeneity

### *i. Population characteristics*

- general versus higher risk populations
- patient population: primary/secondary/specialist unit
- lesion suspicion: general suspicion/atypical/equivocal/NR
- lesion type: any pigmented; melanocytic
- inclusion of multiple lesions per participant
- ethnicity

### *ii. Index test characteristics*

- the nature of and definition of criteria for test positivity
- observer experience with the index test
- approaches to lesion preparation (e.g. the use of oil or antiseptic gel for dermoscopy)

### *iii. Reference standard characteristics*

- reference standard used
- whether histology-reporting meets pathology-reporting guidelines
- use of excisional versus diagnostic biopsy
- whether two independent dermatopathologists reviewed histological diagnosis

### *iv. Study quality*

- consecutive or random sample of participants recruited
- index test interpreted blinded to the reference standard result
- index test interpreted blinded to the result of any other index test
- presence of partial or differential verification bias (whereby only a sample of those subject to the index test are verified by the reference test or by the same reference test with selection dependent on the index test result)
- use of an adequate reference standard
- overall risk of bias

## 6 Final search strategies

**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)

## 7 Full text inclusion criteria

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

| Criterion | Inclusion | Exclusion |
|---|---|---|
| Target condition | • Melanoma<br>• Keratinocyte skin cancer (or non-melanoma skin cancer)<br>  ◦ BCC or epithelioma<br>  ◦ cSCC | • Studies exclusively conducted in children<br>• Studies of non-cutaneous melanoma or SCC |
| Population | For diagnostic reviews<br>• Adults with a skin lesion suspicious for melanoma, BCC, or cSCC (other terms include pigmented skin lesion/nevi, melanocytic, keratinocyte, etc.)<br>• Adults at high risk of developing melanoma skin cancer, BCC, or cSCC<br>For staging reviews<br>• 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<br>• Studies conducted exclusively in children |
| Index tests | For diagnosis<br>• Visual inspection/clinical examination<br>• Dermoscopy/dermatoscopy<br>• Teledermoscpoy<br>• Smartphone/mobile phone applications<br>• Digital dermoscopy/artificial intelligence<br>• Confocal microscopy<br>• Ocular coherence tomography<br>• Exfoliative cytology<br>• High frequency ultrasound<br>• Canine odour detection<br>• DNA expression analysis/gene chip analysis<br>• Other<br>For staging<br>• CT<br>• PET<br>• PET-CT<br>• MRI<br>• Ultrasound +/fine needle aspiration cytology FNAC<br>• SLNB +/high frequency ultrasound<br>• Other<br>Any test combination and in any order<br>Any test positivity threshold<br>Any variation in testing procedure (e.g. radioisotope used) | • Sentinel lymph biopsy for therapeutic rather than staging purposes<br>• Tests to determine melanoma thickness<br>• Tests to determine surgical margins/lesion borders<br>• Tests to improve histopathology diagnose<br>• LND |

| Criterion | Inclusion | Exclusion |
|---|---|---|
| Reference standard | For diagnostic studies<br>• Histopathology of the excised lesion<br>• Clinical follow-up of non-excised/benign appearing lesions with later histopathology if suspicious<br>• Expert diagnosis (studies should not be included if expert diagnosis is the sole reference standard)<br>For studies of imaging tests for staging<br>• Histopathology (via LND or SLMB)<br>• Clinical/radiological follow-up<br>• A combination of the above<br>For studies of SLNB accuracy for staging<br>• LND of both SLN+ and SLn participants to identify all diseased nodes<br>• 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<br>• Exclude if any disease positive participants have diagnosis unconfirmed by histology<br>• Exclude if > 50% of disease negative participants have diagnosis confirmed by expert opinion with no histology or follow-up<br>• 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.

## 8 Quality assessment (based on QUADAS-2)

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

### Patient 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 study to be less representative than studies conducted in a more general population 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 a clinical practice setting, i.e. in-person or face-to-face with the patient, and by a single observer as opposed to a consensus decision or average across multiple observers. Image-based studies were considered to be high concern, although RCM image interpretations where the observer was also supplied with a clinical or dermoscopic image of the lesion

along with some patient characteristics were considered ‘unclear’.

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.

The experience of the examiner will also impact on the applicability of study results. We required studies to describe the test interpreter as ‘experienced’ or ‘expert’ in RCM to have low concern about applicability.

### 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.

### Comparative domain

A comparative domain was added to the QUADAS-2 checklist for studies comparing the accuracy of RCM and dermoscopy. Items were included to assess the presence blinding of interpretation between tests, and to specify a maximum of one month interval between application of index tests, as intervals greater than these may be accompanied by changes in tumour characteristics. As it would not be normal practice for RCM to be interpreted blinded to the clinical or dermoscopic diagnosis, the scoring of this item did not contribute to our overall assessment of risk of bias. We also considered whether both tests were applied and interpreted in a clinically applicable manner.

*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<br>**No** – if paper describes other method of sampling<br>**Unclear** – if participant sampling not described |

| Item | Response (delete as required) |
|---|---|
| **PARTICIPANT SELECTION (1) RISK OF BIAS** | |
| 2) Was a case-control design avoided? | **Yes** – if consecutive or random or case-control design clearly not used<br>**No** – if study described as case-control or describes sampling specific numbers of participants with particular diagnoses<br>**Unclear** – if not described |
| 3) Did the study avoid inappropriate exclusions, e.g.<br>• 'difficult to diagnose' lesions not excluded<br>• lesions not excluded on basis of disagreement between evaluators | **Yes** if inappropriate exclusions were avoided<br>**No** – if lesions were excluded that might affect test accuracy, e.g. 'difficult to diagnose' lesions, or where disagreement between evaluators was observed<br>**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):<br>• **A)** were the same participant selection criteria used for those allocated to each test?<br>• **B)** was the potential for biased allocation between tests avoided through adequate generation of a randomised sequence?<br>• **C)** was the potential for biased allocation between tests avoided through concealment of allocation prior to assignment? | For A)<br>• **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<br>For B)<br>• **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<br>For C)<br>• **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?<br>For non-comparative and within-person-comparative studies<br>1. If answers to all of questions 1), 2), and 3) 'Yes':<br>2. If answers to any 1 of questions 1), 2), or 3) 'No':<br>3. If answers to any 1 of questions 1), 2), or 3) 'Unclear':<br>For between-person comparative studies<br>1. If answers to all of questions 1), 2), 3), and 4) 'Yes':<br>2. If answers to any 1 of questions 1), 2), 3), or 4) 'No':<br>3. If answers to any 1 of questions 1), 2), 3), or 4) 'Unclear': | For non-comparative and within-person-comparative studies<br>1. Risk is low<br>2. Risk is high<br>3. Risk unclear<br>For between-person comparative studies<br>1. Risk is low<br>2. Risk is high<br>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?<br>• 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<br>• 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)**<br>**Yes** – if participants included in the study appear to be generally representative of those who might present in a usual practice setting<br>**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<br>**Unclear –** if insufficient details are provided to determine the generalisability of study participants<br>**B) For studies that will contribute to the analysis of referred participants (i.e. who have already undergone some form of testing)**<br>**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<br>**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<br>**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%<br>**No** – if the difference between the number of included lesions and number of included participants is greater than 5%<br>**Unclear –** if it is not possible to assess |
| Is there concern that the included participants do not match the review question?<br>1. If the answer to question 1) or 2) 'Yes':<br>2. If the answer to question 1) or 2) 'No':<br>3. If the answer to question 1) or 2) 'Unclear': | 1. Concern is low<br>2. Concern is high<br>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<br>**No** – if index test described as interpreted in knowledge of reference standard result<br>**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)<br>**No** – if threshold was not prespecified<br>**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<br>**No** – if the index tests were described as interpreted in the knowledge of the results of the others<br>**Unclear** – if it is not possible to tell whether knowledge of other index tests could have influenced test interpretation<br>**N/A** – if only 1 index test was evaluated |
| Could the conduct or interpretation of the index test have introduced bias?<br>**For non-comparative and between-person comparison studies**<br>1. If answers to questions 1) and 2) 'Yes':<br>2. If answers to either questions 1) or 2) 'No':<br>3. If answers to either questions 1) or 2) 'Unclear':<br>**For within-person comparative studies**<br>1. If answers to all questions 1), 2), for any index test and 3) 'Yes':<br>2. If answers to any 1 of questions 1) or 2) for any index test or 3) 'No':<br>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**<br>1. Risk is low<br>2. Risk is high<br>3. Risk is unclear<br>**For within-person comparative studies**<br>1. Risk is low<br>2. Risk is high<br>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?<br>E.g. previously evaluated/established<br>• algorithm/checklist used<br>• lesion characteristics indicative of melanoma used<br>• 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<br>**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<br>**Unclear** – if insufficient information was reported |
| 2) Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?<br>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<br>**No** – if the criteria for diagnosis of melanoma were not reported in sufficient detail to allow replication<br>**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<br><br>**No** – if the test was not interpreted by an experienced examiner (see above)<br><br>**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<br><br>**N/A** – if system-based diagnosis, i.e. no observer interpretation |
| Is there concern that the index test, its conduct, or interpretation differ from the review question?<br><br>1. If answers to questions 1), 2), and 3) 'Yes':<br>2. If answers to questions 1), 2), or 3) 'No':<br>3. If answers to questions 1), 2), or 3) 'Unclear': | 1. Concern is low<br>2. Concern is high<br>3. Concern is unclear |
| **REFERENCE STANDARD (3) RISK OF BIAS** | |
| 1) Is the reference standard likely to correctly classify the target condition?<br><br>A) Disease-positive – 1 or more of the following:<br><br>• histological confirmation of melanoma following biopsy or lesion excision<br>• 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<br><br>B) Disease-negative – 1 or more of the following:<br><br>• histological confirmation of absence of melanoma following biopsy or lesion excision in at least 80% of disease-negative participants<br>• 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<br><br>**Yes** – if all participants with a final diagnosis of melanoma underwent 1 of the listed reference standards<br><br>**No** – If a final diagnosis of melanoma for any participant was reached without histopathology<br><br>**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<br><br>B) Disease-negative<br><br>**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<br><br>**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<br><br>**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?<br><br>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<br><br>**No** – if the reference standard diagnosis was reached with knowledge of the index test result<br><br>**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?<br>**For visual inspection/dermoscopy evaluations**<br>1. If answer to question 1) 'Yes':<br>2. If answer to question 1) 'No':<br>3. If answer to question 1) 'Unclear':<br>**For all other tests**<br>1. If answers to questions 1) and 2) 'Yes':<br>2. If answers to questions 1) or 2) 'No':<br>3. If answers to questions 1) or 2) 'Unclear': | **For visual inspection/dermoscopy evaluations**<br>1. Risk is low<br>2. Risk is high<br>3. Risk is unclear<br>**For all other tests**<br>1. Risk is low<br>2. Risk is high<br>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<br>**No –** if index test results for the different components of the target condition cannot be disaggregated<br>**Unclear** – if not clearly reported |
| 2) Expert opinion (with no histological confirmation) was not used as a reference standard<br>'Expert opinion' means diagnosis based on the standard clinical examination, with no histology or lesion follow-up<br>***do not complete this item for teledermatology studies | **Yes** – if expert opinion was not used as a reference standard for any participant<br>**No** – if expert opinion was used as a reference standard for any participant<br>**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<br>**No –** if histology interpretation was reported to be carried out by a less experienced histopathologist<br>**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?<br>1. If answers to all questions 1), 2), and 3) 'Yes':<br>2. If answers to any 1 of questions 1), 2), or 3) 'No':<br>3. If answers to any 1 of questions 1), 2), or 3) 'Unclear':<br>*****For teledermatology studies only**<br>1. If answers to all questions 1) and 3) 'Yes':<br>2. If answers to questions 1) or 3) 'No':<br>3. If answers to questions 1) or 3) 'Unclear': | 1. Concern is low<br>2. Concern is high<br>3. Concern is unclear<br>*****For teledermatology studies only**<br>1. Concern is low<br>2. Concern is high<br>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?<br>**A)** For histopathological reference standard, was the interval between index test and reference standard ≤ 1 month?<br>**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)**<br>**Yes** – if study reports ≤ 1 month between index and reference standard<br>**No** – if study reports > 1 month between index and reference standard<br>**Unclear** – if study does not report interval between index and reference standard<br>**B)**<br>**Yes** – if study reports ≥ 3 months' follow-up<br>**No** – if study reports < 3 months' follow-up<br>**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<br>**No** – if more than 1 reference standard was used<br>**Unclear** – if not clearly reported |
| 3) Were all participants included in the analysis? | **Yes** – if all participants were included in the analysis<br>**No** – if some participants were excluded from the analysis<br>**Unclear**– if not clearly reported |
| 4) **For within-person comparisons of index tests**<br>Was the interval between application of index tests ≤ 1 month? | **Yes** – if study reports ≤ 1 month between index tests<br>**No** – if study reports > 1 month between index tests<br>**Unclear** – if study does not report the interval between index tests |
| Could the participant flow have introduced bias?<br>**For non-comparative and between-person comparison studies**<br>1. If answers to questions 1), 2), and 3) 'Yes':<br>2. If answers to any 1 of questions 1), 2), or 3) 'No':<br>3. If answers to any 1 of questions 1), 2), or 3) 'Unclear':<br>**For within-person comparative studies**<br>1. If answers to all questions 1), 2), 3), and 4) 'Yes':<br>2. If answers to any 1 of questions 1), 2), 3), or 4) 'No':<br>3. If answers to any 1 of questions 1), 2), 3), or 4) 'Unclear': | **For non-comparative and between-person comparison studies**<br>1. Risk is low<br>2. Risk is high<br>3. Risk is unclear<br>**For within-person comparative studies**<br>1. Risk is low<br>2. Risk is high<br>3. Risk is unclear |
| BCC = basal cell carcinoma; cSCC = cutaneous squamous cell carcinoma. | |

## 9 Summary study details – in-person evaluations

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MEL) | Exclusio |
|---|---|---|---|---|---|---|---|---|
| **2 – Limited prior testing** | | | | | | | | |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MEL) | Exclusio |
|---|---|---|---|---|---|---|---|---|
| Grimaldi 2009<br>Pathway - clear<br>MM+MiS | WPC<br>P-CS<br>Italy<br>Primary | Cutaneous PSL requiring confirmation of diagnosis by teledermatology. | 197/235 | VI (no algorithm) Dermoscopy (no algorithm)<br>In-person (Single) | Subjective impression ('suspicious for malignancy') | GP (n=13)<br>Assumed to be Low (Expertise NR; simple protocols for diagnosis provided for study purposes) | Histology/Clinical FU (6 months)<br>MM+MiS 5; BCC 0; Benign 230 (NR)<br>20% | NR |
| Menzies 2009<br>Pathway - clear<br>MM+MiS<br>Any | WPC<br>P-CS<br>Australia<br>Primary | PSL that would be biopsied or referred on after routine naked eye examination | NR/374 | VI (no algorithm) Dermoscopy (no algorithm)<br>In-person (Single) | Subjective impression ('correct diagnosis of melanoma') | GP (n=62)<br>Assumed to be Low (trained for study; required history of excision or referral of at least 10 pigmented skin lesions over the previous 12-month period but no prior dermoscopy use) | Histology/Clinical FU (3-6 months)/Expert dx<br>MM+MiS 32; BD 2; Benign 323; Unknown 9<br>4% | 6 BCC a BD exclu by autho 43 exclu as both \ Dermosc diagnose not availa |
| Walter 2012<br>Pathway - clear<br>MM<br>MM+MiS<br>Any | BPC<br>RCT<br>UK<br>Primary | Any suspicious PSL that could not immediately be diagnosed as benign | 654/792 (control arm only) | VI (7-point)<br>Siascope (iv arm)<br>In-person (Single) | 7PCL: >= 3 | GP (n=28)<br>Nurse practitioner (n=2)<br>Low (excluded if specialist dermatology training) | Histology/Clinical FU (3-6 months)/Expert dx<br>Control group only:<br>MM 16; MiS 2<br>BCC 4; SK 20; DF 2; lentigo 5; 'benign' 686; unknown 10<br>6% | 19 (5 due violation recruitme criteria o discontin protocol; died; 4 d not atten for dermatol assessm 2 missing histology not clear accounte for) |
| **3 – Limited prior testing (selected for excision)** | | | | | | | | |
| Collas 1999<br>Pathway – unclear<br>MM+MiS | NC<br>P-CS<br>France<br>Mixed (Private/Hospital) | PSL undergoing excision by dermatologists in private practice, and by hospital dermatologists | 353 / 353 | VI (1. no algorithm; 2. own new algorithm)<br>In-person | 1. subjective impression<br>2. >=1 of 3 characteristics present | Dermatologist (n=NR; exp NR)<br>Single observer | Histology<br>MM+MiS 38<br>BN 249; Other pigmented 55<br>38/353; 11% | None reported |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MEL) | Exclusio |
|---|---|---|---|---|---|---|---|---|
| Gachon 2005<br>Pathway – clear | NC<br>P-CS<br>France<br>Private | Melanocytic skin lesions removed for any reason | NR / 4036 | VI (no algorithm)<br>In-person; Single | Subjective impression ('considered suspicious') | Dermatologists (135/200)<br>Exp. NR | Histology<br>MM 113; MiS 36<br>BN 3887<br>149/4036; 4% | NR |
| McGovern 1992<br>Pathway – clear | WPC-algs<br>P-CS<br>US<br>Community (Army Medical Center DermClinic) | PSL (>10mm) excised to rule out dysplasia, lentigo maligna or malignant melanoma | 179 / 237 | VI (7-point; (A)BCD)<br>In-person; Single | 7-point: >=2, >=3, >=4 characteristics present (A)BCD: >=1, >=2, >=3 characteristics present | NR (presume dermatologist)<br>Exp. NR | Histology<br>MM 6; MiS 6<br>BCC 4; SK 32; BN 138; AK 6; Other 45<br>12/205; 6% | 32 lesion unaccou for; 13 excluded due to le size of 8 or less. 1 evaluate ABCD ar 3-point; 2 evaluate 7 point |
| **4 - Referred for further assessment** | | | | | | | | |
| Barzegari 2005<br>Pathway – clear<br>MM+MiS | WPC<br>NR-CS<br>Iran<br>Secondary | PSL <=15mm diameter referred to dermatology clinic for diagnostic evaluation or cosmetic reasons | 91 / 122 | VI (no algorithm)<br>In-person (consensus diagnosis of 2) | Melanoma likely / melanoma possible | Mixed (n=2; 1 attending dermatologist and a third year dermatology resident) | Histology<br>MM 3; MiS 3<br>SK 2; AK 1; BN 106; DF 7<br>6/122; 5% | None reported |
| Stanganelli 2000<br>Pathway – clear<br>MM+MiS<br>Any | WPC<br>R-CS<br>Italy<br>Specialist clinic | PSL referred by dermatologists and general practitioners either for pre-surgical assessment or consultation | NR / 3372 | VI (ABCD) Dermoscopy (no algorithm)<br>In-person (Single) | NR Subjective impression | NR (assumed dermatologist - described as one of the co-authors; n=1) | Histology / Registry FU<br>MM+MiS 55<br>BCC 43; Benign 3274<br>55/3372; 2% | None reported |
| **5 - Referred for further assessment (selected for excision)** | | | | | | | | |
| Benelli 1999<br>Pathway – unclear<br>MM+MiS | WPC<br>P-CS<br>Italy<br>Secondary | All PSL observed and excised at the Dermatologic Surgery Department | NR / 401 | 1. VI (ABCDE)<br>2. Dermoscopy (7FFM)<br>In-person | 1. >/=1 characteristic present; >/=2 characteristics present; >/=3 characteristics present; >/=4 characteristics present; all 5 characteristics present<br>2. Score >=2 | Dermatologist (n=2; exp NR)<br>Consensus of 2 | Histology<br>MM 54; MiS 6<br>BCC 1<br>BN 337; LS 5; SK 1<br>60/401; 15% | None reported |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MEL) | Exclusio |
|---|---|---|---|---|---|---|---|---|
| Bono 2002<br>Pathway – clear<br>MM+MiS | WPC<br>P-CS<br>Italy<br>Specialist clinic | PSL with a more or less important suspicion for MM on VI and/or dermoscopy | 298 / 313 | VI (no algorithm)<br>Dermoscopy (no algorithm)<br>In-person | VI - subjective impression<br>Dermoscopy - >=1 characteristic present | Surgical oncologist (n=4; high)<br>Single observer | Histology<br>MM 55; MiS 11<br>BCC 6; 8 SK; 3 SN; BN 230<br>66/313; 21% | None reported |
| Bono 2002b<br>Pathway – clear<br>MM+MiS | WPC<br>P-CS<br>Italy<br>Specialist clinic | PSL <=6mm requiring surgical biopsy for diagnosis based on clinical or dermoscopic suspicion of MM | 157 / 161 | VI (no algorithm)<br>Dermoscopy (no algorithm)<br>In-person | VI - subjective impression<br>Dermoscopy - >=1 characteristic present | Surgical oncologist (n=2; high)<br>Single observer | Histology<br>MM 10; MiS 3<br>BCC 2; SK 4; SN 5; BN 124<br>13/161; 8% | None reported |
| Bono 2006<br>Pathway – clear<br>MM+MiS | WPC<br>R-CS<br>Italy<br>Specialist clinic | PSL <=3mm undergoing excision due to a more or less important suspicion for MM on VI and/or dermoscopy | 204 / 206 | VI (no algorithm)<br>Dermoscopy (Menzies)<br>In-person | VI - subjective impression<br>Dermoscopy - NR | NR; assumed surgical oncologist as per Bono 2002; Bono 2002b (n=4; exp NR)<br>Single observer | Histology<br>MM 19; MiS 4<br>SN 3; BN 169; Other 11<br>23/206; 11% | None reported |
| Carli 2002<br>Pathway – unclear<br>MM+MiS | WPC<br>R-CS<br>Italy<br>Secondary | Clinically equivocal and suspicious PSL subjected to excisional biopsy at the Institute of Dermatology | NR / 256 | 1. VI (no algorithm)<br>2. Dermoscopy (pattern)<br>In-person (Dermoscopy – image-based) | Subjective impression | Dermatologist (n=2; High exp – “extensive experience in both clinical and dermoscopic diagnosis”)<br>Consensus of 2 | Histology<br>MM 40; MiS 14<br>BCC 5<br>BN 177; SN 16; SK 4<br>54/256; 21% | None reported |
| Cristofolini 1994<br>Pathway – unclear<br>MM+MiS | WPC<br>P-CS<br>Italy<br>Secondary | Patients with PSL presenting during a campaign for the early diagnosis of cutaneous melanoma at the Dermatology Department | NR / 220 | 1. VI (ABCDE)<br>2. Dermoscopy (pattern)<br>In-person | 1. ≥2 characteristics present<br>2. >=1 characteristic present | Dermatologist (n=4; High exp - dermatologists had all been trained in the recognition of pigmented lesions)<br>Unclear observer interpretation | Histology<br>MM+MiS 33<br>BCC 0<br>BN 181; SK 4; 2 thrombosed angioma<br>33/220; 15% | None reported |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MEL) | Exclusio |
|---|---|---|---|---|---|---|---|---|
| Cristofolini 1997<br>Pathway – unclear<br>MM+MiS | WPC (algs)<br>NR-CS<br>Italy<br>Secondary | Patients with small and flat common and atypical PSL recruited during a health campaign for the early diagnosis of melanoma; all underwent skin biopsy. | 176 / 176 | VI (ABCD)<br>In-person | NR | Dermatologist (n=3; High experience)<br>Consensus of 3 | Histology<br>MM+MiS 35<br>BN 141<br>35/176; 20% | None reported |
| Ek 2005<br>Pathway – clear<br>MM+MiS<br>Any | NC<br>P-CS<br>Australia<br>Specialist clinic | Lesions excised for which malignancy could not be excluded | 1223 / 2582 | VI (no algorithm)<br>In-person | Subjective impression | Plastic surgeon (n=4 or 5; mixed experience; 3 consultants, 1 plastic surgery trainee (usually 1st year, on 6 month rotation) and a clinical assistant)<br>Unclear | Histology<br>MM+MiS 23<br>BCC 1214; SCC 517; BD 188; SK 63; 577 other benign (including 330 solar keratosis)<br>23/2582; 1% | Incomple or incorre entered proforma were excluded 79 patier with 96 lesions |
| Green 1991<br>Pathway – clear<br>MM+MiS | NC<br>NR-CS<br>Australia<br>Secondary | PSL for excision | 81 / 89 | VI (no algorithm)<br>In-person | Subjective impression | NR (n=NR; exp NR "in the majority of cases a surgeon or a dermatologist")<br>Single observer | Histology<br>MM+MiS 5<br>BCC 2; SK 7; BN 54; Other 2<br>5/70; 7% | 19/89 lesions excluded (number participa not repor due to incomple clinical a histology records. |
| Langley 2001<br>Pathway – unclear<br>MM+MiS | NC<br>P-CS<br>US<br>Specialist clinic | Patients with lesions scheduled for excision at the pigmented lesion clinic to either remove atypical nevi or to rule out melanoma or for cosmetic reasons | NR / 38 | VI (no algorithm)<br>In-person | NR | NR (presume dermatologist; n=NR; exp NR)<br>Unclear | Histology<br>MM 3; MiS 3<br>BN 32<br>6/38; 16% | None reported |
| Morales Callaghan 2008<br>Pathway – unclear<br>MM+MiS | WPC<br>P-CS<br>Spain<br>Secondary | Randomly selected melanocytic lesions; melanocytic on both clinical and dermoscopic criteria | 166 / 200 | 1. VI (no algorithm)<br>2. Dermoscopy (no algorithm)<br>In-person | NR | Dermatologist (n=2; high exp – “experience in dermoscopy”)<br>Consensus of 2 | Histology<br>MM+MiS 6<br>BN 184; SN 1; Other 9<br>6/200; 3% | None reported |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MEL) | Exclusio |
|---|---|---|---|---|---|---|---|---|
| Morton 1998a (high exp), Morton 1998b (mod exp) and Morton 1998c (low exp)<br>Pathway – clear<br>MM+MiS | NC<br>R-CS<br>UK<br>Specialist clinic | Patients referred by their GP to the clinic | NR / 1999 | VI (no algorithm)<br>In-person | NR | Dermatologist (n=2; high); Dermatology senior registrar (n=1; moderate); Dermatology registrar (n=1; low)<br>SIngle observer per lesion | Histology<br>MM 104; MiS 24<br>Benign 1871<br>High exp: 69/763; 9%<br>Moderate exp: 31/567; 5%<br>Low exp: 28/669; 4% | None reported |
| Thomas 1998<br>Pathway – unclear<br>MM+MiS | NC<br>CCS<br>France<br>Secondary | All cases of melanoma and a nonselected consecutive group of 'non-melanoma' PSL | NR / 1140 | VI (ABCDE)<br>In-person | >=1 characteristic present<br>>=2 characteristics present<br>>=3 characteristics present<br>>=4 characteristics present<br>all 5 characteristics present | Dermatologist (n=2; High exp - described as 'trained dermatologists')<br>Single observer | Histology<br>MM+MiS 460<br>BCC 8<br>BN 638; SN 2; Other 13<br>460/1140; 40% | None reported |
| Unlu 2014<br>Pathway – unclear<br>MM+MiS | WPC (algs)<br>R-CS<br>Turkey<br>Specialist clinic | Melanocytic lesions excised at Department of Dermatology Pigmented Lesion Clinic | 115 / 115 | 1. VI (no algorithm)<br>2. Dermoscopy (7-point; 3-point; CASH; ABCD)<br>In-person | 1. subjective impression<br>2. score >=3; >=2 characteristics present; score >=8; score >5.44 | NR (presume dermatologist; n=1 for VI; n=3 for dermoscopy; Exp NR for VI)<br>Single observer (VI); consensus of 3 (dermoscopy) | Histology<br>MM+MiS 24<br>BN 91<br>24/115; 21%` | None reported |
| Zaumseil 1983<br>Pathway – unclear<br>MM+MiS | NC<br>NR-CS<br>Germany<br>Secondary | Skin lesions undergoing excision | NR / 7063 | VI (no algorithm)<br>In-person | Subjective impression | NR (n=NR; exp NR)<br>Single observer | Histology<br>MM+MiS 337<br>Not melanoma 6726 (dx listed only for FPs)<br>337/7063; 5% | None reported |

**5* - Equivocal referred for further assessment (selected for excision)**

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MEL) | Exclusio |
|---|---|---|---|---|---|---|---|---|
| Dummer 1993<br>Pathway – clear<br>MM+MiS | WPC<br>P-CS<br>Germany | Patients with melanocytic skin lesions difficult to diagnose clinically | NR / 771 | VI (no algorithm)<br>Dermoscopy (pattern)<br>In-person (I-B for dermoscopy) | NR | NR assume Dermatologist (assumed) (n=2; exp NR)<br>Single observer | Histology<br>MM 19; MiS 4<br>SK 4; BN 706; Benign NML 32; Other 6<br>23/771; 3% | 53 non-melanoc lesions n included the final analysis melanom present i this grou |
| Soyer 1995<br>Pathway – clear<br>MM+MiS | WPC<br>NR-CS<br>Austria | PSL difficult to diagnose on clinical grounds alone | NR / 159 | VI (no algorithm)<br>Dermoscopy (pattern)<br>In-person | NR | Dermatologist (n=2; exp High; "the examination was performed by a dermatologist expert in dermoscopy")<br>Single observer | Histology<br>MM 50; MiS 15<br>BCC 3; SK 18; AK 4; BN 61; Other 7<br>65/159; 41% | None reported |
| Steiner 1987<br>Pathway – unclear<br>MM+MiS | P-CS<br>Austria<br>Specialist clinic | Small (< 10 mm) diagnostically equivocal PSL; no absolute agreement on clinical diagnosis among investigating clinicians at a pigmented lesion clinic. | NR / 318 | 1. VI (no algorithm)<br>2. Dermoscopy (pattern)<br>In-person | Subjective impression | Dermatologists (n=3; High exp - "experienced dermatologists")<br>Consensus diagnosis of 3 observers | Histology<br>MM 49; MiS 24<br>BCC 20<br>BN 143; SK 20; lentigo simplex and nevoid lentigo 19; Other 15<br>73/318; 23% | None reported |

NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; cSCC – cutaneous squamous cell carcinoma; LS – lentigo simplex; SK – seborrheic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BNM – benign non-melanocytic; BD – Bowen' disease; DF – dermatofibroma; FU – follow-up; R –retrospective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; exp - experience; GP - general practitioner; VI - visual inspection; RCT - randomised controlled trial; SCC - squamous cell carcinom dx - diagnosis; AHM - atypical melanocytic naevi; ELM - epiluminescence microscopy.

## 10 Summary QUADAS - in-person evaluations

| | STUDIES CLEARLY PLACED ON CLINICAL PATHWAY | | STUDIES NOT CLEARLY PLACED ON CLINICAL PATHWAY | |
|---|---|---|---|---|
| Pathway | Risk of Bias | Concerns about applicability | Risk of Bias | Concerns about applicability |
| 2 – Limited prior testing | | | | |
| Studies | N=3; Grimaldi 2009 ; Menzies 2009 ; Walter 2012 | | N=0 | |

| | | | | |
|---|---|---|---|---|
| Participant selection | Low (3/3) | High (2/3); Unclear (1/3). Inclusion of multiple lesions per participant (Grimaldi 2009; Walter 2012 ); patient numbers not reported (Menzies 2009). | | |
| Index test | Low (1/3); Unclear (2/3). Lack of clear pre-specification of threshold (Grimaldi 2009; Menzies 2009) | Low (1/3); High (2/3). Lack of description of diagnostic threshold (Grimaldi 2009; Menzies 2009). Non-expert test interpretation (Menzies 2009; Walter 2012); not clear in Grimaldi 2009. | | |
| Reference standard | High (3/3). <80% of disease negative participants had histological or clinical follow up reference standard | High (2/3); Unclear (1/3). Expert diagnosis as reference standard (Menzies 2009; Walter 2012); unclear histopathologist expertise (3/3). | | |
| Flow and timing | High (3/3). Mixed reference standards (3/3); participant exclusions (Menzies 2009; Walter 2012); all unclear on index to reference interval. | | | |
| **3 – Limited prior testing (selected for excision)** | | | | |
| **Studies** | **N=2;** Gachon 2005 **;** McGovern 1992 | | **N=1;** Collas 1999 | |
| Participant selection | Low (1/2); Unclear (1/2). Unclear exclusion criteria (1/2; Gachon 2005). | High (2/2). Restriction to melanocytic (1/2; Gachon 2005) or primarily excised lesions (2/2); multiple lesions per participant (1/2; McGovern 1992); no. patients not reported (1/2; Gachon 2005) | Unclear (1/1). Participant sampling not described; exclusion criteria not reported | High (1/1). Excised only included |
| Index test | Unclear (1/2); High (1/2). Lack of clear pre-specification of the threshold (1/2; Gachon 2005) or testing of multiple thresholds (1/2; McGovern 1992) | High (1/2); Unclear (1/2). Lack of threshold detail (1/2; Gachon 2005); unclear description of observer expertise (2/2) | Low (1/1) | Unclear (1/1). Observer expertise not described |

| | | | | |
|---|---|---|---|---|
| Reference standard | Low (2/2) | Low (1/2); Unclear (1/2). Lack of description of histopathology expertise (1/2; Gachon 2005) | Low (1/1). | Unclear (1/1). Histology expertise not described (histologically analysed by different private and hospital pathologists and reviewed by one of the authors). |
| Flow and timing | High (1/2); Unclear (1/2). Participant exclusions (1/2; McGovern 1992); unclear reference interval (2/2). | | Low (1/1) | |
| **4 - Referred for further assessment** | | | | |
| **Studies** | **N=2;** Barzegari 2005 ; Stanganelli 2000 | | **N=0** | |
| Participant selection | Low (2/2) | High (2/2). Included excisions for cosmetic reasons (1/2; Barzegari 2005), or multiple lesions per participant (2/2). | | |
| Index test | Low (1/2); Unclear (1/2). Lack of clear pre-specification of the threshold (Barzegari 2005) | High (1/2); Unclear (1/2). Consensus result (1/2; Barzegari 2005); insufficient threshold detail (1/2; Barzegari 2005); observer expertise not clear (2/2). | | |
| Reference standard | Low (1/2); High (1/2). <80% of disease negative participants had histological or clinical follow up reference standard (Stanganelli 2000) | Unclear (2/2). Lack of description of histopathology expertise (2/2) | | |
| Flow and timing | High (1/2); Unclear (1/2). Unclear reference interval (2/2); use of different reference standards (1/2; Stanganelli 2000) | | | |
| **5 - Referred for further assessment (selected for excision)** | | | | |
| **Studies** | **N=6;** Bono 2002 ; Bono 2002b ; Bono 2006 ; Ek 2005 ; Green 1991 ; Morton 1998a ; Morton 1998b ; Morton 1998c ***)** | | **N=9;** Benelli 1999 ; Carli 2002b ; Cristofolini 1994 ; Cristofolini 1997 ; Langley 2001 ; Morales Callaghan 2008 ; Thomas 1998 ; Unlu 2014 ; Zaumseil 1983 | |

| | | | | |
|---|---|---|---|---|
| Participant selection | Low (2/6); High (2/6); Unclear (2/6). Inappropriate (2/6; Bono 2002; Ek 2005) or unclear (2/6; Gachon, Morton) exclusions; consecutive recruitment not reported (1/6; Gachon) | High (6/6). Unrepresentative (6/6) participants; all excised. Multiple lesions per participant (2/6; Ek 2005; Green 1991) or number of participants not reported (Morton 1998) | High (4/9); Unclear (5/9). Inappropriate exclusions (4/9) due to restriction to melanocytic only (Morales Callaghan 2008; Unlu 2014 ), disagreement on histology (Zaumseil 1983). Use of case control type design (1/9; Thomas 1998). Unclear participant sampling (6/9; Benelli 1999; Carli 2002b, Cristofolini 1994, Cristofolini 1997, Langley 2001; Zaumseil 1983). | High (9/9). Inclusion of only excised lesions (9/9). Multiple lesions per participant (2/9; Langley 2001, Morales Callaghan 2008); number of participants not reported (6/9; Benelli 1999, Carli 2002b, Cristofolini 1994, Cristofolini 1997, Thomas 1998, Zaumseil 1983) |
| Index test | Low (3/6); Unclear (3/6). Pre-specification of threshold not reported (Ek 2005; Green 1991; Morton 1998) | High (6/6). All clinically applicable application of test. No threshold details (6/6). Observer experience unclear (3/6; Bono 2006; Ek 2005; Green 1991). | Low (2/9); High (2/9), Unclear (5/9). Threshold not prespecified (2/9; Benelli 1999, Thomas 1998) or not clear whether prespecified (Carli 2002b, Cristofolini 1997, Langley 2001, Morales Callaghan 2008, Unlu 2014). | Low (1/9);High (7/9), Unclear (1/9). Test application not clinically applicable (4/9; Benelli 1999; Carli 2002b; Cristofolini 1997; Morales Callaghan 2008) or not clear (Cristofolini 1994; Langley 2001). No threshold detail (5/9; Carli 2002b; Langley 2001; Morales Callaghan 2008; Unlu 2014; Zaumseil 1983) |
| Reference standard | Low (5/6); High (1/6). Inadequate reference standard (1/6; Green 1991) | Low (1/6); High (1/6); Unclear (4/6). Expert diagnosis used (1/6; Green 1991). Lack of description of histopathology expertise (5/6; all except Morton 1998) | Low (9/9) | Low (2/9); High (1/9); Unclear (6/9). Use of expert diagnosis (1/9; Langley 2001). Histopathology expertise not reported (7/9; Benelli 1999; Carli 2002b; Cristofolini 1994; Cristofolini 1997; Langley 2001; Morales Callaghan 2008; Zaumseil 1983) |
| Flow and timing | High (2/6); Unclear (4/6). Index to reference interval not reported (5/6; Bono 2002; Bonon 2002b; Bono 2006; Green 1991; Morton 1998). Participant exclusions due to incomplete data (2/6; Green 1991; Ek 2005) | | Low (3/9);Unclear (6/9). Interval to reference standard not reported (6/9; Benelli 1999; Cristofolini 1994; Langley 2001; Thomas 1998; Unlu 2014; Zaumseil 1983) | |
| **5* - Equivocal referred for further assessment (selected for excision)** | | | | |
| **Studies** | **N=2;** Dummer 1993 **;** Soyer 1995 | | **N=1;** Steiner 1987 | |
| Participant selection | Unclear (2/2). Unclear sampling methods (2/2); Unclear exclusions (1/2; Soyer 1995) | High (1/2); Unclear (1/2). Participants not representative (1/2; Dummer 1993) or unclear (1/2; Soyer 1995). Number of participants not reported (2/2) | Unclear (1/1). Participant sampling not described; exclusion criteria not reported | High (1/1). Restricted to small <10mm pigmented skin lesions; all excised |
| Index test | Unclear (2/2). Pre-specification of threshold not reported (2/2) | High (2/2). No threshold details (2/2). Observer experience unclear (1/2; Dummer 1993) | Unclear (1/1). Pre-specification of threshold not reported. | High (1/1). Consensus decision reported and no threshold detail. |

| | | | | |
|---|---|---|---|---|
| Reference standard | Low (2/2) | Unclear (2/2). Lack of description of histopathology expertise (2/2) | Low (1/1). | Unclear (1/1). Histology expertise not described |
| Flow and timing | High (1/2), Unclear (1/2). Participant exclusions (1/2; Dummer 1993). Index to reference interval not reported (2/2) | | Low (1/1). | |

## 11 Summary study details – image-based evaluations

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| 3 – Limited prior testing (with selection on reference standard) | | | | | | | | |
| Bourne 2012<br>Pathway - clear | WPC-tests<br>R-CS<br>Australia<br>Primary | All skin lesions excised to exclude skin cancer (and 3 examples common lesions assessed as clearly benign and not biopsied) | 46 / 50 | VI (no algorithm) Dermoscopy (3-point; Menzies; BLINCK (excluded))<br>Image-based (blinded) | NR | GP (n=3)<br>Clinical nurse (n=1)<br>Mixed experience “varying levels of dermatoscopic experience”<br>Average | Histology / Clinical FU / Expert dx<br>MM 1; MiS 8<br>BCC 6; SK 5; BN 11; Other 19<br>9/45; 20% | 5 non-pigmented specimens (not further identified) in the set of 50 were excluded from dermoscopic evaluations |
| Rosendahl 2011<br>Pathway – unclear | NC<br>R-CS<br>Australia<br>Primary | PSL submitted for histology from the primary care skin cancer practice of one author | 389 / 463 | 1. VI (no algorithm)<br>2. Dermoscopy (pattern) | 1. subjective impression<br>2. both characteristics present | Dermatologist (n=1)<br>Image-based; High experience (confirmed by author); Single observer | Histology<br>MM 9; MiS 20<br>BCC 72; SCC 5<br>BN 217; BD 18; AK 14*; BNM 140<br>*<br>considered malignant by study authors<br>29/463; 6% | 3 poor quality images excluded |
| 4 - Referred for further assessment | | | | | | | | |
| Stanganelli 2005<br>Pathway - clear<br>MM+MiS | WPC<br>R-CS<br>Italy<br>Specialist clinic | Melanocytic lesions referred to Skin Cancer Unit for clinical and dermoscopic evaluation. | NR / 477 | VI (no algorithm) Dermsocopy (no algorithm)<br>Image-based (Average) | NR | Dermatologist (n=3); GP (n=3)<br>Dermatologists - High experience (“2 years dermoscopy experience”); experience NR for GPs, assumed Low | Histology / Registry FU<br>MM+MiS 31<br>BN 103<br>31/134; 23% | None reported |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| 5 - Referred for further assessment (with selection on reference standard) | | | | | | | | |
| Benelli 2001<br>Pathway – unclear | WPC<br>R-CS<br>Italy<br>Training images | Slides of PSL selected for evaluation during a training course on dermoscopy. Lesions not located on head, palms or soles | NR / 49 | 1. VI (ABCDE)<br>2. Dermoscopy (7FFM) | 1. >=3 & >=2<br>2. >=2 | Expert author (n=1); Dermatologists (n=65)<br>Image-based; Single author - High experience; Average result for dermatologist group; experience NR | Histology<br>MM 10, MiS 2<br>BCC 2<br>BN 25, SN 5, SK 3,<br>Other 2 (1 missing)<br>12/50; 24% | None reported |
| Carli 2002b<br>Pathway – unclear | WPC<br>R-CS<br>Italy<br>Secondary | Clinically suspicious or equivocal PSL undergoing excision for diagnostic purposes; all <= 14mm diameter | NR / 57 | 1. VI (NR)<br>2. Dermoscopy (NR) | NR | Dermatologists (n=2)<br>Image-based; high experience ('with experience in the field of PSL'); consensus of 2 | Histology<br>MM 6, MiS 5<br>BCC 10<br>BN 31, SK 1; Other 4<br>11/57; 19% | 4 'not evaluables' excluded (1MM, 3 benign) |
| Dolianitis 2005<br>Pathway – unclear | WPC<br>CCS<br>Multi-centre<br>Training images | Melanocytic skin lesions selected from a collection of dermoscopic images belonging to one author. | NR / 40 | 1. VI (no algorithm)<br>2. Dermoscopy (Pattern analysis; Menzies Criteria; 7 point; ABCD) | 1. subjective impression<br>2. subjective impression; NR; NR; >4.75 | Dermatologists(n=16); dermatology trainees (n=16); GPs (n=35)<br>Image-based; mixed experience ("range of experience levels with assessment of skin lesions"); average result | Histology (n=39); Expert diagnosis (n=1)<br>MM 18, MiS 2<br>BN 12; SN 3; Other 4<br>20/20; 50% | None reported; poor quality images exclusion criterion |
| Pizzichetta 2004<br>Pathway – unclear | WPC<br>R-CS<br>US/Italy<br>Secondary | Clinical and/or dermoscopic hypomelanotic (extent of pigmentation <=30%) and amelanotic skin lesions | 151 / 151 | 1. VI (no algorithm)<br>2. Dermoscopy (pattern) | subjective impression | NR (presume dermatologist; n=1)<br>Image-based; experience NR; single observer | Histology<br>AHM 34, MiS 5<br>BCC 25, SCC 5<br>BN 47, SN 5, SK 8, Other 18<br>39/108; 36% (analysed) | 23 lesions excluded due to image quality; further 43 lesions were not available for evaluation by clinical images ("mainly benign melanocytic lesions"). |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| Stanganelli 1998<br>Pathway – unclear | WPC<br>R-CS<br>Italy<br>Training images | PSL images selected from computerised files of the skin cancer clinic. | NR / 30 | 1. VI (no algorithm)<br>2. Dermoscopy (no algorithm) | NR | Dermatologists (n=20)<br>Image-based; experience NR ("experience in ELM but (with) no formal training"); average | Histology<br>MM+MiS 10<br>BCC 4<br>BN 10, SK 3, Other 3<br>10/30; 33% | None reported |
| Winkelmann 2016<br>Pathway – unclear | WPC<br>CCS<br>Unclear<br>Training images | Selected images previously analysed by MSDSLA | NR / 12 | 1. VI (no algorithm)<br>2. Dermoscopy (no algorithm) | NR | Dermatologists (n=70)<br>Image-based; experience NR; average | Histology<br>MM 3; MiS 2<br>BN 7<br>5/12; 42% | None reported |
| **5* - Equivocal referred for further assessment (with selection on reference standard)** | | | | | | | | |
| Carli 2003b<br>Pathway – unclear | WPC<br>R-CS<br>Italy<br>Secondary | Clinically difficult to diagnose or equivocal melanocytic lesions randomly selected from image database; all melanomas <1mm thickness. | NR / 200 | 1. VI (no algorithm)<br>2. Dermoscopy (own choice) | subjective impression | Dermatology registrar (n=2); dermatologists (senior experts n=2; practicing dermatologists n=4)<br>Classed as High experience (both dermatologists and registrars "formally trained in dermoscopy"); Average result | Histology<br>MM 40; MiS 24<br>BN 136<br>64/200; 32% | None reported |
| de Giorgi 2012<br>Pathway – unclear | WPC<br>R-CS<br>Italy<br>Secondary | Pigmented melanocytic skin lesions <= 6mm diameter excised at dermatology department | NR / 103 | VI (ABCD) | 1. >=2 characteristics present<br>2. >=3 characteristics present | Dermatologists (n=3)<br>High experience ("more than 5 years of practice in dermoscopy"); consensus of 3 | Histology<br>MM 16; MiS 18<br>BN 69<br>34/103; 33% | None reported |

NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; cSCC – cutaneous squamous cell carcinoma; LS – lentigo simplex; SK – seborrhoeic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BD – Bowen's disease; DF – dermatofibroma; FU – follow-up; R –retrospective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; GP - general practitioner; VI - visual inspection; RCT - randomised controlled trial; SCC - squamous cell carcinoma; dx - diagnosis; ELM – epiluminescence microscopy; BLINCK - Benign Lonely irregular Nervous Change Known Clues; BNM - benign non-melanocytic; AHM - amelanotic/hypomelanotic melanoma; MSDSLA - multispectral digital skin lesion analysis device.

## 12 Summary QUADAS - image-based evaluations

| | STUDIES CLEARLY PLACED ON CLINICAL PATHWAY | STUDIES NOT CLEARLY PLACED ON CLINICAL PATHWAY |
|---|---|---|

| Pathway | Risk of Bias | Concerns about applicability | Risk of Bias | Concerns about applicability |
|---|---|---|---|---|
| **3 – Limited prior testing (with selection on reference standard)** | | | | |
| **Studies** | **N=1;** Bourne 2012 | | **N=1;** Rosendahl 2011 | |
| Participant selection | Unclear (1/1). Unclear exclusion criteria ( Bourne 2012). | High (1/1). Restriction to primarily excised lesions (1/1) | Low (1/1). | High (1/1). Includes excised lesions only; multiple lesions per participant. |
| Index test | Unclear (1/1) Lack of clear pre-specification of the threshold (Bourne 2012). | High (1/1). Blinded image interpretation and average observer result presented (Bourne 2012); lack of threshold detail (Bourne 2012); unclear description of observer expertise | Unclear (1/1). No clear pre-specification of threshold. | High (1/1). Image-based study; no threshold detail. |
| Reference standard | Low (1/1) | High (1/1) Use of expert diagnosis as reference (Bourne 2012); lack of description of histopathology expertise (Bourne 2012) | Low (1/1). | Unclear (1/1). Histopathology experience not reported. |
| Flow and timing | High (1/1) Use of different reference standards (Bourne 2012 ); participant exclusions (Bourne 2012). | | High (1/1). Exclusions on image quality Unclear interval between index and reference. | |
| **4 - Referred for further assessment** | | | | |
| **Studies** | **N=1;** Stanganelli 2005 | | **N=0** | |
| Participant selection | Unclear (1/1). Unclear participant sampling across all items (Stanganelli 2005) | High (1/1). Sample restricted to melanocytic lesions (Stanganelli 2005). Patient numbers not reported, | | |

| | | | | |
|---|---|---|---|---|
| Index test | Unclear (1/1). Lack of clear pre-specification of the threshold (Stanganelli 2005) | High (1/1). Average result presented (Stanganelli 2005); insufficient threshold detail (Stanganelli 2005). | | |
| Reference standard | Low (1/1) | Unclear (1/1). Unclear use of expert diagnosis as reference standard (Stanganelli 2005). Unclear histopathology expertise | | |
| Flow and timing | High (1/1) Use of different reference standards (Stanganelli 2005); unclear reference interval. | | | |
| **5 - Referred for further assessment (with selection on reference standard)** | | | | |
| **Studies** | **N=0** | | **N=6;** Benelli 2001 ; Carli 2002b ; Dolianitis 2005 ; Pizzichetta 2004 ; Stanganelli 1998 ; Winkelmann 2016 | |
| Participant selection | | | High (3/6), Unclear (3/6). Case control type design used (3/3; Dolianitis 2005; Stanganelli 1998, Winkelmann 2016) or unclear design (Benelli 2001; Pizzichetta 2004). Unclear participant sampling (5/6; Benelli 2001, Carli 2002b, Pizzichetta 2004, Stanganelli 1998, Winkelmann 2016), design unclear (1/6), exclusion criteria not clearly reported (5/6; Benelli 2001, Carli 2002b, Dolianitis 2005, Stanganelli 1998, Winkelmann 2016). | High (6/6). Excised only included (6/6), amelanotic/ hypomelanotic lesions only (1/6; Pizzichetta 2004). Number participants not reported (5/6; Benelli 2001, Carli 2002b, Dolianitis 2005, Stanganelli 1998, Winkelmann 2016) |

| | | | | |
|---|---|---|---|---|
| Index test | | | Low (1/6); Unclear (5/6). No clear pre-specification of threshold (5/6; Carli 2002b, Dolianitis 2005, Pizzichetta 2004, Stanganelli 1998, Winkelmann 2016). | High (6/6). Image-based evaluations (6/6), blinded to all other information (5/6; Benelli 2001, Carli 2002b, Dolianitis 2005, Stanganelli 1998, Winkelmann 2016), with consensus (1/6; Carli 2002b) or average result (4/6; Benelli 2001, Dolianitis 2005, Stanganelli 1998, Winkelmann 2016) reported. Threshold not clearly specified (5/6; Carli 2002b, Dolianitis 2005, Pizzichetta 2004, Stanganelli 1998, Winkelmann 2016). Observer expertise not reported (4/6; Dolianitis 2005; Pizzichetta 2004; Stanganelli 1998; Winkelmann 2016) |
| Reference standard | | | Low (6/6) | High (1/6); Unclear (5/6). Use of expert observer diagnosis (1/6; Dolianitis 2005); expertise of histopathologist not described (6/6). |
| Flow and timing | | | Low (1/6); High (2/6); Unclear (3/6). Lesions excluded from analysis (reason not reported) (2/6; Dolianitis 2005; Pizzichetta 2004); different reference standards used (1/6; Dolianitis 2005). Index to reference interval not reported (5/6; Benelli 2001, Dolianitis 2005, Pizzichetta 2004, Stanganelli 1998, Winkelmann 2016). | |
| **5* - Equivocal referred for further assessment (with selection on reference standard)** | | | | |
| **Studies** | **N=0** | | **N=2; Carli 2003b ; de Giorgi 2012** | |
| Participant selection | | | High (2/2). Exclusion of difficult to diagnose, including peculiar lesions (1/2; Carli 2003b), histology disagreement (1/2; de Giorgi 2012). | High (2/2). Restriction to melanocytic only (2/2), excised only (2/2). Patient numbers not reported (2/2) |
| Index test | | | High (1/2), Unclear (1/2). Multiple thresholds tested (1/2; de Giorgi 2012); no clear threshold specification (1/2; Carli 2003b). | High (2/2). Image-based evaluations (2/2), blinded to all other information (1/2; Carli 2003b), with consensus (1/2; de Giorgi 2012) or average result (1/2; Carli 2003b) reported. Threshold not described (1/2; Carli 2003b) |
| Reference standard | | | Low (2/2) | Low (2/2) |
| Flow and timing | | | Unclear (2/2). Index to reference interval not reported (2/2). | |

## 13 Summary study details – detection of invasive melanoma alone

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| **In-person** | | | | | | | | |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| Bono 1996 | WPC-tests<br>Unclear<br>Italy<br>Specialist clinic | Pigmented skin lesions at the Instituto Nazionale Tumori of Milan | 45 / 54 | VI (no algorithm)<br>Single observer | Subjective impression | Plastic surgeon | Histology plus other (31% of benign had expert dx)<br>MM: 18<br>BN: 25<br>18/43; 42% | Only 43 lesions had complete clinical and histological information. 11 lesions not surgically removed had only clinical diagnosis (benign) and were not included in the final accuracy analysis |
| Green 1994 | NC<br>NR-CS<br>Australia<br>Secondary | Pigmented lesions for excision | 129 / 164 | VI (no algorithm)<br>Single observer | Subjective impression; clinical dx recorded | NR | Histology<br>MM 18; MiS 3<br>BN 128; misc pigmented lesions including SK, BCC, lentigines 15<br>18/164; 11% | |
| Kopf 1975 | NC<br>R-CS<br>US<br>Specialist clinic | All lesions subject to biopsy at the Oncology Section of the Skin and Cancer Unit. | NR / 5538 | VI (no algorithm)<br>Single observer | No details; 'clinical diagnosis' | Oncologist | Histology<br>MM 99<br>Other dx listed only for false positives<br>99/5538; 2% | |
| Krahn 1998 | WPC-tests<br>P-CS<br>Germany<br>Secondary | Excised pigmented skin lesions | 80 / 80 | VI (no algorithm)<br>Single observer | no details | Dermatologist (assumed) | Histology<br>MM 39<br>BN 40; SN 1<br>39/80; 49% | |
| McGovern 1992<br>Pathway – clear<br>Pathway - clear | WPC-algs<br>P-CS<br>US<br>Community | PSL (>10mm) excised to rule out dysplasia, lentigo maligna or malignant melanoma | 179 / 237 | VI (7-point; (A)BCD)<br>In-person; Single | 7-point: >=2, >=3, >=4 characteristics present (A)BCD: >=1, >=2, >=3 characteristics present | NR (presume dermatologist)<br>experience.<br>NR | Histology<br>MM 6; MiS 6<br>BCC 4; SK 32; BN 138; AK 6; Other 45<br>6/211; 3% | 32 lesions unaccounted for; 13 excluded due to lesion size of 8mm or less. 192 evaluated for ABCD and 3-point; 205 evaluated for 7 point |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| Viglizzo 2004 | WPC-tests<br>NR-CS<br>Italy<br>Specialist clinic | Pigmented skin lesions examined at the Dermoscopy Service and undergoing excisions; high and medium risk on dermoscopy were selected for excision and 2x2 can be estimated only for melanocytic subgroup. | NR / 79 | VI (no algorithm)<br>Single observer | No details | Dermatologist (assumed) | Histology<br>Melanoma (invasive): 11; Melanoma (in situ): 1 Melanocytic lesion: 57<br>11/67 16% | |
| Walter 2012<br>Pathway - clear<br>MM<br>MM+MiS<br>Any | BPC<br>RCT<br>UK<br>Primary | Any suspicious PSL that could not immediately be diagnosed as benign | 654/792 (control arm only) | VI (7-point)<br>Siascope (iv arm)<br>In-person (Single) | NR | GP (n=28)<br>Nurse practictioner. (n=2)<br>Low (excluded if specialist dermatology training) | Histology/Clinical FU(3-6 months)/Expert dx<br>Control group only:<br>MM 16; MiS 2<br>BCC 4; SK 20; DF 2; lentigo 5; 'benign' 686; unknown 10<br>16/773 2% | 19 (5 due to violation of recruitment criteria or discontinued protocol; 1 died; 4 did not attend for dermatology assessment; 2 missing histology; 7 not clearly accounted for) |
| **Image-based** | | | | | | | | |
| Lorentzen 1999 | WPC-tests<br>P-CS<br>Denmark<br>Secondary | Patients with lesions suspicious for CMM referred to outpatients clinic | 232 / 232 | VI (no algorithm)<br>[Dermoscopy]<br>Single observer | Subjective impression; clinical diagnosis | Dermatologist | Histology<br>MM 49 'malignant melanoma' BCC 16, SK 12; BN: 137 Other: 18 (including SN, BD, and others)<br>49/232; 21% | Poor quality index test image 10 cases excluded |
| Rao 1997 | WPC-algs (tests)<br>R-CS<br>US<br>Private | Patients with atypical melanocytic lesions or suspected early malignant melanoma | 63 / 72 | VI (ABCD) [Dermoscopy]<br>Single observer | Diagnosis of melanoma | Dermatology registrar | Histology<br>MM 21<br>Atypical melanocytic nevus 51<br>21/72; 29% | None |

| Study author<br><br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| Scope 2008 | NC<br>R-CS<br>New Zealand<br>Industry image database | Images of pigmented skin lesions selected from a database of standardised patient images provided by a New Zealand–based teledermatology company (MoleMap); images were selected on the basis that (1) at least 8 clinically atypical nevi were apparent on the back; (2) most of the lesions on the back and all of the atypical nevi had close-up clinical digital images; (3) 1-year follow-up images (close-up clinical and dermoscopic images) were available to show that lesions considered to be benign were in fact biologically indolent by revealing no change; and (4) the image quality of both the overview and the close-up images were acceptable | 12 / 145 | VI (Ugly duckling)<br>Single observer | Lesion id as 'completely different' or somewhat different from the other moles; (Bx) decision. | Dermatologist | Histology or FU<br>MM 5 'malignant melanoma'<br>BN: 140<br>5/145; 3% | Unacceptable image quality |
| Troyanova 2003 | BPC/WPC-tests<br>R-CCS<br>NR<br>Training images (source NR) | Images of pigmented skin lesions selected for a dermoscopy training study | NR / 50 | VI (no algorithm)<br>[Dermoscopy]<br>Single observer | Subjective impression; dx of melanoma | Dermatologist | Histology<br>MM: 25<br>'Benign': 25<br>25/50; 50% | NR |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| Westerhoff 2000 | WPC-tests<br>R-CCS<br>Australia<br>Training images (Specialist unit) | Clinically atypical pigmented skin lesions; 50 invasive melanomas and 50 nonmelanomas randomly selected from the Sydney Melanoma Unit pigmented skin lesions (PSL) image database. | NR / 100 | VI (no algorithm)<br>[Dermoscopy]<br>Single observer | Subjective impression; dx of melanoma | GP | Histology or FU<br>MM 50<br>'Benign':50<br>50/100; 50% | None |
| NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; SK – seborrhoeic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BD – Bowen's disease; DF – dermatofibroma; FU – follow-up; R –retrospective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; VI - visual inspection; RCT - randomised controlled trial; FU - follow-up; Bx - biopsy; CMM - cutaneous malignant melanoma. | | | | | | | | |

## 14 Summary study details – detection of any skin lesion requiring excision

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| **In person** | | | | | | | | |
| Argenziano 2006<br>Any | RCT<br>Italy, Spain<br>Primary | Patients asking for screening or exhibiting one or more skin tumours as seen during routine physical examination (patient-finding screening).<br>Participating PCPs randomised to either visual inspection alone or visual inspection plus dermoscopy; only excised lesions can be included for each arm. | NR / 85 | VI (ABCD)<br>Dermoscopy (3-point checklist)<br>In person (single observer) | Subjective impression ; dx of malignancy | GPs (n=37)<br>All trained in ABCD rule | Histology<br>MM+MiS 6<br>BCC 37; SCC 10<br>Benign 32<br>53/85; 62% | Only those patients who were considered to have lesions suggestive of skin cancer had histology and could be included; rest had expert diagnosis (making full dataset ineligible for this review) |

| Study author<br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br>Diagnostic approach | Threshold | Observer qual. (n)<br>Experience | Reference standard<br>Final diagnoses<br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| Chang 2013<br>Any | NC<br>R-CS<br>Taiwan<br>Secondary | Potentially malignant biopsied or excised skin lesions (nontumour specimens excluded) | 676 / 769 | VI (no algorithm)<br>In person (single observer) | Subjective impression; definitely malignant | Dermatologists; n= 25<br>Board-certified | Histology<br>MM 4; MiS 4<br>BCC: 110; cSCC: 20<br>'Benign' diagnoses: 595<br>152/769; 20% | Poor quality index test image mis-registered or poor quality images (unfocused or containing a motion artifact) |
| Ek 2005<br>Pathway – clear<br>MM+MiS<br>Any | NC<br>P-CS<br>Australia<br>Specialist clinic | Lesions excised for which malignancy could not be excluded | 1223 / 2582 | VI (no algorithm)<br>In person | Subjective impression | Plastic surgeon (n=4 or 5; mixed experience; 3 consultants, 1 plastic surgery trainee (usually 1st year, on 6 month rotation) and a clinical assistant)<br>Unclear | Histology<br>MM+MiS 23<br>BCC 1214; SCC 517; BD 188; SK 63; 577 other benign (including 330 solar keratosis)<br>1754/2582; 68% | Incomplete or incorrectly entered proformas were excluded – 79 patients with 96 lesions |
| McGovern 1992<br>Pathway – clear | WPC-algs<br>P-CS<br>US<br>Community | PSL (>10mm) excised to rule out dysplasia, lentigo maligna or malignant melanoma | 179 / 237 | VI (7-point; (A)BCD)<br>In person; Single | 7-point: >=2, >=3, >=4 characteristics present (A)BCD: >=1, >=2, >=3 characteristics present | NR (presume dermatologist)<br>experience. NR | Histology<br>MM 6; MiS 6<br>BCC 4; SK 32; BN 138; AK 6; Other 45<br>15/192; 8% | 32 lesions unaccounted for; 13 excluded due to lesion size of 8mm or less. 192 evaluated for ABCD and 3-point; 205 evaluated for 7 point |
| Stanganelli 2000<br>Pathway – clear<br>MM+MiS<br>Any | WPC<br>R-CS<br>Italy<br>Specialist clinic | PSL referred by dermatologists and general practitioners either for pre-surgical assessment or consultation | NR / 3372 | VI (ABCD) Dermoscopy (no algorithm)<br>In person (Single) | NR Subjective impression | NR (assumed dermatologist - described as one of the co-authors; n=1) | Histology / Registry FU<br>MM+MiS 55<br>BCC 43; Benign 3274<br>98/3372; 3% | None reported |

| Study author<br><br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br><br>Diagnostic approach | Threshold | Observer qual. (n)<br><br>Experience | Reference standard<br><br>Final diagnoses<br><br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| Steiner 1987<br><br>Pathway – unclear<br><br>MM+MiS | P-CS<br><br>Austria<br><br>Specialist clinic | Small (< 10 mm) diagnostically equivocal PSL; no absolute agreement on clinical diagnosis among investigating clinicians at a pigmented lesion clinic. | NR / 318 | 1. VI (no algorithm)<br><br>2. Dermoscopy (pattern)<br><br>In person | Subjective impression | Dermatologists (n=3; High experience - "experienced dermatologists")<br><br>Consensus diagnosis of 3 observers | Histology<br><br>MM 49; MiS 24<br><br>BCC 20<br><br>BN 143; SK 20; lentigo simplex and nevoid lentigo 19; Other 15<br><br>93/318; 29% | None reported |
| Walter 2012<br><br>Pathway - clear<br><br>MM<br><br>MM+MiS<br><br>Any | BPC<br><br>RCT<br><br>UK<br><br>Primary | Any suspicious PSL that could not immediately be diagnosed as benign | 654/792 (control arm only) | VI (7-point)<br><br>Siascope (iv arm)<br><br>In person (Single) | NR | GP (n=28)<br><br>Nurse practitioner (n=2)<br><br>Low (excluded if specialist dermatology training) | Histology/Clinical FU(3-6 months)/Expert dx<br><br>Control group only:<br><br>MM 16; MiS 2<br><br>BCC 4; SK 20; DF 2; lentigo 5; 'benign' 686; unknown 10<br><br>22/773; 3% | 19 (5 due to violation of recruitment criteria or discontinued protocol; 1 died; 4 did not attend for dermatology assessment; 2 missing histology; 7 not clearly accounted for) |
| **Image-based** | | | | | | | | |
| Carli 2002b<br><br>Pathway – unclear | WPC<br><br>R-CS<br><br>Italy<br><br>Secondary | Clinically suspicious or equivocal PSL undergoing excision for diagnostic purposes; all <= 14mm diameter | NR / 57 | 1. VI (NR)<br>2. Dermoscopy (NR) | NR | Dermatologists (n=2)<br><br>Image-based; high experience ('with experience in the field of PSL'); consensus of 2 | Histology<br><br>MM 6, MiS 5<br><br>BCC 10<br><br>BN 31, SK 1; Other 4<br><br>20/54; 37% | 4 'not evaluables' excluded (1 MM, 3 benign) |
| Rosendahl 2011<br><br>Pathway – unclear | NC<br><br>R-CS<br><br>Australia<br><br>Primary | PSL submitted for histology from the primary care skin cancer practice of one author | 389 / 463 | 1. VI (no algorithm)<br><br>2. Dermoscopy (pattern) | 1. subjective impression<br><br>2. both characteristics present | Dermatologist (n=1)<br><br>Image-based; High experience (confirmed by author); Single observer | Histology<br><br>MM 9; MiS 20<br><br>BCC 72; SCC 5<br><br>BN 217; BD 18; AK 14*; BNM 140<br><br>* considered malignant by study authors<br><br>104/463; 22% | 3 poor quality images excluded |

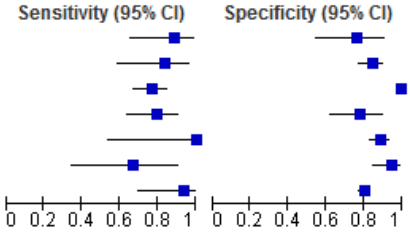
| Study author<br><br>Outcomes reported | Study type<br>Country<br>Setting | Inclusion criteria | No. patients / lesions | Index tests (algorithm)<br><br>Diagnostic approach | Threshold | Observer qual. (n)<br><br>Experience | Reference standard<br><br>Final diagnoses<br><br>Prevalence (MM+MiS) | Exclusions |
|---|---|---|---|---|---|---|---|---|
| Stanganelli 1998<br><br>Pathway – unclear | WPC<br>R-CS<br>Italy<br>Training images | PSL images selected from computerised files of the skin cancer clinic. | NR / 30 | 1. VI (no algorithm)<br>2. Dermoscopy (no algorithm) | NR | Dermatologists (n=20)<br><br>Image-based; experience NR ("experience in ELM but (with) no formal training"); average | Histology<br>MM+MiS 10<br>BCC 4<br>BN 10, SK 3, Other 3<br>14/30; 47% | None reported |

NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; cSCC – cutaneous squamous cell carcinoma; SK – seborrheic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BD – Bowen's disease; DF – dermatofibroma; FU – follow-up; R –retrospective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; GP - general practitioner; VI - visual inspection; RCT - randomised controlled trial; SCC - squamous cell carcinoma; dx - diagnosis; ELM – epiluminescence microscopy; PCP - primary care practitioner; WPC-algs - within person comparison of algorithms.

## Graphs

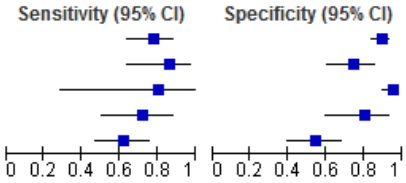
**Visual inspection - in-person (MM)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Bono 1996 | 16 | 6 | 2 | 19 | 0.89 [0.65, 0.99] | 0.76 [0.55, 0.91] |
| Green 1994 | 15 | 23 | 3 | 123 | 0.83 [0.59, 0.96] | 0.84 [0.77, 0.90] |
| Kopf 1975 | 76 | 19 | 23 | 5420 | 0.77 [0.67, 0.85] | 1.00 [0.99, 1.00] |
| Krahn 1998 | 31 | 9 | 8 | 32 | 0.79 [0.64, 0.91] | 0.78 [0.62, 0.89] |
| McGovern 1992 | 6 | 21 | 0 | 165 | 1.00 [0.54, 1.00] | 0.89 [0.83, 0.93] |
| Viglizzo 2004 | 8 | 3 | 4 | 52 | 0.67 [0.35, 0.90] | 0.95 [0.85, 0.99] |
| Walter 2012 | 15 | 152 | 1 | 605 | 0.94 [0.70, 1.00] | 0.80 [0.77, 0.83] |



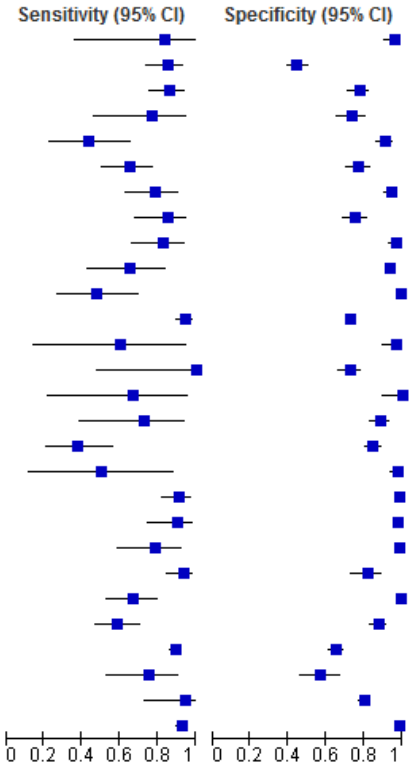

**Visual inspection - image-based (MM)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Lorentzen 1999 | 38 | 20 | 11 | 163 | 0.78 [0.63, 0.88] | 0.89 [0.84, 0.93] |
| Rao 1997 | 18 | 13 | 3 | 38 | 0.86 [0.64, 0.97] | 0.75 [0.60, 0.86] |
| Scope 2008 | 4 | 7 | 1 | 133 | 0.80 [0.28, 0.99] | 0.95 [0.90, 0.98] |
| Troyanova 2003 | 18 | 5 | 7 | 20 | 0.72 [0.51, 0.88] | 0.80 [0.59, 0.93] |
| Westerhoff 2000 | 31 | 23 | 19 | 27 | 0.62 [0.47, 0.75] | 0.54 [0.39, 0.68] |



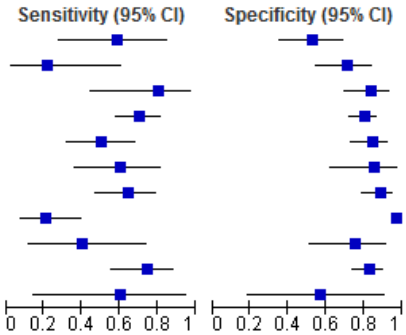
**Visual inspection - in-person (MM+MiS)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Barzegari 2005 | 5 | 5 | 1 | 111 | 0.83 [0.36, 1.00] | 0.96 [0.90, 0.99] |
| Benelli 1999 | 51 | 189 | 9 | 152 | 0.85 [0.73, 0.93] | 0.45 [0.39, 0.50] |
| Bono 2002 | 57 | 56 | 9 | 191 | 0.86 [0.76, 0.94] | 0.77 [0.72, 0.82] |
| Bono 2002b | 10 | 39 | 3 | 109 | 0.77 [0.46, 0.95] | 0.74 [0.66, 0.81] |
| Bono 2006 | 10 | 16 | 13 | 167 | 0.43 [0.23, 0.66] | 0.91 [0.86, 0.95] |
| Carli 2002 | 35 | 46 | 19 | 156 | 0.65 [0.51, 0.77] | 0.77 [0.71, 0.83] |
| Collas 1999 | 30 | 19 | 8 | 296 | 0.79 [0.63, 0.90] | 0.94 [0.91, 0.96] |
| Cristofolini 1994 | 28 | 46 | 5 | 141 | 0.85 [0.68, 0.95] | 0.75 [0.69, 0.81] |
| Cristofolini 1997 | 29 | 4 | 6 | 137 | 0.83 [0.66, 0.93] | 0.97 [0.93, 0.99] |
| Dummer 1993 | 15 | 49 | 8 | 699 | 0.65 [0.43, 0.84] | 0.93 [0.91, 0.95] |
| Ek 2005 | 11 | 23 | 12 | 2536 | 0.48 [0.27, 0.69] | 0.99 [0.99, 0.99] |
| Gachon 2005 | 141 | 1058 | 8 | 2829 | 0.95 [0.90, 0.98] | 0.73 [0.71, 0.74] |
| Green 1991 | 3 | 2 | 2 | 63 | 0.60 [0.15, 0.95] | 0.97 [0.89, 1.00] |
| Grimaldi 2009 | 5 | 63 | 0 | 167 | 1.00 [0.48, 1.00] | 0.73 [0.66, 0.78] |
| Langley 2001 | 4 | 0 | 2 | 34 | 0.67 [0.22, 0.96] | 1.00 [0.90, 1.00] |
| McGovern 1992 | 8 | 21 | 3 | 160 | 0.73 [0.39, 0.94] | 0.88 [0.83, 0.93] |
| Menzies 2009 | 12 | 46 | 20 | 253 | 0.38 [0.21, 0.56] | 0.85 [0.80, 0.89] |
| Morales Callaghan 2008 | 3 | 5 | 3 | 189 | 0.50 [0.12, 0.88] | 0.97 [0.94, 0.99] |
| Morton 1998a | 63 | 10 | 6 | 684 | 0.91 [0.82, 0.97] | 0.99 [0.97, 0.99] |
| Morton 1998b | 28 | 14 | 3 | 522 | 0.90 [0.74, 0.98] | 0.97 [0.96, 0.99] |
| Morton 1998c | 22 | 11 | 6 | 630 | 0.79 [0.59, 0.92] | 0.98 [0.97, 0.99] |
| Soyer 1995 | 61 | 17 | 4 | 77 | 0.94 [0.85, 0.98] | 0.82 [0.73, 0.89] |
| Stanganelli 2000 | 37 | 21 | 18 | 3296 | 0.67 [0.53, 0.79] | 0.99 [0.99, 1.00] |
| Steiner 1987 | 43 | 31 | 30 | 214 | 0.59 [0.47, 0.70] | 0.87 [0.83, 0.91] |
| Thomas 1998 | 411 | 236 | 49 | 444 | 0.89 [0.86, 0.92] | 0.65 [0.62, 0.69] |
| Unlu 2014 | 18 | 39 | 6 | 52 | 0.75 [0.53, 0.90] | 0.57 [0.46, 0.67] |
| Walter 2012 | 17 | 152 | 1 | 603 | 0.94 [0.73, 1.00] | 0.80 [0.77, 0.83] |
| Zaumseil 1983 | 312 | 76 | 25 | 6650 | 0.93 [0.89, 0.95] | 0.99 [0.99, 0.99] |



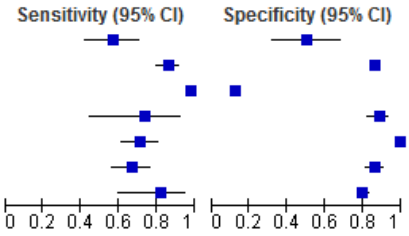

**Visual inspection - image-based (MM+MiS)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 2001 | 7 | 18 | 5 | 20 | 0.58 [0.28, 0.85] | 0.53 [0.36, 0.69] |
| Bourne 2012 | 2 | 12 | 7 | 29 | 0.22 [0.03, 0.60] | 0.71 [0.54, 0.84] |
| Carli 2002b | 8 | 7 | 2 | 36 | 0.80 [0.44, 0.97] | 0.84 [0.69, 0.93] |
| Carli 2003b | 45 | 27 | 19 | 109 | 0.70 [0.58, 0.81] | 0.80 [0.72, 0.86] |
| de Giorgi 2012 | 17 | 11 | 17 | 58 | 0.50 [0.32, 0.68] | 0.84 [0.73, 0.92] |
| Dolianitis 2005 | 12 | 3 | 8 | 17 | 0.60 [0.36, 0.81] | 0.85 [0.62, 0.97] |
| Pizzichetta 2004 | 25 | 8 | 14 | 61 | 0.64 [0.47, 0.79] | 0.88 [0.78, 0.95] |
| Rosendahl 2011 | 6 | 14 | 23 | 420 | 0.21 [0.08, 0.40] | 0.97 [0.95, 0.98] |
| Stanganelli 1998 | 4 | 5 | 6 | 15 | 0.40 [0.12, 0.74] | 0.75 [0.51, 0.91] |
| Stanganelli 2005 | 23 | 18 | 8 | 85 | 0.74 [0.55, 0.88] | 0.83 [0.74, 0.89] |
| Winkelmann 2016 | 3 | 3 | 2 | 4 | 0.60 [0.15, 0.95] | 0.57 [0.18, 0.90] |



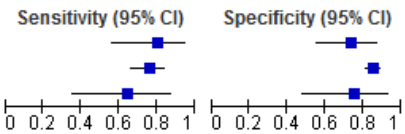

**Visual inspection - in-person (Any)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Argenziano 2006 | 30 | 16 | 23 | 16 | 0.57 [0.42, 0.70] | 0.50 [0.32, 0.68] |
| Chang 2013 | 131 | 84 | 21 | 533 | 0.86 [0.80, 0.91] | 0.86 [0.83, 0.89] |
| Ek 2005 | 1711 | 722 | 43 | 106 | 0.98 [0.97, 0.98] | 0.13 [0.11, 0.15] |
| McGovern 1992 | 11 | 21 | 4 | 156 | 0.73 [0.45, 0.92] | 0.88 [0.82, 0.93] |
| Stanganelli 2000 | 70 | 29 | 28 | 3245 | 0.71 [0.61, 0.80] | 0.99 [0.99, 0.99] |
| Steiner 1987 | 62 | 31 | 31 | 194 | 0.67 [0.56, 0.76] | 0.86 [0.81, 0.90] |
| Walter 2012 | 18 | 152 | 4 | 599 | 0.82 [0.60, 0.95] | 0.80 [0.77, 0.83] |



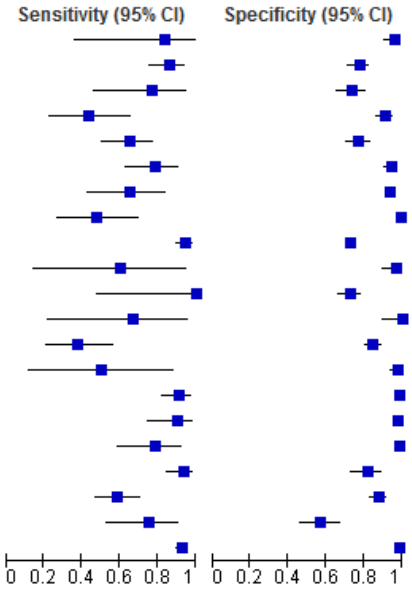

**Visual inspection - image-based (Any)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Carli 2002b | 16 | 9 | 4 | 25 | 0.80 [0.56, 0.94] | 0.74 [0.56, 0.87] |
| Rosendahl 2011 | 79 | 54 | 25 | 305 | 0.76 [0.67, 0.84] | 0.85 [0.81, 0.88] |
| Stanganelli 1998 | 9 | 4 | 5 | 12 | 0.64 [0.35, 0.87] | 0.75 [0.48, 0.93] |



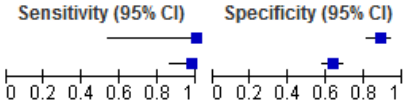
**MM2- VI - in-person - no algorithm**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Barzegari 2005 | 5 | 5 | 1 | 111 | 0.83 [0.36, 1.00] | 0.96 [0.90, 0.99] |
| Bono 2002 | 57 | 56 | 9 | 191 | 0.86 [0.76, 0.94] | 0.77 [0.72, 0.82] |
| Bono 2002b | 10 | 39 | 3 | 109 | 0.77 [0.46, 0.95] | 0.74 [0.66, 0.81] |
| Bono 2006 | 10 | 16 | 13 | 167 | 0.43 [0.23, 0.66] | 0.91 [0.86, 0.95] |
| Carli 2002 | 35 | 46 | 19 | 156 | 0.65 [0.51, 0.77] | 0.77 [0.71, 0.83] |
| Collas 1999 | 30 | 19 | 8 | 296 | 0.79 [0.63, 0.90] | 0.94 [0.91, 0.96] |
| Dummer 1993 | 15 | 49 | 8 | 699 | 0.65 [0.43, 0.84] | 0.93 [0.91, 0.95] |
| Ek 2005 | 11 | 23 | 12 | 2536 | 0.48 [0.27, 0.69] | 0.99 [0.99, 0.99] |
| Gachon 2005 | 141 | 1058 | 8 | 2829 | 0.95 [0.90, 0.98] | 0.73 [0.71, 0.74] |
| Green 1991 | 3 | 2 | 2 | 63 | 0.60 [0.15, 0.95] | 0.97 [0.89, 1.00] |
| Grimaldi 2009 | 5 | 63 | 0 | 167 | 1.00 [0.48, 1.00] | 0.73 [0.66, 0.78] |
| Langley 2001 | 4 | 0 | 2 | 34 | 0.67 [0.22, 0.96] | 1.00 [0.90, 1.00] |
| Menzies 2009 | 12 | 46 | 20 | 253 | 0.38 [0.21, 0.56] | 0.85 [0.80, 0.89] |
| Morales Callaghan 2008 | 3 | 5 | 3 | 189 | 0.50 [0.12, 0.88] | 0.97 [0.94, 0.99] |
| Morton 1998a | 63 | 10 | 6 | 684 | 0.91 [0.82, 0.97] | 0.99 [0.97, 0.99] |
| Morton 1998b | 28 | 14 | 3 | 522 | 0.90 [0.74, 0.98] | 0.97 [0.96, 0.99] |
| Morton 1998c | 22 | 11 | 6 | 630 | 0.79 [0.59, 0.92] | 0.98 [0.97, 0.99] |
| Soyer 1995 | 61 | 17 | 4 | 77 | 0.94 [0.85, 0.98] | 0.82 [0.73, 0.89] |
| Steiner 1987 | 43 | 31 | 30 | 214 | 0.59 [0.47, 0.70] | 0.87 [0.83, 0.91] |
| Unlu 2014 | 18 | 39 | 6 | 52 | 0.75 [0.53, 0.90] | 0.57 [0.46, 0.67] |
| Zaumseil 1983 | 312 | 76 | 25 | 6650 | 0.93 [0.89, 0.95] | 0.99 [0.99, 0.99] |



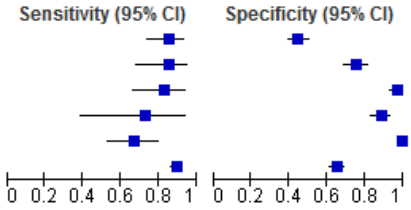

**MM2- VI - in-person - no algorithm (alt thresholds)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Barzegari 2005 | 6 | 13 | 0 | 103 | 1.00 [0.54, 1.00] | 0.89 [0.82, 0.94] |
| Collas 1999 | 37 | 114 | 1 | 201 | 0.97 [0.86, 1.00] | 0.64 [0.58, 0.69] |



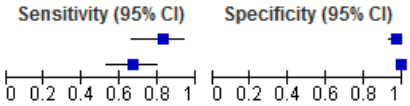

**MM2- VI - in-person - (A)BCD(E) at NR or standard threshold**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 1999 | 51 | 189 | 9 | 152 | 0.85 [0.73, 0.93] | 0.45 [0.39, 0.50] |
| Cristofolini 1994 | 28 | 46 | 5 | 141 | 0.85 [0.68, 0.95] | 0.75 [0.69, 0.81] |
| Cristofolini 1997 | 29 | 4 | 6 | 137 | 0.83 [0.66, 0.93] | 0.97 [0.93, 0.99] |
| McGovern 1992 | 8 | 21 | 3 | 160 | 0.73 [0.39, 0.94] | 0.88 [0.83, 0.93] |
| Stanganelli 2000 | 37 | 21 | 18 | 3296 | 0.67 [0.53, 0.79] | 0.99 [0.99, 1.00] |
| Thomas 1998 | 411 | 236 | 49 | 444 | 0.89 [0.86, 0.92] | 0.65 [0.62, 0.69] |



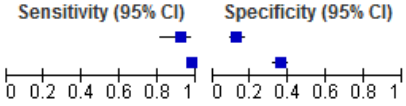

**MM2-VI - in-person - ABCD at NR**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Cristofolini 1997 | 29 | 4 | 6 | 137 | 0.83 [0.66, 0.93] | 0.97 [0.93, 0.99] |
| Stanganelli 2000 | 37 | 21 | 18 | 3296 | 0.67 [0.53, 0.79] | 0.99 [0.99, 1.00] |



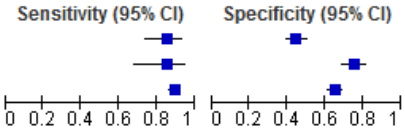

**MM2-VI - in-person - ABCDE at >=1**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 1999 | 55 | 297 | 5 | 44 | 0.92 [0.82, 0.97] | 0.13 [0.10, 0.17] |
| Thomas 1998 | 448 | 435 | 12 | 245 | 0.97 [0.95, 0.99] | 0.36 [0.32, 0.40] |



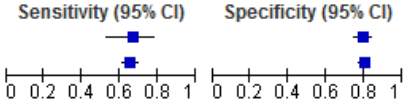

**MM2-VI - in-person - ABCDE at >=2**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 1999 | 51 | 189 | 9 | 152 | 0.85 [0.73, 0.93] | 0.45 [0.39, 0.50] |
| Cristofolini 1994 | 28 | 46 | 5 | 141 | 0.85 [0.68, 0.95] | 0.75 [0.69, 0.81] |
| Thomas 1998 | 411 | 236 | 49 | 444 | 0.89 [0.86, 0.92] | 0.65 [0.62, 0.69] |



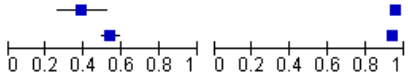

**MM2-VI - in-person - ABCDE at >=3**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 1999 | 40 | 70 | 20 | 271 | 0.67 [0.53, 0.78] | 0.79 [0.75, 0.84] |
| Thomas 1998 | 301 | 136 | 159 | 544 | 0.65 [0.61, 0.70] | 0.80 [0.77, 0.83] |



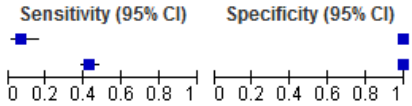
**MM2-VI - in-person - ABCDE at >=4**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Benelli 1999 | 23 | 15 | 37 | 326 | 0.38 [0.26, 0.52] | 0.96 [0.93, 0.98] | | |
| Thomas 1998 | 248 | 44 | 212 | 636 | 0.54 [0.49, 0.59] | 0.94 [0.91, 0.95] | | |



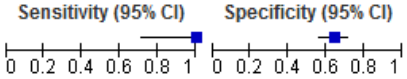

**MM2-VI - in-person - ABCDE at >=5**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Benelli 1999 | 4 | 2 | 56 | 339 | 0.07 [0.02, 0.16] | 0.99 [0.98, 1.00] | | |
| Thomas 1998 | 198 | 3 | 262 | 677 | 0.43 [0.38, 0.48] | 1.00 [0.99, 1.00] | | |



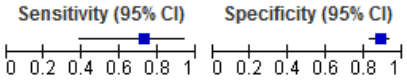

**MM2-VI - in-person - BCD at >=1**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| McGovern 1992 | 11 | 65 | 0 | 116 | 1.00 [0.72, 1.00] | 0.64 [0.57, 0.71] | | |



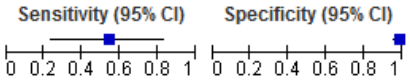

**MM2-VI - in-person - BCD at >=2**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| McGovern 1992 | 8 | 21 | 3 | 160 | 0.73 [0.39, 0.94] | 0.88 [0.83, 0.93] | | |



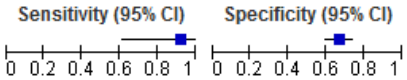

**MM2-VI - in-person - BCD at >=3**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| McGovern 1992 | 6 | 3 | 5 | 178 | 0.55 [0.23, 0.83] | 0.98 [0.95, 1.00] | | |



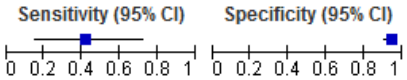

**MM2-VI - in-person - 7point at >=2**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| McGovern 1992 | 11 | 64 | 1 | 129 | 0.92 [0.62, 1.00] | 0.67 [0.60, 0.73] | | |



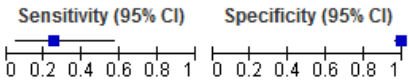

**MM2-VI - in-person - 7point at >=3**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| McGovern 1992 | 5 | 11 | 7 | 182 | 0.42 [0.15, 0.72] | 0.94 [0.90, 0.97] | | |



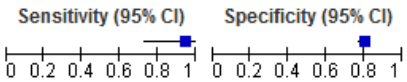

**MM2-VI - in-person - 7point at >=4**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| McGovern 1992 | 3 | 2 | 9 | 191 | 0.25 [0.05, 0.57] | 0.99 [0.96, 1.00] | | |



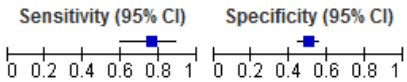

**MM2-VI - in-person - 7point(rev) at >=3**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Walter 2012 | 17 | 152 | 1 | 603 | 0.94 [0.73, 1.00] | 0.80 [0.77, 0.83] | | |



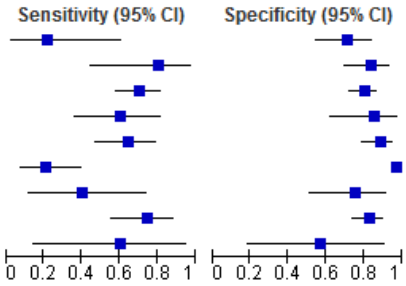

**MM2-VI - in-person - Collas at >=1**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Collas 1999 | 29 | 157 | 9 | 158 | 0.76 [0.60, 0.89] | 0.50 [0.44, 0.56] | | |



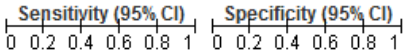
**MM2- VI - image-based - no algorithm**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Bourne 2012 | 2 | 12 | 7 | 29 | 0.22 [0.03, 0.60] | 0.71 [0.54, 0.84] |
| Carli 2002b | 8 | 7 | 2 | 36 | 0.80 [0.44, 0.97] | 0.84 [0.69, 0.93] |
| Carli 2003b | 45 | 27 | 19 | 109 | 0.70 [0.58, 0.81] | 0.80 [0.72, 0.86] |
| Dolianitis 2005 | 12 | 3 | 8 | 17 | 0.60 [0.36, 0.81] | 0.85 [0.62, 0.97] |
| Pizzichetta 2004 | 25 | 8 | 14 | 61 | 0.64 [0.47, 0.79] | 0.88 [0.78, 0.95] |
| Rosendahl 2011 | 6 | 14 | 23 | 420 | 0.21 [0.08, 0.40] | 0.97 [0.95, 0.98] |
| Stanganelli 1998 | 4 | 5 | 6 | 15 | 0.40 [0.12, 0.74] | 0.75 [0.51, 0.91] |
| Stanganelli 2005 | 23 | 18 | 8 | 85 | 0.74 [0.55, 0.88] | 0.83 [0.74, 0.89] |
| Winkelmann 2016 | 3 | 3 | 2 | 4 | 0.60 [0.15, 0.95] | 0.57 [0.18, 0.90] |



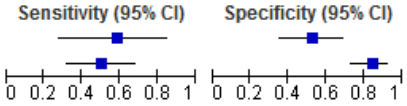

**MM2- VI - image-based - no algorithm (alt threshold)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|



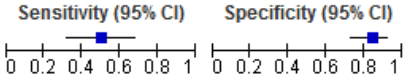

**MM2-VI - image-based - ABCD(E) at standard**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 2001 | 7 | 18 | 5 | 20 | 0.58 [0.28, 0.85] | 0.53 [0.36, 0.69] |
| de Giorgi 2012 | 17 | 11 | 17 | 58 | 0.50 [0.32, 0.68] | 0.84 [0.73, 0.92] |



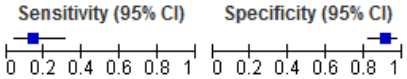

**MM2-VI - image-based - ABCD at >=2**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| de Giorgi 2012 | 17 | 11 | 17 | 58 | 0.50 [0.32, 0.68] | 0.84 [0.73, 0.92] |



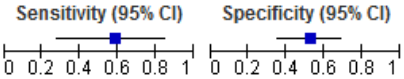

**MM2-VI - image-based - ABCD at >=3**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| de Giorgi 2012 | 5 | 6 | 29 | 63 | 0.15 [0.05, 0.31] | 0.91 [0.82, 0.97] |



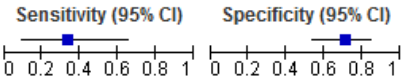

**MM2-VI - image-based - ABCDE at >=2**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 2001 | 7 | 18 | 5 | 20 | 0.58 [0.28, 0.85] | 0.53 [0.36, 0.69] |



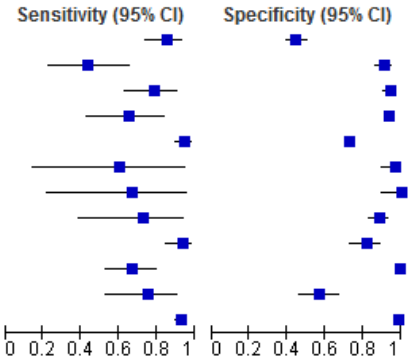

**MM2-VI - image-based - ABCDE at >=3**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 2001 | 4 | 11 | 8 | 27 | 0.33 [0.10, 0.65] | 0.71 [0.54, 0.85] |



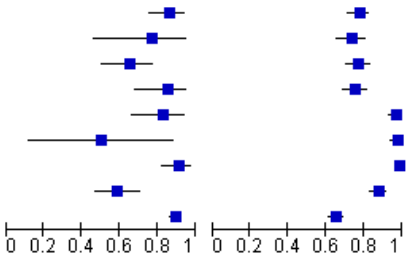

**MM2- VI - in-person - experience NR**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 1999 | 51 | 189 | 9 | 152 | 0.85 [0.73, 0.93] | 0.45 [0.39, 0.50] |
| Bono 2006 | 10 | 16 | 13 | 167 | 0.43 [0.23, 0.66] | 0.91 [0.86, 0.95] |
| Collas 1999 | 30 | 19 | 8 | 296 | 0.79 [0.63, 0.90] | 0.94 [0.91, 0.96] |
| Dummer 1993 | 15 | 49 | 8 | 699 | 0.65 [0.43, 0.84] | 0.93 [0.91, 0.95] |
| Gachon 2005 | 141 | 1058 | 8 | 2829 | 0.95 [0.90, 0.98] | 0.73 [0.71, 0.74] |
| Green 1991 | 3 | 2 | 2 | 63 | 0.60 [0.15, 0.95] | 0.97 [0.89, 1.00] |
| Langley 2001 | 4 | 0 | 2 | 34 | 0.67 [0.22, 0.96] | 1.00 [0.90, 1.00] |
| McGovern 1992 | 8 | 21 | 3 | 160 | 0.73 [0.39, 0.94] | 0.88 [0.83, 0.93] |
| Soyer 1995 | 61 | 17 | 4 | 77 | 0.94 [0.85, 0.98] | 0.82 [0.73, 0.89] |
| Stanganelli 2000 | 37 | 21 | 18 | 3296 | 0.67 [0.53, 0.79] | 0.99 [0.99, 1.00] |
| Unlu 2014 | 18 | 39 | 6 | 52 | 0.75 [0.53, 0.90] | 0.57 [0.46, 0.67] |
| Zaumseil 1983 | 312 | 76 | 25 | 6650 | 0.93 [0.89, 0.95] | 0.99 [0.99, 0.99] |



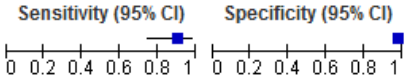
**MM2- VI - in-person - experience High**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Bono 2002 | 57 | 56 | 9 | 191 | 0.86 [0.76, 0.94] | 0.77 [0.72, 0.82] |
| Bono 2002b | 10 | 39 | 3 | 109 | 0.77 [0.46, 0.95] | 0.74 [0.66, 0.81] |
| Carli 2002 | 35 | 46 | 19 | 156 | 0.65 [0.51, 0.77] | 0.77 [0.71, 0.83] |
| Cristofolini 1994 | 28 | 46 | 5 | 141 | 0.85 [0.68, 0.95] | 0.75 [0.69, 0.81] |
| Cristofolini 1997 | 29 | 4 | 6 | 137 | 0.83 [0.66, 0.93] | 0.97 [0.93, 0.99] |
| Morales Callaghan 2008 | 3 | 5 | 3 | 189 | 0.50 [0.12, 0.88] | 0.97 [0.94, 0.99] |
| Morton 1998a | 63 | 10 | 6 | 684 | 0.91 [0.82, 0.97] | 0.99 [0.97, 0.99] |
| Steiner 1987 | 43 | 31 | 30 | 214 | 0.59 [0.47, 0.70] | 0.87 [0.83, 0.91] |
| Thomas 1998 | 411 | 236 | 49 | 444 | 0.89 [0.86, 0.92] | 0.65 [0.62, 0.69] |



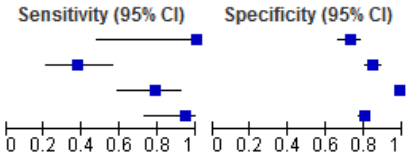

**MM2- VI - in-person - experience Moderate**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Morton 1998b | 28 | 14 | 3 | 522 | 0.90 [0.74, 0.98] | 0.97 [0.96, 0.99] |



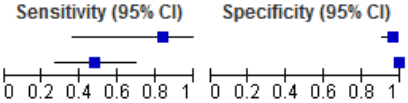

**MM2- VI - in-person - experience Low**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Grimaldi 2009 | 5 | 63 | 0 | 167 | 1.00 [0.48, 1.00] | 0.73 [0.66, 0.78] |
| Menzies 2009 | 12 | 46 | 20 | 253 | 0.38 [0.21, 0.56] | 0.85 [0.80, 0.89] |
| Morton 1998c | 22 | 11 | 6 | 630 | 0.79 [0.59, 0.92] | 0.98 [0.97, 0.99] |
| Walter 2012 | 17 | 152 | 1 | 603 | 0.94 [0.73, 1.00] | 0.80 [0.77, 0.83] |



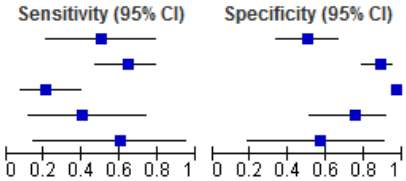

**MM2- VI - in-person - experience Mixed**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Barzegari 2005 | 5 | 5 | 1 | 111 | 0.83 [0.36, 1.00] | 0.96 [0.90, 0.99] |
| Ek 2005 | 11 | 23 | 12 | 2536 | 0.48 [0.27, 0.69] | 0.99 [0.99, 0.99] |



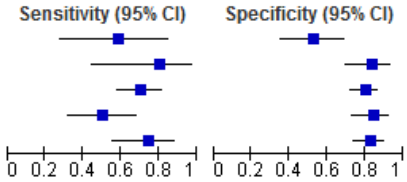

**MM2- VI - image-based - experience NR**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 2001 | 6 | 19 | 6 | 19 | 0.50 [0.21, 0.79] | 0.50 [0.33, 0.67] |
| Pizzichetta 2004 | 25 | 8 | 14 | 61 | 0.64 [0.47, 0.79] | 0.88 [0.78, 0.95] |
| Rosendahl 2011 | 6 | 14 | 23 | 420 | 0.21 [0.08, 0.40] | 0.97 [0.95, 0.98] |
| Stanganelli 1998 | 4 | 5 | 6 | 15 | 0.40 [0.12, 0.74] | 0.75 [0.51, 0.91] |
| Winkelmann 2016 | 3 | 3 | 2 | 4 | 0.60 [0.15, 0.95] | 0.57 [0.18, 0.90] |



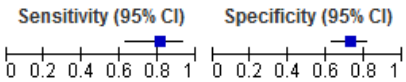

**MM2- VI - image-based - experience High**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 2001 | 7 | 18 | 5 | 20 | 0.58 [0.28, 0.85] | 0.53 [0.36, 0.69] |
| Carli 2002b | 8 | 7 | 2 | 36 | 0.80 [0.44, 0.97] | 0.84 [0.69, 0.93] |
| Carli 2003b | 45 | 27 | 19 | 109 | 0.70 [0.58, 0.81] | 0.80 [0.72, 0.86] |
| de Giorgi 2012 | 17 | 11 | 17 | 58 | 0.50 [0.32, 0.68] | 0.84 [0.73, 0.92] |
| Stanganelli 2005 | 23 | 18 | 8 | 85 | 0.74 [0.55, 0.88] | 0.83 [0.74, 0.89] |



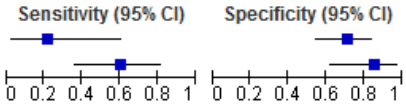

**MM2- VI - image-based - experience Low**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Stanganelli 2005 | 25 | 28 | 6 | 75 | 0.81 [0.63, 0.93] | 0.73 [0.63, 0.81] |



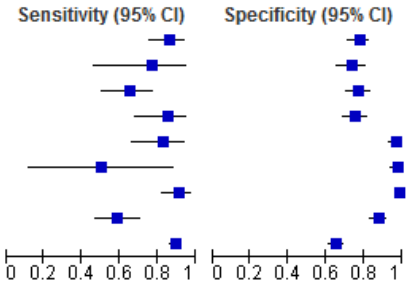

**MM2- VI - image-based - experience Mixed**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Bourne 2012 | 2 | 12 | 7 | 29 | 0.22 [0.03, 0.60] | 0.71 [0.54, 0.84] |
| Dolianitis 2005 | 12 | 3 | 8 | 17 | 0.60 [0.36, 0.81] | 0.85 [0.62, 0.97] |



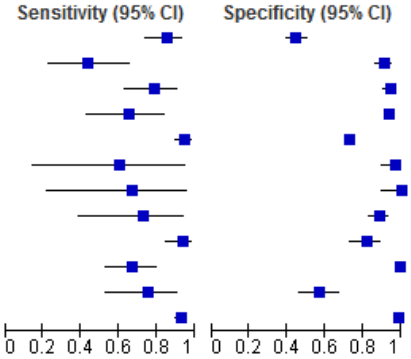
**VI - in-person - Expert consultant (MM+MiS)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Bono 2002 | 57 | 56 | 9 | 191 | 0.86 [0.76, 0.94] | 0.77 [0.72, 0.82] |
| Bono 2002b | 10 | 39 | 3 | 109 | 0.77 [0.46, 0.95] | 0.74 [0.66, 0.81] |
| Carli 2002 | 35 | 46 | 19 | 156 | 0.65 [0.51, 0.77] | 0.77 [0.71, 0.83] |
| Cristofolini 1994 | 28 | 46 | 5 | 141 | 0.85 [0.68, 0.95] | 0.75 [0.69, 0.81] |
| Cristofolini 1997 | 29 | 4 | 6 | 137 | 0.83 [0.66, 0.93] | 0.97 [0.93, 0.99] |
| Morales Callaghan 2008 | 3 | 5 | 3 | 189 | 0.50 [0.12, 0.88] | 0.97 [0.94, 0.99] |
| Morton 1998a | 63 | 10 | 6 | 684 | 0.91 [0.82, 0.97] | 0.99 [0.97, 0.99] |
| Steiner 1987 | 43 | 31 | 30 | 214 | 0.59 [0.47, 0.70] | 0.87 [0.83, 0.91] |
| Thomas 1998 | 411 | 236 | 49 | 444 | 0.89 [0.86, 0.92] | 0.65 [0.62, 0.69] |



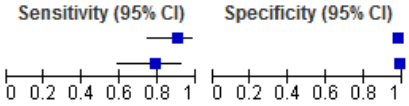

**VI - in-person - Consultant (MM+MiS)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 1999 | 51 | 189 | 9 | 152 | 0.85 [0.73, 0.93] | 0.45 [0.39, 0.50] |
| Bono 2006 | 10 | 16 | 13 | 167 | 0.43 [0.23, 0.66] | 0.91 [0.86, 0.95] |
| Collas 1999 | 30 | 19 | 8 | 296 | 0.79 [0.63, 0.90] | 0.94 [0.91, 0.96] |
| Dummer 1993 | 15 | 49 | 8 | 699 | 0.65 [0.43, 0.84] | 0.93 [0.91, 0.95] |
| Gachon 2005 | 141 | 1058 | 8 | 2829 | 0.95 [0.90, 0.98] | 0.73 [0.71, 0.74] |
| Green 1991 | 3 | 2 | 2 | 63 | 0.60 [0.15, 0.95] | 0.97 [0.89, 1.00] |
| Langley 2001 | 4 | 0 | 2 | 34 | 0.67 [0.22, 0.96] | 1.00 [0.90, 1.00] |
| McGovern 1992 | 8 | 21 | 3 | 160 | 0.73 [0.39, 0.94] | 0.88 [0.83, 0.93] |
| Soyer 1995 | 61 | 17 | 4 | 77 | 0.94 [0.85, 0.98] | 0.82 [0.73, 0.89] |
| Stanganelli 2000 | 37 | 21 | 18 | 3296 | 0.67 [0.53, 0.79] | 0.99 [0.99, 1.00] |
| Unlu 2014 | 18 | 39 | 6 | 52 | 0.75 [0.53, 0.90] | 0.57 [0.46, 0.67] |
| Zaumseil 1983 | 312 | 76 | 25 | 6650 | 0.93 [0.89, 0.95] | 0.99 [0.99, 0.99] |



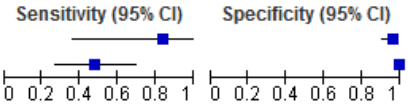

**VI - in-person - Resident/registrar (MM+MiS)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Morton 1998b | 28 | 14 | 3 | 522 | 0.90 [0.74, 0.98] | 0.97 [0.96, 0.99] |
| Morton 1998c | 22 | 11 | 6 | 630 | 0.79 [0.59, 0.92] | 0.98 [0.97, 0.99] |



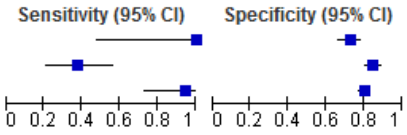

**VI - in-person - Mixed qualifications (secondary care) (MM+MiS)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Barzegari 2005 | 5 | 5 | 1 | 111 | 0.83 [0.36, 1.00] | 0.96 [0.90, 0.99] |
| Ek 2005 | 11 | 23 | 12 | 2536 | 0.48 [0.27, 0.69] | 0.99 [0.99, 0.99] |



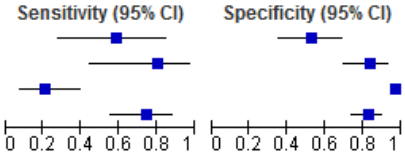

**VI - in-person - GP (MM+MiS)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Grimaldi 2009 | 5 | 63 | 0 | 167 | 1.00 [0.48, 1.00] | 0.73 [0.66, 0.78] |
| Menzies 2009 | 12 | 46 | 20 | 253 | 0.38 [0.21, 0.56] | 0.85 [0.80, 0.89] |
| Walter 2012 | 17 | 152 | 1 | 603 | 0.94 [0.73, 1.00] | 0.80 [0.77, 0.83] |



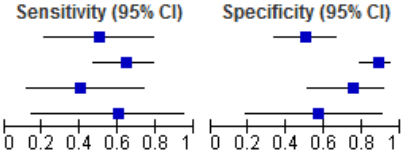

**MM2- VI - image-based - Expert consultant**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 2001 | 7 | 18 | 5 | 20 | 0.58 [0.28, 0.85] | 0.53 [0.36, 0.69] |
| Carli 2002b | 8 | 7 | 2 | 36 | 0.80 [0.44, 0.97] | 0.84 [0.69, 0.93] |
| Rosendahl 2011 | 6 | 14 | 23 | 420 | 0.21 [0.08, 0.40] | 0.97 [0.95, 0.98] |
| Stanganelli 2005 | 23 | 18 | 8 | 85 | 0.74 [0.55, 0.88] | 0.83 [0.74, 0.89] |



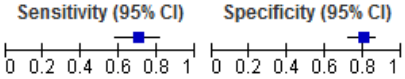

**MM2- VI - image-based - Consultant**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|
| Benelli 2001 | 6 | 19 | 6 | 19 | 0.50 [0.21, 0.79] | 0.50 [0.33, 0.67] |
| Pizzichetta 2004 | 25 | 8 | 14 | 61 | 0.64 [0.47, 0.79] | 0.88 [0.78, 0.95] |
| Stanganelli 1998 | 4 | 5 | 6 | 15 | 0.40 [0.12, 0.74] | 0.75 [0.51, 0.91] |
| Winkelmann 2016 | 3 | 3 | 2 | 4 | 0.60 [0.15, 0.95] | 0.57 [0.18, 0.90] |



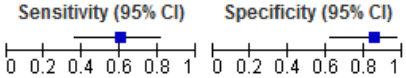
**MM2- VI - image-based - Mixed qualifications (secondary care)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Carli 2003b | 45 | 27 | 19 | 109 | 0.70 [0.58, 0.81] | 0.80 [0.72, 0.86] | | |



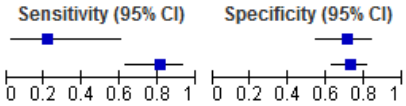

**MM2- VI - image-based - Mixed qualifications (secondary/primary care)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Dolianitis 2005 | 12 | 3 | 8 | 17 | 0.60 [0.36, 0.81] | 0.85 [0.62, 0.97] | | |



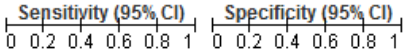

**MM2- VI - image-based - Mixed qualifications (primary care)**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Bourne 2012 | 2 | 12 | 7 | 29 | 0.22 [0.03, 0.60] | 0.71 [0.54, 0.84] | | |
| Stanganelli 2005 | 25 | 28 | 6 | 75 | 0.81 [0.63, 0.93] | 0.73 [0.63, 0.81] | | |



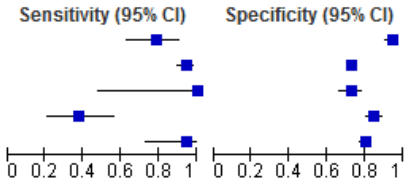

**MM2 - VI - image-based qual not reported**

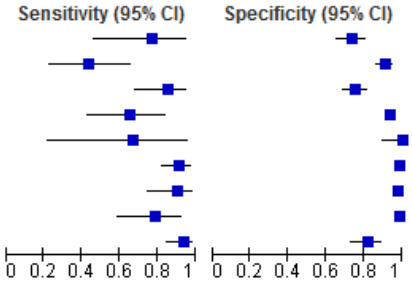
| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|




**MM2 - Selected on quality - pathway 2 or 3**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Collas 1999 | 30 | 19 | 8 | 296 | 0.79 [0.63, 0.90] | 0.94 [0.91, 0.96] | | |
| Gachon 2005 | 141 | 1058 | 8 | 2829 | 0.95 [0.90, 0.98] | 0.73 [0.71, 0.74] | | |
| Grimaldi 2009 | 5 | 63 | 0 | 167 | 1.00 [0.48, 1.00] | 0.73 [0.66, 0.78] | | |
| Menzies 2009 | 12 | 46 | 20 | 253 | 0.38 [0.21, 0.56] | 0.85 [0.80, 0.89] | | |
| Walter 2012 | 17 | 152 | 1 | 603 | 0.94 [0.73, 1.00] | 0.80 [0.77, 0.83] | | |




**MM2 - Selected on quality - pathway 5**

| Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Bono 2002b | 10 | 39 | 3 | 109 | 0.77 [0.46, 0.95] | 0.74 [0.66, 0.81] | | |
| Bono 2006 | 10 | 16 | 13 | 167 | 0.43 [0.23, 0.66] | 0.91 [0.86, 0.95] | | |
| Cristofolini 1994 | 28 | 46 | 5 | 141 | 0.85 [0.68, 0.95] | 0.75 [0.69, 0.81] | | |
| Dummer 1993 | 15 | 49 | 8 | 699 | 0.65 [0.43, 0.84] | 0.93 [0.91, 0.95] | | |
| Langley 2001 | 4 | 0 | 2 | 34 | 0.67 [0.22, 0.96] | 1.00 [0.90, 1.00] | | |
| Morton 1998a | 63 | 10 | 6 | 684 | 0.91 [0.82, 0.97] | 0.99 [0.97, 0.99] | | |
| Morton 1998b | 28 | 14 | 3 | 522 | 0.90 [0.74, 0.98] | 0.97 [0.96, 0.99] | | |
| Morton 1998c | 22 | 11 | 6 | 630 | 0.79 [0.59, 0.92] | 0.98 [0.97, 0.99] | | |
| Soyer 1995 | 61 | 17 | 4 | 77 | 0.94 [0.85, 0.98] | 0.82 [0.73, 0.89] | | |