









# BMJ Open Improving Skin cancer Management with ARTificial Intelligence (SMARTI): protocol for a preintervention/postintervention trial of an artificial intelligence system used as a diagnostic aid for skin cancer management in a specialist dermatology setting

Claire Felmingham <sup>1,2</sup> Samantha MacNamara,<sup>1</sup> William Cranwell <sup>2</sup> Narelle Williams,<sup>3</sup> Miki Wada <sup>1</sup> Nikki R Adler <sup>1</sup> Zongyuan Ge <sup>4,5</sup> Alastair Sharfe,<sup>6</sup> Adrian Bowling,<sup>6</sup> Martin Haskett <sup>7</sup> Rory Wolfe <sup>1</sup> Victoria Mar <sup>1,2</sup>

**To cite:** Felmingham C, MacNamara S, Cranwell W, *et al.* Improving Skin cancer Management with ARTificial Intelligence (SMARTI): protocol for a preintervention/postintervention trial of an artificial intelligence system used as a diagnostic aid for skin cancer management in a specialist dermatology setting. *BMJ Open* 2022;**12**:e050203. doi:10.1136/bmjopen-2021-050203

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-050203>).

Received 28 February 2021  
Accepted 13 December 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Claire Felmingham;  
[clairefelmingham@gmail.com](mailto:clairefelmingham@gmail.com)

## ABSTRACT

**Introduction** Convolutional neural networks (CNNs) can diagnose skin cancers with impressive accuracy in experimental settings, however, their performance in the real-world clinical setting, including comparison to teledermatology services, has not been validated in prospective clinical studies.

**Methods and analysis** Participants will be recruited from dermatology clinics at the Alfred Hospital and Skin Health Institute, Melbourne. Skin lesions will be imaged using a proprietary dermoscopic camera. The artificial intelligence (AI) algorithm, a CNN developed by MoleMap Ltd and Monash eResearch, classifies lesions as benign, malignant or uncertain. This is a preintervention/postintervention study. In the preintervention period, treating doctors are blinded to AI lesion assessment. In the postintervention period, treating doctors review the AI lesion assessment in real time, and have the opportunity to then change their diagnosis and management. Any skin lesions of concern and at least two benign lesions will be selected for imaging. Each participant's lesions will be examined by a registrar, the treating consultant dermatologist and later by a teledermatologist. At the conclusion of the preintervention period, the safety of the AI algorithm will be evaluated in a primary analysis by measuring its sensitivity, specificity and agreement with histopathology where available, or the treating consultant dermatologists' classification. At trial completion, AI classifications will be compared with those of the teledermatologist, registrar, treating dermatologist and histopathology. The impact of the AI algorithm on diagnostic and management decisions will be evaluated by: (1) comparing the initial management decision of the registrar with their AI-assisted decision and (2) comparing the benign to malignant ratio (for lesions biopsied) between the preintervention and postintervention periods.

## Strengths and limitations of this study

- The first prospective clinical trial to evaluate the safety and performance of an artificial intelligence (AI) diagnostic aid for skin cancer detection and management in the real-world clinical setting.
- Participants are recruited on a consecutive basis from routine attendance at melanoma and skin cancer assessment clinics, forming a representative sample of patients and lesion phenotypes from which to evaluate AI algorithm performance.
- AI performance will be compared with teledermatologists' assessment, as well as to face-to-face assessors of varying clinical experience (registrars and consultant dermatologists), and with histopathology results for biopsied lesions.
- Longitudinal follow-up is not undertaken for lesions labelled 'benign' and not actively 'monitored', hence the ultimate malignancy status of these lesions will not be evaluated in this study.
- Inherent differences in application of AI in the specialist setting may limit generalisability of study findings (regarding AI utility) to primary care settings, necessitating further research to establish feasibility for broader clinical implementation.

**Ethics and dissemination** Human Research Ethics Committee (HREC) approval received from the Alfred Hospital Ethics Committee on 14 February 2019 (HREC/48865/Alfred-2018). Findings from this study will be disseminated through peer-reviewed publications, non-peer reviewed media and conferences.

**Trial registration number** NCT04040114.

## INTRODUCTION

Skin cancer, including melanoma and keratinocyte carcinoma, is the most common type of cancer in Caucasian populations, and its incidence is increasing worldwide.<sup>1-3</sup> The incidence of keratinocyte carcinoma is difficult to establish precisely due to a lack of nationwide cancer registry data, although Australia is thought to have the highest incidence worldwide, with over 1000 cases per 100 000 person-years.<sup>4</sup> Similarly, Australia has one of the highest incidence rates of melanoma in the world, with almost 14 000 Australians diagnosed with invasive and more than 20 000 with in-situ melanoma each year.<sup>5</sup>

In Australia there is a shortage of dermatology services in rural and remote areas, where there are consequently long wait times to see a dermatologist. Travel to urban centres can be logistically challenging and expensive for patients. The MoleMap model of care involves total body and dermoscopic imaging by a melanographer. Images are sent to a teledermatologist for reporting. If a lesion is suspicious for malignancy, or if there is diagnostic uncertainty, a recommendation is made to monitor or biopsy the lesion and the patient is advised to consult their doctor. This teledermatology model is particularly useful for people living in areas poorly serviced by dermatologists.<sup>6</sup> It is, however, labour intensive, and it is hoped that artificial intelligence (AI) may reduce workload for teledermatologists in the future.

Melanoma is the third most commonly diagnosed invasive cancer irrespective of gender and is responsible for over 1600 deaths in Australia each year.<sup>5</sup> Early diagnosis of skin cancer reduces morbidity and, in the case of melanoma, is associated with significantly improved survival.<sup>3 7</sup> More accurate and timely skin cancer diagnosis and management could be brought about by the use of new AI-based diagnostic aids.<sup>8-10</sup>

A subset of AI is machine learning. Machine learning refers to the ability of a computer system to write its own programming for a task, and to automatically learn and improve through training data. Deep learning is a branch of machine learning which is becoming increasingly utilised in medicine.<sup>11</sup> Convolutional neural networks (CNNs) are a class of artificial neural networks that are most often used to analyse visual imagery through deep learning. They are especially effective at automated image recognition.

CNNs have been tested with the task of diagnosing skin cancers in multiple studies, and have displayed impressive accuracy equal or superior to that of the dermatologists with whom they have been compared.<sup>12-21</sup> However, these studies have thus far been undertaken in experimental (in silico) settings, and the use of AI as a diagnostic aid has not been adequately evaluated in the real-world clinical setting and in the hands of clinician end-users.<sup>9 22</sup>

AI algorithms should be tested with datasets separate to those with which they are trained, in order to avoid over-fitting or prior dataset bias, which can lead to over-estimation of an algorithm's accuracy.<sup>23 24</sup> In particular, AI algorithms should be tested on the end-target patients

or lesions to ensure their reliability and safety in their intended setting.

Furthermore, in the real-world, dermatologists have additional clinical information (eg, patient demographics and skin cancer history), which improves their diagnostic accuracy.<sup>25</sup> Previous studies comparing AI and dermatologist diagnostic accuracy without provision of this clinical information have therefore disadvantaged dermatologists.

Additionally, these experimental studies positing AI and dermatologists as opponents have been unable to assess the impact of AI algorithms, when used by clinicians, on clinicians' diagnoses and management decisions.

There is a need for prospective clinical trials to validate performance and ensure generalisability of the algorithms, and to evaluate the safety, utility and feasibility of implementing an AI diagnostic aid for skin cancer detection in the clinical setting.<sup>9 12 13 26</sup>

This validation study will evaluate the utility of AI as a diagnostic aid for skin cancer detection and management in the specialist dermatology setting, prior to a larger trial of the intervention in the primary care setting.

If this diagnostic aid for skin cancer management is proven safe, consistent and reliable in a specialist setting, and comparable to a teledermatologist diagnostic assessment, AI-assistance may be appropriate for use in specialist clinics including teledermatology-based services. Further research will be required to determine safety in a primary care setting prior to more widespread implementation, because there will be inherent differences in disease prevalence and clinician experience in this setting when compared with a specialist dermatology setting.

## Objectives

### Primary objective

Assess accuracy of the AI diagnostic aid compared with teledermatologist skin lesion assessment.

### Secondary objectives

- ▶ Evaluate the impact of the AI device when used as a diagnostic aid on the appropriateness of skin cancer management decisions.
- ▶ Evaluate the accuracy and safety of the AI device when used as a diagnostic aid for skin cancer detection in specialist clinics.
- ▶ Assess the feasibility of implementing the AI device as a diagnostic aid for skin cancer detection and management in specialist settings, including teledermatology services.

## METHODS AND ANALYSIS

### Study design and setting

A preintervention/postintervention study of an AI diagnostic aid for skin cancer detection and management.

Participants will be recruited between October 2019 and May 2021 from the patient population attending specialist dermatology and melanoma clinics at two Australian tertiary centres: Skin Health Institute and the

Alfred Hospital in Melbourne, Australia. Participants attending these clinics have a suspected or confirmed diagnosis of skin cancer, or are attending for routine skin surveillance.

Testing the algorithm in specialist dermatology settings allows for comparison of AI lesion classifications with the classifications of both experts (consultant dermatologists) and less-expert clinicians (dermatology registrars). The impact of the AI on less-expert (dermatology registrar) classification and management decisions can be assessed using the expert (consultant dermatologist's) management decision and histopathology as the reference standard. Having established this knowledge, the AI algorithm could subsequently be applied and studied in a primary care setting more safely.

### Participant and public involvement

The study protocol is endorsed by the Melanoma and Skin Cancer Trials Ltd (MASC Trials), a registered not-for-profit Australian and New Zealand's Cancer Collaborative Trials Group member and affiliate of Monash University. All MASC Trials endorsed protocols are subject to review by consumer group representatives, including members of the Australian Melanoma Consumer Alliance.

### Eligibility criteria

Patients aged 18 or over, who are able to provide written informed consent, with at least one skin lesion of concern (to either the patient or treating doctor, excluding acral or scalp lesions), and are willing to have multiple lesions imaged are eligible to participate.

### Recruitment

Willing patients who meet eligibility criteria are provided with a copy of the Participant Information and Consent Form (PICF) and guided through informed consent by their treating dermatology registrar during their clinic

consultation. Participants are recruited on a consecutive basis via convenience sampling from routine attendance at specialist clinics.

### Randomisation and blinding

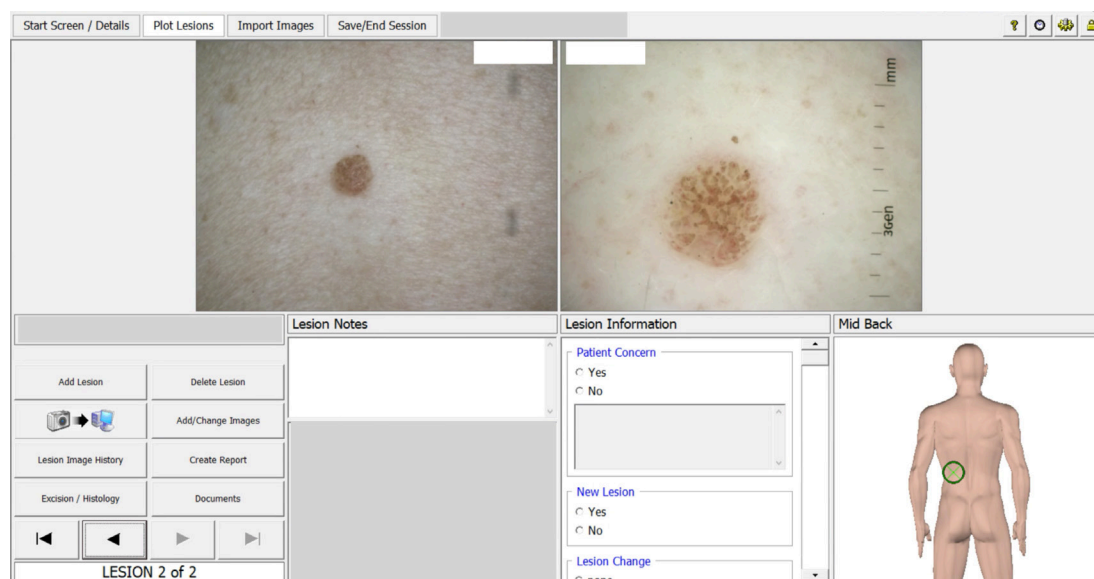
In this preintervention/postintervention study design, the preintervention period will provide an estimate of skin cancer management parameters as a comparator (control) for assessing the impact of AI in the postintervention period. Participants are recruited on a consecutive basis during each of the preintervention and postintervention periods; there is no randomisation. Data are collected on participant risk factors and potentially relevant confounders to be considered during analysis.

In the preintervention period, treating doctors remain blinded to each other's lesion assessment and are unexposed to the AI assessment. Teledermatologists are blinded to the treating doctors' diagnoses and management plans, and to the AI assessment.

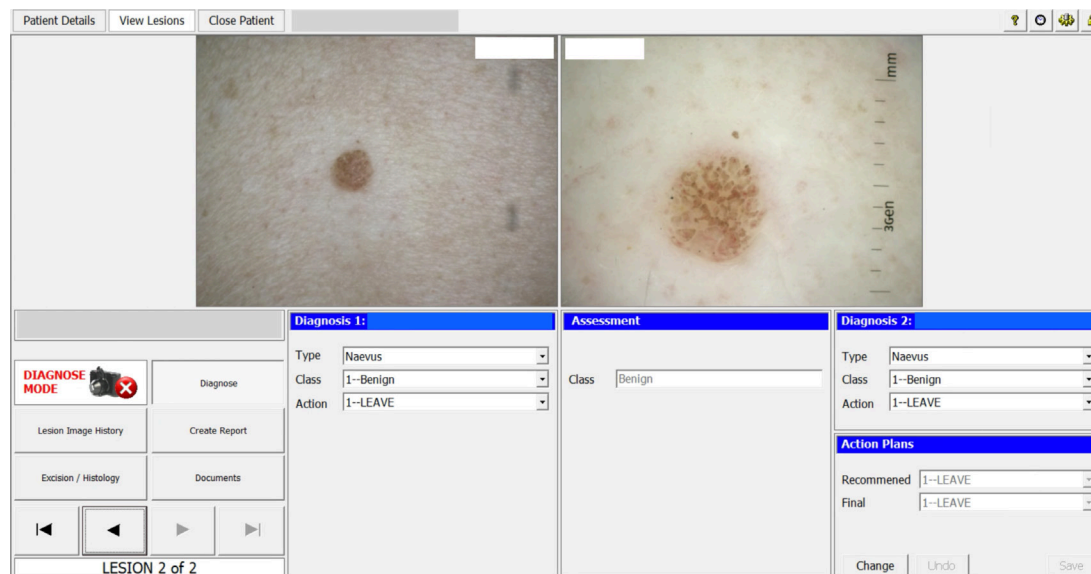
In the postintervention period, treating doctors record their initial diagnosis and management plan decision, and are then exposed to the AI assessment prior to recording a final AI-assisted diagnosis and management plan. The teledermatologists remain blinded to the treating doctors' diagnoses and management plans, and to the AI assessment.

### Description of the intervention: the Skin cancer Management with ARTificial Intelligence AI system

The investigational device includes a proprietary MoleMap Ltd camera capable of taking dermoscopic and macroscopic images and uploading them to an adjacent conventional computer, and the AI software that performs lesion assessments. The computer displays the participant's avatar and lesion images, along with diagnostic and management plan options from which the doctor



**Figure 1** The Skin cancer Management with ARTificial Intelligence computer display: participant avatar indicating the lesion location.



**Figure 2** The Skin cancer Management with ARTificial Intelligence computer display: clinician diagnosis and management plan entry, where: ‘Diagnosis 1’ is the clinician’s initial assessment; ‘Assessment’ is the artificial intelligence (AI) algorithm’s classification; ‘Diagnosis 2’ is the clinician’s AI-assisted assessment and ‘Action Plans’ detail the recommended and final agreed-upon plan.

chooses (figures 1 and 2). Prior to the commencement of the study, research and medical staff working in the clinics receive training on use of the camera, uploading of images and use of the computer software for making diagnoses and management plans.

The Skin cancer Management with ARTificial Intelligence (SMARTI) AI system is a CNN trained to classify lesions using a three-point scale: benign, malignant or uncertain. Figures 1 and 2 demonstrate the SMARTI computer displays and participant avatar indicating the lesion location.

In a laboratory setting, when compared with teledermatologist lesion classification, the first version of the CNN demonstrated a sensitivity of 85%, specificity of 78% and area under the receiver operating characteristic curve (AUROC) of 0.91 for detection of melanoma; and a sensitivity of 72%, specificity of 88% and AUROC of 0.89 for distinguishing a ‘cancer’ from a benign lesion in a binary decision task. These results are comparable to those in pre-existing literature.<sup>12–14</sup> The AUROC is a statistical measure used to assess the discrimination ability of a diagnostic test when there is a dichotomous outcome.<sup>27</sup> An AUROC of 1.00 would mean that the test can discriminate perfectly between the two outcomes. The algorithm was tested with different images to those with which it was trained, however, they were derived from the same dataset of images from MoleMap Ltd. Both macroscopic and dermoscopic images were used to train the algorithm.

The CNN has since been updated to improve its sensitivity and specificity. The algorithm used in the postintervention period will be the algorithm which classifies the lesions imaged during the preintervention period with the greatest accuracy, as assessed by the interim quality assurance analysis.

### Preintervention period

In the preintervention period, lesion assessments made by the AI algorithm are not visible to the treating doctors and therefore do not contribute to diagnostic or management decisions applicable to each lesion.

Participants receive standard of care according to Australian Guidelines,<sup>28 29</sup> including a full skin examination. The participant is first examined by a registrar who selects all skin lesions of concern for imaging, along with two or more non-suspicious lesions. These randomly selected non-suspicious lesions are included to enable analysis of the AI algorithm’s specificity.

Acral and scalp lesions are excluded as these are inherently difficult areas to image, affecting reliability of diagnostic assessment. If approved for use, the algorithm would therefore not be appropriate to use for assessment of lesions at these sites in practice (unless further studies were undertaken) and this would need to be made clear to clinicians.

Macroscopic and polarised dermoscopic images are obtained for each lesion, and are uploaded to an electronic Case Report Form (eCRF) containing the participant’s unique numerical study identifier, with the location of each lesion recorded on a digital avatar. The registrar records their initial favoured diagnosis and management plan for each lesion in the eCRF. Once entered, the diagnostic classification and management plan is locked and cannot be altered.

The treating consultant dermatologist then assesses the participant, recording their favoured diagnosis and management plan for each lesion in the eCRF. If the consultant identifies additional lesions of concern, these are imaged and uploaded to the eCRF and are assessed by the consultant only.

The participant receives recommended management advice from the consultant dermatologist for each lesion, and the final patient-agreed management plan is recorded in the eCRF.

All lesion images are reviewed remotely by one of three experienced teledermatologists. The teledermatologist records their favoured diagnosis and management plan in the eCRF for each lesion. This information is not visible to the treating doctors.

At the conclusion of the preintervention period, the AI algorithm will be applied to generate assessment of all lesions for an interim quality assurance analysis to evaluate safety of the AI algorithm prior to its use in the postintervention period. The algorithm's sensitivity, specificity and agreement (using Kappa statistics) will be calculated, using histopathology as gold standard for biopsied lesions, and treating dermatologists' classifications as gold standard for lesions which are not biopsied to ensure acceptable accuracy prior to proceeding to the intervention phase. That is, whether the algorithm performs with a similar accuracy to the laboratory setting (sensitivity of 72%, specificity of 88%); and with a similar accuracy to that of other AI algorithms which have been shown to classify skin cancer with a sensitivity (ranging 76%–96.3%) and specificity (ranging 53.5%–92%) equal or superior to that of dermatologists.<sup>30</sup> Images collected during the preintervention period will not be used for algorithm retraining.

### Postintervention period

Following the same procedure described above for the preintervention period, participants will be examined by the registrar. Lesions of concern and non-suspicious lesions will be selected, photographed and uploaded to the eCRF. The registrar will record their initial favoured diagnosis and management plan for each lesion and will then submit the images to be analysed by the AI algorithm. The AI assessment will be visible to the registrar in the form of a benign, malignant or uncertain classification for each lesion. On review of the AI assessment, if they choose to, the registrar can update their diagnosis and management plan for each lesion, which will be recorded as an additional AI-assisted diagnosis and management plan in the eCRF.

The consultant dermatologist will then assess the participant and record their favoured diagnosis and management plan for each lesion in the eCRF. The consultant dermatologist will also submit the same images to be analysed by the AI algorithm. The AI assessment will then become visible to the consultant. On review of the AI assessment, if they choose to, the consultant dermatologist may update their diagnosis and management plan for each lesion, which will be recorded as an additional AI-assisted diagnosis and management plan in the eCRF.

The participant will then receive recommended management advice from the consultant dermatologist, which will be recorded on the eCRF. The final plan agreed on between the participant and treating doctors

will be recorded. If either the consultant dermatologist initial or AI-assisted management plan included the decision to biopsy, the biopsy will be undertaken. This is to ensure that standard of care is provided.

The teledermatologists will assess all lesion images remotely following the patient visit and record their favoured diagnosis and management plan in the eCRF, maintaining blinding to the AI assessments. The teledermatologists' diagnoses and plans will not be visible to the treating doctors during either period. The teledermatologists' diagnoses and plans will therefore not influence management decisions in the clinic. Rather, they will be collected for the purpose of comparing and evaluating the accuracy of the AI assessments. All management decisions will ultimately be determined by the treating consultant dermatologist in the clinic (after discussion and agreement with the participant), in line with the standard of care.

### Participant timeline and follow-up procedures

The participant will exit the study after the single study visit is completed if the participant's lesions have all been managed by either: (1) reassurance that no action is required; or (2) non-surgical treatment, such as cryotherapy or imiquimod cream.

If a participant has lesions which have been biopsied or surgically treated, and has no lesions to be monitored, they will exit the study at the time of receipt of the histopathology result.

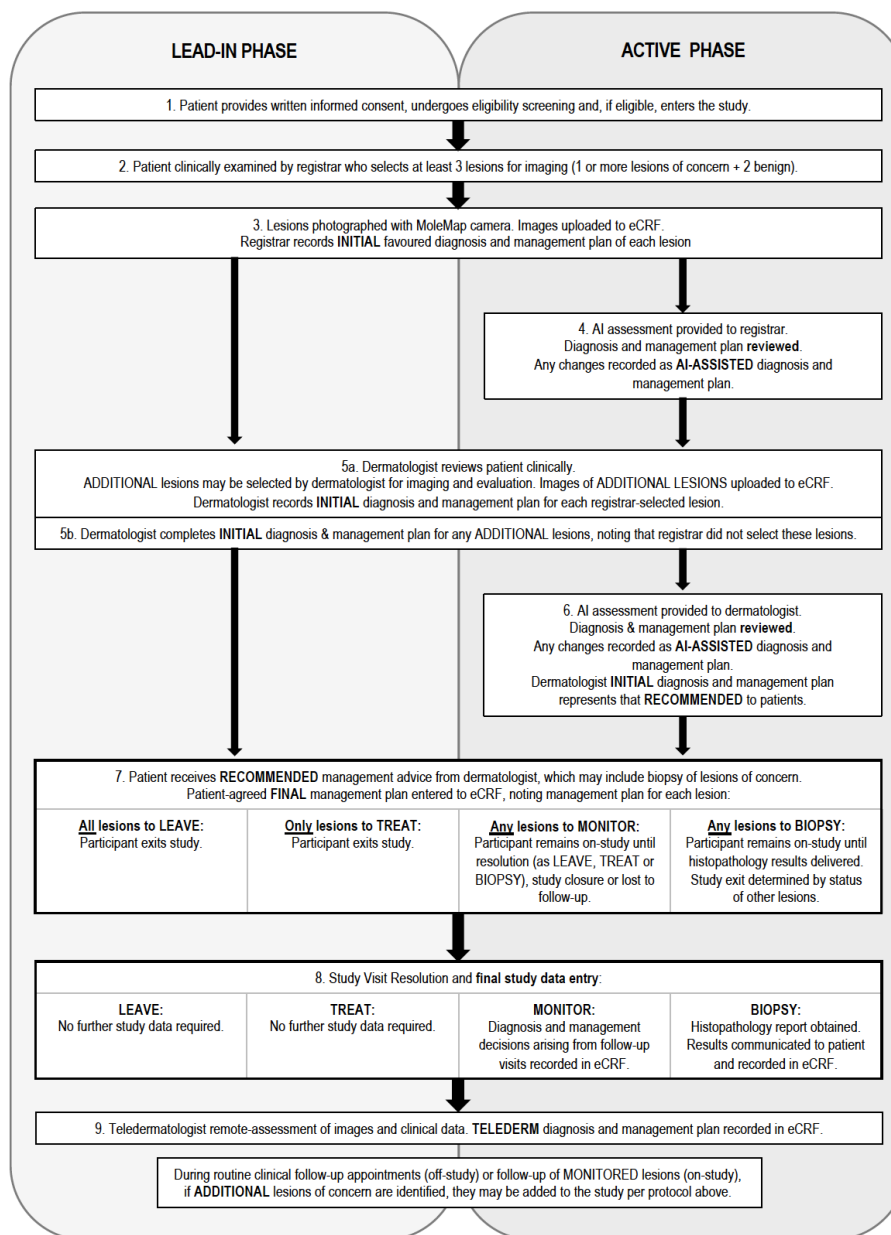
If any lesions are to be monitored, participants will exit the study when either: (1) the monitored lesion(s) progress to biopsy at the 3-month or 6-month follow-up, and the histopathology results are received; (2) the monitored lesion(s) are classified as benign at the 3-month or 6-month follow-up or (3) the participant is lost to follow-up (figure 3).

On study completion, participants will continue to undergo routine surveillance depending on their level of risk and will receive treatment for all lesions as per Australian Guidelines (figure 3).

### Primary outcomes

The primary outcome measure for this study is lesion classification, using a 3-point scale: benign, uncertain or malignant. Definitions and examples for these classifications are given in table 1. The intention of the 'uncertain' classification option for clinicians is to highlight lesions for which a diagnostic tool is most likely to be called on. The aim of the 'uncertain' class for the algorithm is to enable AI categorisation of lesions which are not definitely benign or malignant (eg, severely dysplastic naevi or low grade actinic keratoses), without misleading the clinician.

The primary analysis to evaluate AI performance will compare lesion classification accuracy determined by the AI algorithm to lesion classification accuracy according to teledermatologist assessment, using histopathology as reference standard where available, and the treating



**Figure 3** Participant flow chart. AI, artificial intelligence; eCRF, electronic Case Report Form.

dermatologist's assessment as reference standard where histopathology is not available. The rationale behind this comparison of AI and teledermatologist accuracy is

that: (1) AI and teledermatologists have the same available information (lesion images are available, although they cannot feel the lesion and cannot assess the rest of

**Table 1** Classification definitions

Classification	Definition/situation	Examples
Benign	When the clinician is confident that the lesion is benign	Benign naevus or seborrheic keratosis
Uncertain	When the clinician is unsure and would like a second opinion	Any skin lesion about which the clinician is not confident with regards to its benign/malignant status
Malignant	When the clinician is confident that the lesion is malignant	Melanoma, basal cell carcinoma, squamous cell carcinoma, actinic keratosis*

\*The malignant classification includes premalignant conditions, such as actinic keratosis.

**Table 2** Management decision definitions

Management decision	Definition	Example
Leave	Reassure patient and take no further action.	Benign lesion requiring no further monitoring or medical management.
Manage—monitor	Reassessment of lesion at later time point according to Australian Guidelines.	Patient advised to self-monitor for period of 3 months prior to follow-up monitoring visit.
Manage—biopsy	Partial or complete biopsy of the lesion required to confirm diagnosis.	Shave or excisional biopsy of suspected malignancy.
Treat—elective	Benign or pre-cancerous lesion where treatment is not essential.	Patient requesting cryotherapy of a benign seborrheic keratosis
Treat—essential	Malignancy requiring non-surgical intervention.	Cryotherapy, pharmacotherapy or non-surgical intervention to treat malignancy.

the patient's skin and non-imaged lesions); and (2) an AI diagnostic aid could serve a function similar to a teledermatologist in the future, reducing workload for specialists and improving access to people living in areas poorly serviced by dermatologists.

The primary safety measures include: (1) for all lesions, the proportion of false positive lesion classifications of the AI algorithm that lead to inappropriate registrar management decisions; and (2) for all biopsied lesions, the proportion of false negative lesion classifications of the AI algorithm, using histopathology as the reference standard.

### Secondary outcomes

The secondary outcome is the management decision made by treating doctors, per lesion using the five categories: leave; manage—monitor; manage—biopsy; treat—elective or treat—essential. [Table 2](#) describes management decision outcome categories.

There are seven secondary endpoints: (1) lesion classification of the AI algorithm compared with dermatologist classification; (2) lesion classification of the AI algorithm compared with registrar classification; (3) lesion classification of the AI algorithm compared with histopathology results of any lesions biopsied; (4) initial management decision of the registrar compared with their AI-assisted management decision, using the consultant dermatologist's initial management decision as the reference standard; (5) discordance in the initial and AI-assisted dermatologist management decision during the postintervention period; (6) management decision of the teledermatologist compared with the AI-assisted registrar, using the initial consult dermatologist management decision as the reference standard and (7) the benign to malignant ratio for lesions biopsied in the postintervention period compared with the preintervention period.

### Data collection and management

Participant demographic and risk factor data, including personal and family history of melanoma and keratinocyte carcinoma, ascertained by participant recall will be collected during interview by study staff, recorded directly

to paper CRFs and transcribed to the electronic CRFs at study visit completion.

Pathology reports will be obtained from participants' medical records and relevant histopathology data will be transcribed directly to the eCRF.

Data entered to the custom eCRF platform by study site staff will be automatically synchronised to the electronic database tables built in Microsoft Access. The database will contain only de-identified, re-identifiable data appended to the participant's unique numerical study identifier. The database will be securely stored and backed-up within an approved data-sharing platform with infrastructure enabling at rest encryption using 256-bit Advanced Encryption Standard and Secure Sockets Layer/Transport Layer Security to protect data in transit with 128-bit or higher Advanced Encryption Standard encryption.

### Data monitoring

Routine risk-based monitoring will be undertaken by MASC Research Centre at Monash University for the purpose of source data verification at regular intervals throughout the trial. Data management is also centralised to MASC Research Centre at Monash University, who will be responsible for ongoing surveillance of data quality and integrity.

The Trial Management Committee will conduct regular meetings to review all aspects of study conduct, compliance and progress, in addition to data quality assurance, protocol deviation and monitoring of adverse events and device safety where relevant. Adverse events and protocol violations will be reported to the approving Human Research Ethics Committee (HREC) according to HREC-specific guidelines.

### Statistical methods

#### Sample size

The study aims to recruit 220 participants, providing a minimum of three lesions per participant to the final analysis, thus providing sufficient power to estimate, with reasonable precision, the AI algorithm lesion classification accuracy using teledermatologist assessment as the reference standard. Sample calculations are based on

the assumption that 20% of lesions will be categorised as malignant and 10% will be categorised as uncertain; therefore, approximately 30% of lesions will be categorised as 'not benign' by teledermatologist assessment. If a kappa statistic of 0.8 signifies 'almost perfect' agreement,<sup>31</sup> we will require approximately 220 participants in order to achieve a 95% CI of  $\pm 0.05$  (ie, 95% CI 0.75 to 0.85).

## Statistical analysis

### AI algorithm lesion classification accuracy

The AI algorithm lesion classification accuracy will be compared with relevant physician assessors and histopathology results (for lesions biopsied). Kappa statistics will be used to evaluate agreement between benign/uncertain/malignant lesion classification, with quadratic weights used for kappa calculation. Standard validity indices will be used to evaluate discriminatory ability of the AI algorithm for malignant lesions, including sensitivity, specificity, and positive and negative predictive values.

Performance errors of the CNN will be examined closely. Specifically, all lesions which are classified as benign by the CNN and malignant by the consultant dermatologist or histopathology, and all which are classified as malignant by the CNN and benign by the consultant dermatologist or histopathology, will be reviewed by a dermatologist to determine the nature of these errors.

### Appropriateness of AI-assisted management

The impact of the AI diagnostic aid on appropriateness of the registrar's management decision will be evaluated by measuring the proportion of false positive lesion classifications of the AI algorithm that lead to inappropriate registrar management decisions; comparing the initial registrar management decision with the AI-assisted registrar management decision; and comparing the management decision of the teledermatologist with the AI-assisted registrar decision (all using the dermatologist's initial management decision as the reference standard). The appropriateness of the AI-assisted management will be further assessed by measuring discordance between the initial and AI-assisted management decisions of the dermatologist; and by comparing the benign to malignant ratio (for lesions biopsied) between the preintervention and postintervention periods. Appropriate management of a malignant lesion may vary depending on the diagnosis and the patient's situation. Appropriateness of management decisions will be reviewed for all lesions biopsied or monitored where there is discordance with the dermatologists' initial diagnostic assessment.

Where a lesion's follow-up is unavailable the lesion will be included in analysis according to the treatment path (eg, a lesion that was planned for biopsy will be considered malignant if histopathology is not available). This approach will be supplemented by sensitivity analyses in which the opposite status is assumed (ie, a lesion that was

planned for biopsy will be considered benign if histopathology is not available).

### Interim quality assurance analysis

Following the conclusion of the preintervention period, an interim quality assurance analysis will be conducted to evaluate safety of the AI algorithm to be implemented in the clinical setting during the postintervention period. The safety of the AI algorithm will be evaluated by its agreement with the consultant dermatologists' classification (as benign, malignant or uncertain) for all lesions, and with the histopathology classification for biopsied or excised lesions. Kappa statistics and standard validity indices will be used to assess agreement, evaluating safety of the AI diagnostic aid with reference to gold-standard clinical care provided by consultant dermatologists. The focus of this analysis will be to ensure that the accuracy of the AI algorithm is on par with that of previously produced algorithms.<sup>30</sup>

### Ethics and dissemination

Ethics approval was obtained from the Alfred Hospital Ethics Committee. The protocol has been developed to comply with international standards of Good Clinical Practice (ICH-GCP E6(R2) and TGA Annotation 2016), NHMRC *National Statement* (2018) and *The Code* (2018), and all relevant national, state and local legislative requirements governing data privacy and handling. Study conduct will adhere to principles set out in Declaration of Helsinki 1962 (rev. 2000) and the aforementioned standards.

The findings from this study will be disseminated through peer-reviewed publications, non-peer reviewed media outlets and conferences.

The PICF requests participants indicate whether they consent for their de-identified skin lesion images to be used freely for other research studies. Participants can indicate their consent by completing an additional check box on the PICF.

### Author affiliations

<sup>1</sup>School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

<sup>2</sup>Victorian Melanoma Service, Alfred Health, Melbourne, Victoria, Australia

<sup>3</sup>Melanoma and Skin Cancer Trials Ltd, Melbourne, Victoria, Australia

<sup>4</sup>Monash eResearch Centre, Monash University, Clayton, Victoria, Australia

<sup>5</sup>Department of Electrical and Computer Systems Engineering, Monash University Faculty of Engineering, Clayton, Victoria, Australia

<sup>6</sup>MoleMap Ltd, Auckland, New Zealand

<sup>7</sup>MoleMap Ltd, Melbourne, Victoria, Australia

**Acknowledgements** The authors would like to thank Gabrielle Byars for her valuable contribution.

**Contributors** CF, SM, WC, NW, MW, NA, ZG, AS, AB, MH, RW and VM were all involved in developing the study protocol. VM, RW, MH and ZG worked together on the funding proposal. ZG developed the AI algorithm to be used in the study. RW provided support for the development of the statistical analysis plan. AB and AS provided technical support with the MoleMap computer software. All authors reviewed, edited and approved the final version.

**Funding** The research is funded by the Victorian Medical Research Acceleration Fund, Department of Health and Human Services, State Government of Victoria and MoleMap Ltd.



**Competing interests** VM is supported by an NHMRC Early Career Fellowship. VM reports personal fees from Novartis, personal fees from Bristol-Myers-Squibb, personal fees from Merck, outside the submitted work. MH reports personal fees from MoleMap Ltd, during the conduct of the study; and is a shareholder in MoleMap Ltd. AB reports personal fees from MoleMap Ltd, during the conduct of the study; personal fees from Molemap Ltd, outside the submitted work; and is a shareholder in Molemap Ltd. AS reports personal fees from MoleMap Ltd, during the conduct of the study; personal fees from Molemap Ltd, outside the submitted work. ZG reports personal fees from MoleMap Ltd. NW and SM are former employees of the Cancer Collaborative Trials Group contracted to implement the SMARTI Study—Melanoma and Skin Cancer Trials (MASC Trials) Ltd. CF is supported by a Monash University Research Training Program Scholarship. RW, NA, WC and MW have nothing to disclose. The study is sponsored by Monash University and endorsed by MASC Trials Ltd.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Consent obtained directly from patient(s).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Claire Felmingham <http://orcid.org/0000-0002-3443-8065>

William Cranwell <http://orcid.org/0000-0001-6368-5738>

Miki Wada <http://orcid.org/0000-0002-6337-3619>

Nikki R Adler <http://orcid.org/0000-0002-7972-9050>

Zongyuan Ge <http://orcid.org/0000-0002-5880-8673>

Martin Haskett <http://orcid.org/0000-0002-3357-5826>

Rory Wolfe <http://orcid.org/0000-0002-2126-1045>

Victoria Mar <http://orcid.org/0000-0001-9423-3435>

#### REFERENCES

- Apalla Z, Lallas A, Sotiriou E, *et al*. Epidemiological trends in skin cancer. *Dermatol Pract Concept* 2017;7:1–6.
- Leiter U, Eigentler T, Garbe C. Epidemiology of skin cancer. *Adv Exp Med Biol* 2014;810:120–40.
- Schadendorf D, van Akkooi ACJ, Berking C, *et al*. Melanoma. *Lancet* 2018;392:971–84.
- Perera E, Gnaneswaran N, Staines C, *et al*. Incidence and prevalence of non-melanoma skin cancer in Australia: a systematic review. *Australas J Dermatol* 2015;56:258–67.
- Australian Institute of Health and Welfare. *Cancer in Australia*, 2019. <https://www.aihw.gov.au/getmedia/8c9fcf52-0055-41a0-96d9-f81b0feb98cf/aihw-can-123.pdf.aspx?inline=true>
- Kozera EK, Yang A, Murrell DF. Patient and practitioner satisfaction with tele-dermatology including Australia's Indigenous population: a systematic review of the literature. *Int J Womens Dermatol* 2016;2:70–3.
- Gershenwald JE, Scolyer RA, Hess KR, *et al*. Melanoma staging: evidence-based changes in the American joint Committee on cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017;67:472–92.
- Gilmore SJ. Automated decision support in melanocytic lesion management. *PLoS One* 2018;13:e0203459.
- Tschandl P, Rinner C, Apalla Z, *et al*. Human-computer collaboration for skin cancer recognition. *Nat Med* 2020;26:1229–34.
- Mar VJ, Soyer HP. Artificial intelligence for melanoma diagnosis: how can we deliver on the promise? *Ann Oncol* 2018;29:1625–8.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
- Esteva A, Kuprel B, Novoa RA, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
- Haenssle HA, Fink C, Schneiderbauer R, *et al*. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018;29:1836–42.
- Tschandl P, Rosendahl C, Akay BN, *et al*. Expert-level diagnosis of nonpigmented skin cancer by combined Convolutional neural networks. *JAMA Dermatol* 2019;155:58–65.
- Marchetti MA, Codella NCF, Dusza SW, *et al*. Results of the 2016 international skin imaging collaboration International Symposium on biomedical imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *J Am Acad Dermatol* 2018;78:270–7.
- Fujisawa Y, Otomo Y, Ogata Y, *et al*. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol* 2019;180:373–81.
- Han SS, Kim MS, Lim W, *et al*. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol* 2018;138:1529–38.
- Brinker TJ, Hekler A, Enk AH, *et al*. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer* 2019;119:11–17.
- Brinker TJ, Hekler A, Enk AH, *et al*. A convolutional neural network trained with dermoscopic images performed on PAR with 145 dermatologists in a clinical melanoma image classification task. *Eur J Cancer* 2019;111:148–54.
- Brinker TJ, Hekler A, Enk AH, *et al*. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer* 2019;113:47–54.
- Yu C, Yang S, Kim W, *et al*. Acral melanoma detection using a convolutional neural network for Dermoscopy images. *PLoS One* 2018;13:e0193321.
- Du-Harpur X, Watt FM, Luscombe NM, *et al*. What is AI? applications of artificial intelligence to dermatology. *Br J Dermatol* 2020.
- Navarrete-Dechent C, Dusza SW, Liopyris K, *et al*. Automated Dermatological Diagnosis: Hype or Reality? *J Invest Dermatol* 2018;138:2277–9.
- Tschandl P, Codella N, Akay BN, *et al*. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol* 2019;20:938–47.
- Haenssle HA, Fink C, Toberer F, *et al*. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. *Ann Oncol* 2020;31:137–43.
- Nagendran M, Chen Y, Lovejoy CA, *et al*. Artificial intelligence versus clinicians: systematic review of design, reporting Standards, and claims of deep learning studies. *BMJ* 2020;368:m689.
- Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med* 2013;4:627–35.
- Cancer Council Australia Keratinocyte Cancers Guideline Working Party. *Clinical practice guidelines for keratinocyte cancer*. Sydney: Cancer Council Australia. [https://wiki.cancer.org.au/australia/Guidelines:Keratinocyte\\_carcinoma](https://wiki.cancer.org.au/australia/Guidelines:Keratinocyte_carcinoma)
- Cancer Council Australia Melanoma Guidelines Working Party. *Clinical practice guidelines for the diagnosis and management of melanoma*. Sydney: Cancer Council Australia, 2016. <https://wiki.cancer.org.au/australia/Guidelines:Melanoma>
- Wada M, Ge Z, Gilmore SJ, *et al*. Use of artificial intelligence in skin cancer diagnosis and management. *Med J Aust* 2020;213:256–9.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.