

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Training general practitioners in melanoma diagnosis: a scoping review of the literature.
<b>AUTHORS</b>	Harkemanne, Evelyne; Baeck, Marie; Tromme, Isabelle

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Machet, Laurent Centre Hospitalier Regional Universitaire and University of Tours, France
<b>REVIEW RETURNED</b>	28-Sep-2020

<b>GENERAL COMMENTS</b>	This article aims to determine whether teaching general practitioners to diagnose melanoma is associated with better, more accurate and earlier recognition of melanoma. The methodology is adequate. The results are in line with what could be expected, i.e. education programs based on physical examination or dermoscopy improve diagnostic accuracy. The question that remains open is the optimal form and duration of the training, and the need to repeat the training at regular intervals
-------------------------	--

<b>REVIEWER</b>	Owain Jones Department of Public Health & Primary Care, University of Cambridge, UK
<b>REVIEW RETURNED</b>	05-Oct-2020

<b>GENERAL COMMENTS</b>	<p>The paper sets out to perform a scoping review of training programs for general practitioners in melanoma diagnosis, in particular to look for evidence on the long-term efficacy of the training. You establish the logic for taking this approach well in the introduction. The methods including search strategies are described well and in detail. Overall I feel that the study addresses a relevant aspect of training in this area, and draws some important conclusions on the long-term efficacy of training and the need to establish the required frequency of training updates to maintain clinical skills.</p> <p>Further comments:</p> <ol style="list-style-type: none"><li>1. In the abstract it states that the databases were searched for relevant articles between February and May 2020 - this may be confusing and it may be better to clarify in the abstract that the search was conducted from 1995 to May 2020.</li><li>2. There is no mention of the methodology that you follow for the scoping review. The framework developed by Arksey and O'Malley (2005) and subsequently refined by Levac (2010) is often cited.</li></ol>
-------------------------	--

	<p>3. There is also no mention of a Risk of Bias assessment of the included studies. This is not compulsory in scoping reviews, but is sometimes performed. Did you assess the papers for Bias in any way?</p> <p>4. You signify in table 4 whether a result was statistically significant or not, did all of the studies included use the same definition of statistical significance, or was there variability between different studies?</p> <p>5. Very few of the studies used the 7-point checklist as their clinical algorithm, this is slightly surprising as in the UK this is probably the most widely used diagnostic algorithm. However, the studies you have identified are international, and this result may simply reflect that it is not so widely used around the world.</p> <p>6. You don't comment on the relative validity of the various assessment methods employed by the included studies. For example, is confidence or number of total-body skin examinations performed a good measure of diagnostic ability, and thus a useful measure with which to evaluate a training program?</p> <p>7. A large proportion of papers identified were from other sources according to the PRISMA diagram. It may be worth commenting on this in the manuscript. Was there a reason for this? Were many of the additional studies included, or were the included studies mostly from the original literature searches?</p> <p>8. I find the discussion slightly confusing. I think there are a couple of reasons for this. There are elements of the discussion that could be included in the results section, paragraph 2 for example and some of the information on page 11. I also think the structure could be clarified and the writing more concise. There are also some sentences which might benefit from input from an English language editor, paragraphs 1 and 2 on page 10 for example. This last point is stylistic rather than a reflection on the accuracy or quality of the paper, but I think could help the overall reading of the paper.</p> <p>Overall I feel that this is a good paper, and is suitable for publication with some minor revisions.</p>
--	--

<b>REVIEWER</b>	Tine Vestergaard Department of Dermatology and Allergy Centre, Odense University Hospital, Denmark
<b>REVIEW RETURNED</b>	16-Nov-2020

<b>GENERAL COMMENTS</b>	<p>Thank you for an interesting paper on a very important topic. I have the following comments and questions:</p> <p>I am a little concerned, that the search strategy may not have been sufficient, since three times as many papers were identified through references. In this case, relevant articles may have been missed.</p> <p>Outcomes are not clearly defined. It is mentioned that “results from studies on educational programs on melanoma diagnosis for GPs presented with a wide range of study designs and heterogeneous outcome measures, which made it impossible to formally assess the quality of these studies and to conduct a meta-analysis.” This is fair enough; however, some sort of outcomes for this scoping review should be formulated, e.g. to examine educational</p>
-------------------------	--

	<p>programs for GPs regarding content, teaching method and duration.</p> <p>P8, l. 14-21: This does not add up 11+19+3+3. Is it 36 or 37 studies?</p> <p>In the objectives, training programs on clinical, dermoscopic or combined diagnostics is mentioned (three categories). However, in the results and table 4 only clinical and dermoscopic +/- clinical is reported (two categories). This should be consistent throughout the article, or the reason for combining the two dermoscopy groups explained.</p> <p>P12, l.21: (+-2.5h) How should this be understood?</p> <p>Figure 1: Change pubmed to MEDLINE</p> <p>Please check the English language throughout the article.</p>
--	--

### VERSION 1 – AUTHOR RESPONSE

Responses (answers in blue) to Dr. Owain Jones (reviewer 2):

1. In the abstract, it states that the databases were searched for relevant articles between February and May 2020 - this may be confusing and it may be better to clarify in the abstract that the search was conducted from 1995 to May 2020.  
I clarified the research dates in the abstract (page 2 line 26) and in the study selection paragraph of the material and methods section (page 5 line 54).
2. There is no mention of the methodology that you follow for the scoping review. The framework developed by Arksey and O'Malley (2005) and subsequently refined by Levac (2010) is often cited.  
Page 5 lines 17-21: Indeed, we used the methodology by Arskey and O'Malley to perform the scoping review. The appropriate references have been added to the material and method section.
3. There is also no mention of a Risk of Bias assessment of the included studies. This is not compulsory in scoping reviews, but is sometimes performed. Did you assess the papers for Bias in any way?  
Page 5 lines 57-60: The studies were not assessed for any bias as the 2018 PRISMA-ScR guidelines indicate that the risk of bias assessment is not applicable to scoping reviews.
4. You signify in table 4 whether a result was statistically significant or not, did all of the studies included use the same definition of statistical significance, or was there variability between different studies?  
Caption of Table 4: There was only little variability in the statistical significance defined in the studies. A P-value < 0.05 was considered significant in all studies, except for Grimaldi *et al.* (P<0.001). It should be noted that Shariff *et al.* and Peuvrel *et al.* only provided descriptive statistics of their study results.

5. Very few of the studies used the 7-point checklist as their clinical algorithm, this is slightly surprising as in the UK this is probably the most widely used diagnostic algorithm. However, the studies you have identified are international, and this result may simply reflect that it is not so widely used around the world.  
Of the 31 educational programs, only four trainings were designed and/or evaluated in the UK (Gulati *et al.*; Shariff *et al.*; Harris *et al.*; and Bedlow *et al.*). Of these training programs, only Harris *et al.* specified that they used the 7-point checklist combined with the ABCD rule as the clinical algorithm to teach melanoma recognition. Indeed, the results show that the ABCD(E) rule is more widely used by GPs in Western countries than the 7-point-checklist for clinical assessment of pigmented skin lesions.
  
6. You don't comment on the relative validity of the various assessment methods employed by the included studies. For example, is confidence or number of total-body skin examinations performed a good measure of diagnostic ability, and thus a useful measure with which to evaluate a training program?  
Page 10 lines 34-52: We have added comments on the validity of some assessment methods in the third paragraph of the "Types of educational programs with positive long-term outcomes" section in the discussion.
  
7. A large proportion of papers identified were from other sources according to the PRISMA diagram. It may be worth commenting on this in the manuscript. Was there a reason for this? Were many of the additional studies included, or were the included studies mostly from the original literature searches?  
Figure 1: Most included papers were from the original literature search (29 papers out of 45) including 21 studies and the five systemic reviews about melanoma diagnostic trainings for GPs. The additional studies (15), included from the reference lists of preselected papers, were mostly studies discussed in the five systemic reviews. We misunderstood the definition of papers identified through references. In Figure 1, the number of items identified in the reference list included the total number of references found in the reference lists of the included article. I corrected this and the new number only takes into account relevant articles for our review.
  
8. **Discussion:** I find the discussion slightly confusing. I think there are a couple of reasons for this. There are elements of the discussion that could be included in the results section, paragraph 2 for example and some of the information on page 11. I also think the structure could be clarified and the writing more concise. There are also some sentences which might benefit from input from an English language editor, paragraphs 1 and 2 on page 10 for example. This last point is stylistic rather than a reflection on the accuracy or quality of the paper, but I think could help the overall reading of the paper.  
We revised the discussion. Some of the content were added to the results section (Page 8 lines 7-14) and called upon an English language editor.

Responses (answers in blue) to Dr. Tine Vestergaard (reviewer 3):

1. **Search strategy:** I am a little concerned, that the search strategy may not have been sufficient, since three times as many papers were identified through references. In this case, relevant articles may have been missed.  
Figure 1: The number of items identified in the reference list included the total number of references that were found. This is the result of a misunderstanding on our part. I corrected this and the new number only takes into account relevant articles for our review.

2. Outcomes are not clearly defined. It is mentioned that “results from studies on educational programs on melanoma diagnosis for GPs presented with a wide range of study designs and heterogeneous outcome measures, which made it impossible to formally assess the quality of these studies and to conduct a meta-analysis.” This is fair enough; however, some sort of outcomes for this scoping review should be formulated, e.g. to examine educational programs for GPs regarding content, teaching method and duration.  
Page 4 line 60 - Page 5 line 4: As suggested, the main objectives of this scoping review have been detailed in the last paragraph of the introduction.
3. P8, l. 14-21: This does not add up 11+19+3+3. Is it 36 or 37 studies?  
Page 7 line 19: In total, there are 36 studies and not 37 as mentioned. This has been rectified.
4. In the objectives, training programs on clinical, dermoscopic or combined diagnostics is mentioned (three categories). However, in the results and table 4 only clinical and dermoscopic +/- clinical is reported (two categories). This should be consistent throughout the article, or the reason for combining the two dermoscopy groups explained.  
Abstract page 2 lines 16-18: Introduction page 4 lines 56-59; Material and methods page 5 line 30: To ensure consistency throughout the article, the following two categories of training programs were chosen: clinical diagnosis and dermoscopic +/- clinical diagnosis.
5. P12, l.21: (+-2.5h) How should this be understood?  
Page 10 line 30: +-2.5h means that an average duration of 2.5 hours was observed for the duration of clinical diagnostic training in this review.
6. Figure 1: Change pubmed to MEDLINE  
Figure 1: Pubmed has been changed to MEDLINE.
7. Please check the English language throughout the article.  
As suggested, the manuscript has undergone English language editing.

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Owain Jones University of Cambridge, UK
<b>REVIEW RETURNED</b>	16-Feb-2021

<b>GENERAL COMMENTS</b>	<p>I enjoyed reading your paper. I feel that you have addressed a useful research question, and have discussed it in a clear manner. Overall I have recommended some minor revisions to the paper before publication. I have made some suggestions for you to consider below.</p> <p>Abstract:</p> <ol style="list-style-type: none"> <li>1. I feel there is no need to say ‘the GPs’ it could just be ‘GPs’, on line 4 (Background)</li> <li>2. Did you think about searching the CINAHL database, or including nursing training programmes in the review. When we did a recent systematic review regarding dermoscopy use the systematic review team in our medical library recommended CINAHL for literature on nursing training programmes. In the UK many nurse practitioners will regularly see skin lesions in primary care.</li> </ol>
-------------------------	--

	<p><b>Main text:</b></p> <ul style="list-style-type: none"> <li>• Well done for using the PRISMA-ScR checklist. This is often missed in my experience</li> </ul> <p><b>Introduction:</b></p> <ul style="list-style-type: none"> <li>• I completely agree with your point about diagnostic skills needing to be sustainable, and that there is a lack of research into this.</li> </ul> <p><b>Materials and Method:</b></p> <ol style="list-style-type: none"> <li>1. The first sentence “To carry out this literature review...the most appropriate”. I feel you need to specify what is most appropriate, maybe it is the most appropriate “research method”.</li> <li>2. The second sentence “Indeed, the results...heterogeneous outcomes,” doesn’t quite make sense to me. Do you mean that the studies found in this review were heterogeneous, or the studies found in previous reviews. From what is written, my interpretation is that you mean: “The studies we identified, which provided evidence on the efficacy of education programs in melanoma diagnosis for GPs, demonstrated a wide range of study designs and heterogeneous outcome measures”. But I think you could describe this more clearly</li> <li>3. Further to point 2 above. If you were unable to perform a meta-analysis, it would be good practice to explain here what type of analysis you did do (e.g. narrative synthesis).</li> <li>4. Eligibility criteria: why exclude studies that included training programs for HCPs other than GPs? (and see my previous point about other data sources that could be searched if you included other HCPs)</li> <li>5. Data sources and study selection: I like the way you have described your search criteria. It is very clear. Was the study selection independent? It may be worth specifying this if it was.</li> </ol> <p><b>Results:</b></p> <ol style="list-style-type: none"> <li>1. I find it surprising that many of the training programmes didn’t teach recognition algorithms to aid melanoma recognition.</li> <li>2. GP’s confidence in their own ability does not equal diagnostic competence. I see you allude to this in your discussion. You could consider making this point more strongly.</li> </ol> <p><b>Discussion:</b></p> <ul style="list-style-type: none"> <li>• Given the aims of your study, you have not commented on the relative benefits of different refresher training methods and frequency of refresher training as much as I expected. Although I note you comment on the teaching of clinical and dermoscopic examination seems to produce more sustainable improvements in diagnostic skills. Was there a training modality, or a frequency of refresher training, that the data suggested was better for producing sustainable improvements in GPs diagnostic skills? I recognise that study designs and outcome measures used in the identified studies were very heterogeneous, but if that is limiting your ability to draw stronger conclusions from the data then I would consider stating it more definitively in your discussion and conclusion.</li> </ul> <p><b>Table 4:</b></p> <ul style="list-style-type: none"> <li>• Section A and C uses B/M lesion ratio, section B uses M/B ratio. Is there any reason these are different?</li> </ul>
--	--

<b>REVIEWER</b>	Tine Vestergaard Department of Dermatology and Allergy Centre
-----------------	--

	Odense University Hospital Denmark
<b>REVIEW RETURNED</b>	03-Feb-2021

<b>GENERAL COMMENTS</b>	Thank you for this revised manuscript. I think it is well presented and explores an important area of melanoma diagnostics.
-------------------------	---

### VERSION 2 – AUTHOR RESPONSE

Response to Dr. Tine Vestergaard (reviewer 3):

Thank you for your feedback and contribution to the improvement of our manuscript.

Responses (answers in blue) to Dr. Owain Jones (reviewer 2):

First of all, I want to thank you, on behalf of all the authors, for your careful reviewing of our manuscript. Your comments have certainly contributed to improve the quality of our review.

#### Abstract

1. I feel there is no need to say ‘the GPs’ it could just be ‘GPs’, on line 4 (Background).

Line 11 page 2: I corrected this in the revised manuscript.

2. Did you think about searching the CINAHL database, or including nursing training programmes in the review? When we did a recent systematic review regarding dermoscopy use the systematic review team in our medical library recommended CINAHL for literature on nursing training programmes. In the UK many nurse practitioners will regularly see skin lesions in primary care.

We did not search the CINAHL database or include nursing training programmes in our literature review because we decided to focus on GPs (first paragraph of the introduction). The reason for this is that this literature review is part of my thesis in Medical Sciences which aims to improve early melanoma detection by general practitioners in Belgium.

#### Main text

3. Well done for using the PRISMA-ScR checklist. This is often missed in my experience. Thank you! :-)

#### Introduction

4. I completely agree with your point about diagnostic skills needing to be sustainable, and that there is a lack of research into this.

I am currently working on this and hope to launch a study very soon to assess whether short refresher training sessions at regular intervals are able to prevent the loss of diagnostic skills over time.

## Materials and Method

5. The first sentence “To carry out this literature review...the most appropriate”. I feel you need to specify what is most appropriate, maybe it is the most appropriate “research method”.

Line 12 page 5: I clarified that the scoping review was the most appropriate research method.

6. The second sentence “Indeed, the results...heterogeneous outcomes”, doesn’t quite make sense to me. Do you mean that the studies found in this review were heterogeneous, or the studies found in previous reviews? From what is written, my interpretation is that you mean: “The studies we identified, which provided evidence on the efficacy of education programs in melanoma diagnosis for GPs, demonstrated a wide range of study designs and heterogeneous outcome measures. But I think you could describe this more clearly.

Lines 13-20 page 5: I clarified the sentence using your words which sum up the general idea behind it very well.

7. Further to point 2 above. If you were unable to perform a meta-analysis, it would be good practice to explain here what type of analysis you did do (e.g. narrative synthesis).

Line 21 page 5: I added the type of analysis (narrative synthesis) we performed in the text.

8. Eligibility criteria: why exclude studies that included training programs for HCPs other than GPs? (and see my previous point about other data sources that could be searched if you included other HCPs).

See point 2 for the explanation.

9. Data sources and study selection: I like the way you described your search criteria. It is very clear. Was the study selection independent? It may be worth specifying this if it was.

Line 56 page 5: The study selection was carried out independently by the three authors and specified this in the study selection paragraph.

## Results

10. I find it surprising that many of the training programmes didn’t teach recognition algorithms to aid melanoma recognition.

I only included information on the use of a clinical or dermoscopic algorithm to teach melanoma diagnosis when it was mentioned in the scientific article. I also found this low use of algorithms for teaching melanoma recognition surprising. The use of algorithms may have been more systematic, but the authors did not mention it in their study/educational plan.

11. GP’s confidence in their own ability does not equal diagnostic competence. I see you allude to this in your discussion. You could consider making this point more strongly.

Lines 50-57 page 12: I emphasized this statement in the discussion.

## Discussion



12. Given the aims of your study, you have not commented on the relative benefits of different refresher training methods and frequency of refresher training as much as I expected. Although I note you comment on the teaching of clinical and dermoscopic examination seems to produce more sustainable improvements in diagnostic skills. Was there a training modality, or a frequency of refresher training that the data suggested was better for producing sustainable improvements in GPs diagnostic skills? I recognize that study designs and outcome measures used in the identified studies were very heterogeneous, but if that is limiting your ability to draw stronger conclusions from the data then I would consider stating it more definitively in your discussion and conclusion.

Lines 19-24 page 14 and Lines 10-15 page 15: I stated in the discussion and conclusion that there were no strong conclusions possible on the best training modality to be made due to the variability of study designs, refresher training material provided and heterogeneity of outcome measures.

Table 4

13. Section A and C uses B/M lesion ratio, Section B uses M/B ratio. Is there any reason these are different?

There is no difference. I corrected this in Table 4 pages 28-30.

**VERSION 3 – REVIEW**

<b>REVIEWER</b>	Owain Jones University of Cambridge, UK
<b>REVIEW RETURNED</b>	08-Mar-2021
<b>GENERAL COMMENTS</b>	many thanks for re-submitting this paper, I enjoyed reading it again. I think that the changes you have made clarify your approach and conclusions and add to the paper. i have no further changes to recommend at this stage. I have recommended the paper be accepted for publication.