NIHR | National Institute for Health Research

Journals Library

Check for updates

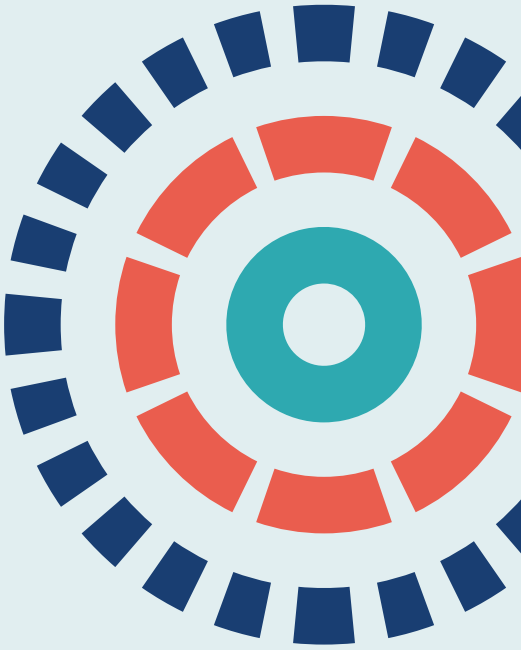
Health Technology Assessment

Volume 25 • Issue 64 • November 2021

ISSN 1366-5278

Optimal surveillance strategies for patients with stage 1 cutaneous melanoma post primary tumour excision: three systematic reviews and an economic model

Luke Vale, Patience Kunonga, Diarmuid Coughlan, Vasileios Kontogiannis, Margaret Astin, Fiona Beyer, Catherine Richmond, Dor Wilson, Dalvir Bajwa, Mehdi Javanbakht, Andrew Bryant, Wanwuri Akor, Dawn Craig, Penny Lovat, Marie Labus, Batoul Nasr, Timothy Cunliffe, Helena Hinde, Mohamed Shawgi, Daniel Saleh, Pam Royle, Paul Steward, Rachel Lucas and Robert Ellis



DOI 10.3310/hta25640

Optimal surveillance strategies for patients with stage 1 cutaneous melanoma post primary tumour excision: three systematic reviews and an economic model

Luke Vale,[1] Patience Kunonga,[1] Diarmuid Coughlan,[1] Vasileios Kontogiannis,[1] Margaret Astin,[1] Fiona Beyer,[1] Catherine Richmond,[1] Dor Wilson,[1] Dalvir Bajwa,[2] Mehdi Javanbakht,[1] Andrew Bryant,[1] Wanwuri Akor,[3] Dawn Craig,[1] Penny Lovat,[4] Marie Labus,[5] Batoul Nasr,[6] Timothy Cunliffe,[7] Helena Hinde,[7] Mohamed Shawgi,[8] Daniel Saleh,[9,10] Pam Royle,[11] Paul Steward,[12] Rachel Lucas[12] and Robert Ellis[4,13*]

[1]Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK
[2]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK
[3]Northumbria Healthcare NHS Foundation Trust, North Shields, UK
[4]Institute of Translation and Clinical Studies, Newcastle University, Newcastle upon Tyne, UK
[5]Business Development and Enterprise, Newcastle University, Newcastle upon Tyne, UK
[6]Dermatological Sciences, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK
[7]Dermatology Department, James Cook University Hospital, Middlesbrough, UK
[8]Radiology Department, James Cook University Hospital, Middlesbrough, UK
[9]Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
[10]Princess Alexandra Hospital Southside Clinical Unit, Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia
[11]Patient representative, ITV Tyne Tees, Gateshead, UK
[12]Patient representative, Dermatology Department, James Cook University Hospital, Middlesbrough, UK
[13]South Tees Hospitals NHS Foundation Trust, Middlesbrough, UK

*Corresponding author

Declared competing interests of authors: Luke Vale was a member of the National Institute for Health Research (NIHR) Clinical Evaluation and Trial Panel from 2015 to 2018. His partner is the Chief Executive Officer (employed by Newcastle University) of a Newcastle University spin-out company called AMLo Biosciences Ltd (Newcastle upon Tyne, UK). She is also a shareholder of the same, and holds patents for biomarkers in this area. AMLo Biosciences Ltd is developing a prognostic biomarker for melanoma. Dawn Craig is a member of the NIHR Health Services and Delivery Research programme's Prioritisation Committee (Researcher-led). Robert Ellis reports personal fees from AbbVie Inc. (Lake Bluff, IL, USA) and AMLo Biosciences Ltd outside the submitted work. Penny Lovat and Marie Labus report personal fees from AMLo Biosciences Ltd, outside the submitted work. In addition, Penny Lovat has a patent family arising from PCT/GB2015/053347 (biomarkers for disease progression in melanoma; assignee: AMLo Biosciences Ltd; inventors: Penny Lovat, Robert Ellis and Marie Labus; issued), patent 1818168.9 (monoclonal antibodies against AMBRA1; assignee: AMLo Biosciences Ltd; inventors: Penny Lovat, Robert Ellis, Ashleigh McConnell and Marie Labus; issued) and patent 118622.1 (monoclonal antibodies against loricrin; assignee: AMLo Biosciences Ltd; inventors: Penny Lovat, Robert Ellis, Ashleigh McConnell and Marie Labus; pending).

Published November 2021
DOI: 10.3310/hta25640

This report should be referenced as follows:

Vale L, Kunonga P, Coughlan D, Kontogiannis V, Astin M, Beyer F, *et al*. Optimal surveillance strategies for patients with stage 1 cutaneous melanoma post primary tumour excision: three systematic reviews and an economic model. *Health Technol Assess* 2021;**25**(64).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.014

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

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

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

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

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

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

HTA programme

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

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

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 16/166/05. The contractual start date was in March 2018. The draft report began editorial review in October 2019 and was accepted for publication in July 2020. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

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

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

NIHR Journals Library Editor-in-Chief

NIHR Journals Library Editors

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

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

Abstract

Optimal surveillance strategies for patients with stage 1 cutaneous melanoma post primary tumour excision: three systematic reviews and an economic model

Luke Vale,[1] Patience Kunonga,[1] Diarmuid Coughlan,[1] Vasileios Kontogiannis,[1] Margaret Astin,[1] Fiona Beyer,[1] Catherine Richmond,[1] Dor Wilson,[1] Dalvir Bajwa,[2] Mehdi Javanbakht,[1] Andrew Bryant,[1] Wanwuri Akor,[3] Dawn Craig,[1] Penny Lovat,[4] Marie Labus,[5] Batoul Nasr,[6] Timothy Cunliffe,[7] Helena Hinde,[7] Mohamed Shawgi,[8] Daniel Saleh,[9,10] Pam Royle,[11] Paul Steward,[12] Rachel Lucas[12] and Robert Ellis[4,13*]

[1]Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK
[2]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK
[3]Northumbria Healthcare NHS Foundation Trust, North Shields, UK
[4]Institute of Translation and Clinical Studies, Newcastle University, Newcastle upon Tyne, UK
[5]Business Development and Enterprise, Newcastle University, Newcastle upon Tyne, UK
[6]Dermatological Sciences, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK
[7]Dermatology Department, James Cook University Hospital, Middlesbrough, UK
[8]Radiology Department, James Cook University Hospital, Middlesbrough, UK
[9]Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
[10]Princess Alexandra Hospital Southside Clinical Unit, Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia
[11]Patient representative, ITV Tyne Tees, Gateshead, UK
[12]Patient representative, Dermatology Department, James Cook University Hospital, Middlesbrough, UK
[13]South Tees Hospitals NHS Foundation Trust, Middlesbrough, UK

*Corresponding author robellis@nhs.net

Background: Malignant melanoma is the fifth most common cancer in the UK, with rates continuing to rise, resulting in considerable burden to patients and the NHS.

Objectives: The objectives were to evaluate the effectiveness and cost-effectiveness of current and alternative follow-up strategies for stage IA and IB melanoma.

Review methods: Three systematic reviews were conducted. (1) The effectiveness of surveillance strategies. Outcomes were detection of new primaries, recurrences, metastases and survival. Risk of bias was assessed using the Cochrane Collaboration's Risk-of-Bias 2.0 tool. (2) Prediction models to stratify by risk of recurrence, metastases and survival. Model performance was assessed by study-reported measures of discrimination (e.g. D-statistic, Harrel's *c*-statistic), calibration (e.g. the Hosmer–Lemeshow 'goodness-of-fit' test) or overall performance (e.g. Brier score, R^2). Risk of bias was assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). (3) Diagnostic test accuracy of fine-needle biopsy and ultrasonography. Outcomes were detection of new primaries, recurrences, metastases and

overall survival. Risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies–2 (QUADAS–2) tool. Review data and data from elsewhere were used to model the cost-effectiveness of alternative surveillance strategies and the value of further research.

Results: (1) The surveillance review included one randomised controlled trial. There was no evidence of a difference in new primary or recurrence detected (risk ratio 0.75, 95% confidence interval 0.43 to 1.31). Risk of bias was considered to be of some concern. Certainty of the evidence was low. (2) Eleven risk prediction models were identified. Discrimination measures were reported for six models, with the area under the operating curve ranging from 0.59 to 0.88. Three models reported calibration measures, with coefficients of ≥ 0.88. Overall performance was reported by two models. In one, the Brier score was slightly better than the American Joint Committee on Cancer scheme score. The other reported an R^2 of 0.47 (95% confidence interval 0.45 to 0.49). All studies were judged to have a high risk of bias. (3) The diagnostic test accuracy review identified two studies. One study considered fine-needle biopsy and the other considered ultrasonography. The sensitivity and specificity for fine-needle biopsy were 0.94 (95% confidence interval 0.90 to 0.97) and 0.95 (95% confidence interval 0.90 to 0.97), respectively. For ultrasonography, sensitivity and specificity were 1.00 (95% confidence interval 0.03 to 1.00) and 0.99 (95% confidence interval 0.96 to 0.99), respectively. For the reference standards and flow and timing domains, the risk of bias was rated as being high for both studies. The cost-effectiveness results suggest that, over a lifetime, less intensive surveillance than recommended by the National Institute for Health and Care Excellence might be worthwhile. There was considerable uncertainty. Improving the diagnostic performance of cancer nurse specialists and introducing a risk prediction tool could be promising. Further research on transition probabilities between different stages of melanoma and on improving diagnostic accuracy would be of most value.

Limitations: Overall, few data of limited quality were available, and these related to earlier versions of the American Joint Committee on Cancer staging. Consequently, there was considerable uncertainty in the economic evaluation.

Conclusions: Despite adoption of rigorous methods, too few data are available to justify changes to the National Institute for Health and Care Excellence recommendations on surveillance. However, alternative strategies warrant further research, specifically on improving estimates of incidence, progression of recurrent disease; diagnostic accuracy and health-related quality of life; developing and evaluating risk stratification tools; and understanding patient preferences.

Study registration: This study is registered as PROSPERO CRD42018086784.

Funding: This project was funded by the National Institute for Health Research Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol 25, No. 64. See the NIHR Journals Library website for further project information.

Contents

List of tables

List of figures

List of boxes

List of supplementary material

Report Supplementary Material 1 Data extraction and risk of bias tools

Supplementary material can be found on the NIHR Journals Library report page (https://doi.org/10.3310/hta25640).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

AGREE II	Appraisal of Guidelines for Research & Evaluation II
AJCC	American Joint Committee on Cancer
AUC	area under the curve
AUROC	area under the receiver operating characteristic
BAD	British Association of Dermatologists
CCG	Clinical Commissioning Group
CEAC	cost-effectiveness acceptability curve
CHARMS	CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies
CI	confidence interval
CNS	clinical nurse specialist
CrI	credible interval
CT	computerised tomography
EPV	event per variable
EQ-5D	EuroQol-5 Dimensions
ESMO	European Society of Medical Oncologists
EVPI	expected value of perfect information
EVPPI	expected value of partial perfect information
FNAC	fine-needle aspiration cytology
FNB	fine-needle biopsy
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HERC	Health Economics Research Centre
HR	hazard ratio
ICER	incremental cost-effectiveness ratio
ILI	isolated limb infusion
ILP	isolated limb perfusion
mCM	modified Connor–Mosimann
MRI	magnetic resonance imaging
NCCN®	National Comprehensive Cancer Network®
NCRAS	National Cancer Registration and Analysis Service
NICE	National Institute for Health and Care Excellence
NMB	net monetary benefit
OR	odds ratio
PDF	Portable Document Format
PET	positron emission tomography
PET-CT	positron emission tomography-computerised tomography
PHE	Public Health England
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST	Prediction model Risk Of Bias ASsessment Tool
PSA	probabilistic sensitivity analysis
QALY	quality-adjusted life-year
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies-2
RCT	randomised controlled trial
RGP	radial growth phase
ROC	receiver operating curve
RR	risk ratio
SAVI	Sheffield Accelerated Value of Information

ScHARRHUD	School of Health and Related Research Health Utilities Database
SD	standard deviation
SE	standard error
SLNB	sentinel lymph node biopsy
SSE	skin self-examination
TA	technology appraisal
TNM	tumour node metastasis
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
UV	ultraviolet
WTP	willingness to pay

Plain English summary

Malignant melanoma is the deadliest of skin cancers; in the UK, > 2500 people die from it every year. Initially, the cancer is removed surgically, which cures it for most people, but, for some, the cancer returns. For this reason, after a melanoma is removed, patients are followed up to see if the melanoma reoccurs or if new melanomas have developed. It is felt that early cancer detection improves the chance of future treatment working. A key question is how best to follow up patients after initial melanoma surgery. This study concentrates on the earliest stage of melanoma (American Joint Committee on Cancer stage I), which accounts for more than 7 out of 10 of all melanoma diagnoses. The study also investigates if new ways of follow-up could be at least as good as current practice and a better use of NHS money.

We systematically reviewed studies comparing different ways of organising follow-up, and then methods to identify those patients at high risk of developing a further melanoma and how good different tests are at detecting this cancer. We then compared different possible follow-up strategies. For each strategy, we considered its impact on quality and length of life, and how well it used NHS resources.

We found little evidence to support a change in how follow-up should be organised currently. There were some ways of organising follow-up that might be better than current care, but further research is needed. We found that new research on whether or not follow-up should be performed by a cancer nurse specialist, rather than a dermatologist or surgeon, would be worthwhile. We also found that more research could be worthwhile on how frequently melanoma recurs and spreads, as well as how accurately a diagnosis of further cancer is made and how to identify those most at risk of further melanoma spread.

Scientific summary

Background

Cutaneous melanoma is a cancer that develops from pigment-producing cells (melanocytes) in the skin, and is one of the deadliest skin cancers. It is aggressive, rapidly disseminates and, until recently, had a median overall survival of between 6 and 10 months once metastasis had occurred. The recent introduction of targeted immunotherapies has improved outcomes, with median overall survival now reaching at least 2 years. Cutaneous melanoma is the fifth most common cancer in the UK, with 17,000 patients diagnosed annually. It is also the UK's leading cause of cancer-related death among people aged 20–35 years.

After removal of the primary tumour, the majority of melanomas are cured. However, up to 30% of all primary melanomas progress to metastatic disease, with extremely poor 5-year survival rates of only 14%. Consequently, there are 2500 melanoma-associated deaths in the UK annually. The total annual cost due to skin cancer to the NHS was £106M–112M in 2008 and is expected to rise to > £180M by 2020.

Primary melanomas are staged according to the American Joint Committee on Cancer staging criteria. These include Breslow depth (the distance into the skin of the tumour invasion) and the presence of ulceration (loss of the epidermis overlying the tumour) to allow disease-risk stratification. In 2017, the eighth edition of American Joint Committee on Cancer staging, but when this study was conducted, most data used the seventh or earlier editions of staging. In the seventh edition of the American Joint Committee on Cancer, stage I tumours are identified as tumours up to 2 mm thick, with no ulceration, or < 1 mm thick, if ulceration is present. American Joint Committee on Cancer stage I disease represents the lowest mortality risk, compared with other stages of disease: up to 14% over 10 years.

Although surgical treatment of primary melanoma is effective, the pace of development is rapid, with the introduction of additional early investigatory techniques (e.g. sentinel lymph node biopsy and various radiological modalities) and advances with the treatment of metastatic disease. A structured, evidence-based model of patient follow-up after initial diagnosis is lacking. Current guidelines for surveillance vary across the world, with most based on anecdotal evidence and expert opinion. The recommendations make the assumption that earlier disease detection results in improved outcome, but often do not consider all elements used in the diagnosis and management of the condition, as well as the potential physical, psychological and financial costs of these surveillance regimens.

With low rates of metastasis, and early physiological stage of development, targeting American Joint Committee on Cancer stage I melanomas for appropriate follow-up strategies could improve, or at least maintain, outcomes at lower costs. Limited evidence suggests that low-risk patients may not need intensive clinician follow-up, as recommended. Conversely, a more appropriately structured follow-up regime for higher-risk patients may allow earlier detection of metastatic disease with associated benefits from earlier treatment.

With the rapid increase in melanoma rates, there is a need to develop a robust, evidence-based model of follow-up care for American Joint Committee on Cancer stage I patients: the majority of people affected by melanoma. The increase in diagnostic accuracy, development of potential prognostic biomarkers, new radiological modalities and the introduction of personalised systemic treatments could transform melanoma care. However, without a robust, evidence-based framework for implementation of such interventions, the potential health and economic benefits for the NHS will not be achieved.

Objectives

The aim of this research was to evaluate the effectiveness and cost-effectiveness of different surveillance strategies for patients with American Joint Committee on Cancer stage I melanoma after surgical excision of the primary cutaneous tumour. The objectives were to:

1. systematically review different strategies for surveillance and follow-up after surgical excision of the primary cutaneous tumour, including their effectiveness and cost-effectiveness
2. systematically review the prognostic performance of risk models used to determine the prognosis and risk stratification of patients with American Joint Committee on Cancer stage I melanoma after surgical excision of the primary cutaneous tumour
3. systematically review the diagnostic performance of tests used in surveillance and follow-up strategies in detecting new primaries, recurrence and metastatic diseases in patients with American Joint Committee on Cancer stage I melanoma after surgical excision of the primary cutaneous tumour
4. develop a decision-analytic model to estimate the effectiveness and cost-effectiveness of the surveillance and follow-up strategies after surgical excision of the primary cutaneous tumour
5. undertake value of information analysis to assess the need for further primary research.

The results of the systematic reviews conducted to meet the first three objectives were used to inform the design and conduct of an economic evaluation based on a decision model, which addressed the fourth and fifth objectives.

Surveillance review

Methods

A systematic review of comparative studies was conducted to identify various surveillance and follow-up strategies after surgical excision of American Joint Committee on Cancer stage I primary cutaneous melanomas in adults, and to assess their relative effectiveness. The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. Ten bibliographic databases, grey literature and guidelines from January 2011 to July 2019 were searched. To identify studies published prior to 2011, we assessed all references of an earlier review on melanoma surveillance by Cromwell *et al.* (Cromwell KD, Ross MI, Xing Y, Gershenwald JE, Royal RE, Lucci A, *et al.* Variability in melanoma post-treatment surveillance practices by country and physician specialty: a systematic review. *Melanoma Res* 2012;**22**:376–85), which searched up to that point. Furthermore, references from the clinical guidelines identified in an earlier component of this project were assessed. Included studies had to compare surveillance strategies (relevant strategies compared included a no active surveillance option). Outcomes included detection of new primary tumours, recurrence, metastases or overall survival.

Results

Searches identified 6205 records. One randomised controlled trial from the USA met the inclusion criteria. This trial evaluated the effect of surveillance using structured skin self-examination. New primaries, recurrences or metastases were detected in 49 out of 258 (19%) patients with stage IA or IB melanoma: 36 out of 203 (18%) in the intervention group and 13 out of 55 (24%) in the control group. The overall risk of bias for the trial was identified as being of some concern. Overall certainty of the evidence was low and future trials would be very likely to influence results.

Prediction model review

Methods

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses and the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling

Studies (CHARMS) guidelines, and assessed the prognostic accuracy of risk prediction models to predict recurrence, new primary tumours and metastases. Searches of 10 databases were conducted, searching from 2000 (when sentinel lymph node biopsy was introduced in melanoma) to July 2019. Model performance measures included discrimination (ability to differentiate between high and low risk), calibration (agreement between observed and predicted risk) and overall performance (combination of the discrimination and calibration measures), assessed by the Brier score (a statistical measure of the accuracy of the measure; a higher score means higher inaccuracy) and R^2 statistic (a statistic describing the percentage of the variance to measure overall model performance).

Results

Searches identified 20,878 records and 11 different risk prediction models. The number of predictors per model ranged from 3 to 11. The most common were age, tumour site, tumour thickness, sex and ulceration. Discrimination was reported in six studies and the area under the operating curve (whereby 0.5 is fail and 1 is perfect) ranged from 0.59 to 0.88. Calibration measures were reported in three studies. One study reported a calibration slope of 0.88 ($p = 0.5$), and another reported concordance correlation coefficients of 0.9 and 0.93 for 5- and 10-year survival rates, both demonstrating high accuracy of the models. Two studies measured the overall performance of the model. One study assessed the Brier score of a new model and showed a slightly better (i.e. lower) Brier score than the American Joint Committee on Cancer scheme. The other study assessed the R^2 statistic and reported it as 0.47 (95% confidence interval 0.45 to 0.49), indicating that the model explains an estimated 47% of the variation. All studies were retrospective, and so were rated as having a high risk of bias; eight studies conducted internal validation using data from their development set.

Diagnostic performance review

Methods

This systematic review explored the diagnostic test accuracy of fine-needle biopsy and ultrasonography to detect new primaries, recurrence and locoregional metastases during follow-up of stage I melanoma. Searches of electronic databases were conducted from inception to July 2019. Data were extracted on study/participant characteristics and index test accuracy statistics. Risk of bias was independently assessed using Quality Assessment of Diagnostic Accuracy Studies-2 for each included study.

A bivariate random-effects meta-analysis model was planned. This approach would have enabled the calculation of summary estimates of sensitivity and specificity across different studies. Owing to paucity of data, a narrative approach was used and estimates of sensitivity and specificity for each study estimated.

Results

Database searches retrieved 2250 records. Two studies assessing different index tests relevant at different stages of diagnosis met the inclusion criteria. One Australian study (Doubrovsky A, Scolyer RA, Murali R, McKenzie PR, Watson GF, Lee CS, *et al.* Diagnostic accuracy of fine needle biopsy for metastatic melanoma and its implications for patient management. *Ann Surg Oncol* 2008;**15**:323–32.) assessed the accuracy of fine-needle biopsy (the study was rated as having a high risk of bias). Data were reported for stage I disease by the number of fine-needle biopsies performed ($n = 323$) in those with stage I melanoma. A German study (Krüger U, Kretschmer L, Thoms KM, Padeken M, Bertsch PH, Schön MP, Zutt M. Lymph node ultrasound during melanoma follow-up significantly improves metastasis detection compared with clinical examination alone: a study on 433 patients. *Melanoma Res* 2011;**21**:457–63.) from assessed ultrasound (the study was rated as having a high risk of bias) and included 669 investigations among individuals with stage I melanoma.

For the study assessing the diagnostic performance of fine-needle biopsy, the sensitivity was 0.93 (95% confidence interval 0.88 to 0.97) and specificity was 0.98 (95% confidence interval 0.95 to 1.00).

Results in the ultrasonography study were reported as the number of paired investigations of both clinical examination and ultrasonography conducted on the same day, with an average of three paired investigations per patient. These data were converted so that the unit of analysis was the participants. Sensitivity was reported as 1.00 (95% confidence interval 0.03 to 1.00) and specificity was 0.99 (95% confidence interval 0.96 to 1.00).

Sensitivity analyses were conducted for both study analyses; results did not change.

Economic evaluation

Methods

A review of cost-effectiveness studies identified 15 possibly relevant studies, but none directly addressed the study question. Therefore, an economic model was developed to assess the cost-effectiveness of alternative surveillance strategies and to estimate the value of information. The model took a lifetime horizon and an NHS and Personal Social Services perspective. A Markov microsimulation model, with monthly cycles, was developed in TreeAge 2019 R1.0 (TreeAge Software, Inc., Williamstown, MA, USA). Quality-adjusted life-years and costs were estimated and discounted at an annual rate of 3.5%. All costs are reported in 2018 Great British pounds.

Based on consultation with clinical team members and current National Institute for Health and Care Excellence surveillance guidelines, a total of 75 alternative NHS strategies for stage IA and 87 strategies for stage IB were identified for initial modelling. The main probabilities used in the model were the probabilities of recurrence, the probabilities of treatment success, patient self-diagnosis, 'false alarms' resulting in emergency visits, transition probabilities between the different stages of melanoma and mortality rates taken from various international data sources. EuroQol-5 Dimensions utility estimates were derived from a systematic review and meta-analysis of studies from Australia, Europe and North America. Both deterministic and probabilistic sensitivity analyses were used to explore uncertainty. The age and sex information from the individual patient data obtained from the University Hospital of North Durham (NREC 19/NE/004) were used to estimate the mortality rate of the simulated cohort.

Results

Initial modelling showed that strategies involving cancer nurse specialists providing clinical examinations were unlikely to be cost-effective, primarily because of the comparatively poorer diagnostic accuracy assumed.

From initial modelling, 20 surveillance strategies each for stages IA and IB were evaluated in more depth. For both stages, the evaluated strategies were similar in terms of quality-adjusted life-years, reflecting the relatively low rates of recurrence expected. The strategy of follow-up once for 1 year by a dermatologist was the least costly and most likely to be considered cost-effective if society were willing to pay £20,000 per quality-adjusted life-year (AJCC stage IA, 13%; AJCC stage IB, 13%). For stage IA, the strategy recommended by the National Institute for Health and Care Excellence performed similarly (12%). For stage IB, the strategy recommended by the National Institute for Health and Care Excellence performed poorly (4%). Although these probabilities are low, a large number of different surveillance strategies were compared. A sensitivity analysis showed that there may be value in improving the diagnostic performance of cancer nurse specialists and in the use of low-cost risk prediction tools for prognosis.

The highest value for research came from removing all uncertainty around probabilities of transitioning between the different stages of melanoma (stage IA, £380M; stage IB, £457M) and diagnostic accuracy (stage IA, £276M; stage IB, £193M). The value of removing uncertainty in utilities was lower, but still substantial.

Conclusions

Few data were available specific to surveillance of people after treatment for melanoma. Furthermore, few data were available for key components of a surveillance strategy that could be used to model alternative strategies. What data were available mainly related to studies using cancer staging classifications predating the publication of the American Joint Committee on Cancer eighth edition. Therefore, results are imprecise and considerable uncertainty exists. There is insufficient evidence to recommend any changes to the current surveillance guidelines produced by the National Institute for Health and Care Excellence. There are plausible surveillance strategies that may perform better than current recommendations for surveillance. However, for those treated for stage IA disease, the National Institute for Health and Care Excellence's strategy still performs comparatively well. For stage IB disease, the strategy recommended by the National Institute for Health and Care Excellence of follow-up every 3 months for 3 years then every 6 months for a further 2 years performs poorly, compared with other strategies considered, but there is insufficient evidence to support any changes.

Surveillance strategies whereby the clinical follow-up is conducted by a cancer nurse specialist may ease pressure on dermatologists and plastic surgeons. However, methods to enhance cancer nurse specialists' diagnostic performance may be needed, as the current limited evidence base suggests that their ability to correctly identify who does or does not have a recurrence is not as good as that of dermatologists. Likewise, encouraging and supporting patients in making accurate self-diagnosis of recurrence in stage I disease may reduce the need for any active surveillance strategy for those initially treated for stage I disease.

Recommendations for research

It is tempting to recommend that a randomised controlled trial should be conducted to compare surveillance strategies. However, a surveillance strategy is a complex intervention and research should first establish what sensible comparators there should be against current practice. What an appropriate comparator would be may vary between stage IA and stage IB disease, and establishing this requires improved evidence on how disease in patients with stage I melanoma develops over time. The economic modelling shows that both the incidence of recurrent and metastatic disease over time, and how it progresses are important. Further research would also be valuable on how well recurrent and metastatic disease is diagnosed, improving the diagnostic performance of practitioner groups like cancer nurse specialists, developing and evaluating low-cost tools that can better stratify patients into low or high risk of future recurrence and metastasis, and identifying the patient preferences for alternative methods of surveillance and on the impacts on health-related quality of life in patients with melanoma.

Study registration

This study is registered as PROSPERO CRD42018086784.

Funding

This project was funded by the National Institute for Health Research Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol 25, No. 64. See the NIHR Journals Library website for further project information.

Chapter 1 Background

Description of the underlying health problem

Melanoma is one of the most deadly of all skin cancers.[1,2] Metastatic melanoma is a highly aggressive disease with rapid dissemination and, until recently, had a median overall survival of between 6 and 10 months once metastasis had occurred.[3] The introduction of targeted therapies and immunotherapies has improved outcomes for these patients, with median overall survival now reaching at least 2 years.[4]

The worldwide incidence of melanoma is estimated to be approximately 2% of the population per annum,[5] which continues to increase across the globe, with the highest incidence rates being seen in Australia, New Zealand, northern Europe and the USA. In Australia/New Zealand, the incidence has been reported as being > 33 cases per 100,000 population, followed by northern Europe (Norway, Denmark and the Netherlands) with > 25 cases per 100,000.[6] In Australia, the lifetime risk of developing melanoma is 1 in 25 for men and 1 in 34 for women.

Melanoma affects a disproportionate number of people aged < 50 years, compared with other cancers.[7] For example, 11% of all melanomas are diagnosed in those aged < 50 years, compared with 5% for other cancer types.[8,9]

In the UK, 2019 figures estimate that the incidence rate has increased 134% since the 1990s, and melanoma is now the fifth most common cancer, accounting for 5% of all new cancer cases, which is on a par with the incidence in other European countries.[10] Globally, there are approximately 232,000 new cases of melanoma diagnosed annually, of which > 140,000 are in Europe.[11–13]

Cutaneous melanoma is a cancer that develops from pigmented cells (melanocytes) in the skin. Melanocytes are responsible for production of the main pigment in the skin: melanin. The proportion of the darker eumelanin and lighter pheomelanin play key roles in offering protection against deoxyribonucleic acid damage induced by ultraviolet (UV) radiation. The development of melanoma may occur de novo from melanocytes, or in a stepwise manner from benign naevus to invasive melanoma.[14,15]

Description of current service provision

Staging of disease

There have been great advances in the earlier detection of primary melanoma through increased public awareness, the adoption of dermatoscopic examinations and a rapid '2-week wait' referral system in the UK.[16] There is also widespread belief that earlier detection of metastatic disease results in improved patient outcomes.[17] However, at present, there is no internationally accepted standardised model of follow-up of patients diagnosed with cutaneous melanoma, with wide variations in care across North America, Australia, Europe and the UK.[18]

When it comes to follow-up of those treated for melanoma, disease staging and judgements on the risk of spread (metastasis) are based on the microscopic appearance and depth of the original tumour. Currently, this is based on the American Joint Committee on Cancer (AJCC) eighth edition staging criteria, which were published in 2016 and formally implemented on 1 January 2018.[19] However, as this is an evidence synthesis project, the definitions used in the 2010 seventh edition[20] are also pertinent, as existing data would have based decisions on staging using this edition or earlier editions. The seventh edition included mitotic count (the number of actively dividing cells in the tumour), as this was thought to be an important prognostic feature for thin melanomas,[21] but this has been dropped from the eighth edition staging

guidelines because of a lack of evidence supporting its prognostic significance. The key definitions of stage I and II disease for both the seventh and eighth editions are set out in *Table 1*.

Specific changes of note between the seventh and eighth AJCC staging criteria affecting stage I melanoma criteria are as follows:

- T1a has had the Breslow depth reduced to 0.8 mm when non-ulcerated
- T1b is now any ulcerated tumour of < 0.8 mm Breslow depth or between 0.8 and 1.00 mm Breslow depth regardless of ulceration status
- mitotic count has no role in the defining stage.

One further change of note is the distinction between clinical and pathological staging for non-ulcerated tumours that are between 0.8 and 1 mm. These tumours are clinical stage IB, but if they undergo a sentinel lymph node biopsy (SLNB) that is negative, the tumour is 'downgraded' to pathological stage IA,

TABLE 1 The AJCC staging of primary cutaneous tumours

TNM stage	Breslow thickness (mm)	Ulceration	Mitotic count	AJCC stage
AJCC – seventh edition[20]				
T1				
T1a	< 1.00	Absent	< 1 mitosis/mm^2	IA
T1b	< 1.00	Present	≥ 1 mitosis/mm^2	IB
T2				
T2a	1.01–2.00	Absent	N/A	IB
T2b	1.01–2.00	Present	N/A	IIA
T3				
T3a	2.01–4.00	Absent	N/A	IIA
T3b	2.01–4.00	Present	N/A	IIB
T4				
T4a	> 4.00	Absent	N/A	IIB
T4b	> 4.00	Present	N/A	IIC
AJCC – eighth edition[19]				
T1				
T1a	< 0.80	Absent	Not included	IA
T1b	< 0.80	Absent	Not included	IB
T2				
T2a	1.01–2.00	Absent	Not included	IB
T2b	1.01–2.00	Present	Not included	IIA
T3				
T3a	2.01–4.00	Absent	Not included	IIA
T3b	2.01–4.00	Present	Not included	IIB
T4				
T4a	> 4.0	Absent	Not included	IIB
T4b	> 4.0	Present	Not included	IIC

N/A, not applicable; TNM, tumour node metastasis.

with associated changes in overall prognostication. This subcohort of patients is likely to represent a tiny proportion of UK patients, given that, under the National Institute for Health and Care Excellence (NICE) guidelines,[16] they would not be routinely offered SLNB if their tumour is of < 1 mm Breslow depth.

The AJCC stage I disease encompasses both stages IA and IB disease and represents the thinnest tumours. At initial diagnosis, 70% of melanomas are classified as AJCC stage I. Early-stage tumours are treated by surgical excision. However, the 5-year overall survival of patients with stage I disease is only 95%.[22] Stage II disease encompasses thicker, but still localised, tumours. Stages III/IV patients have evidence of local and distant metastases; 2-year mortality is up to 82% in stage IV disease, although, with the introduction of new systemic agents, this is now falling.

Surveillance strategies

Given the relatively low rates of local or systemic recurrence, those with AJCC stage I melanoma may not need the same level of clinician follow-up as is generally recommended,[23] whereas patients at higher risk following surgical treatment may benefit from more intensive future surveillance to detect recurrent or metastatic disease early. However, balanced against this is the fact that approximately 10% of patients with AJCC stage I disease develop metastases and the prognosis for these people remains poor. Currently, this results in rigorous, routine follow-up for all melanoma patients.

The potential interventions and investigations used as part of a post-surgical treatment surveillance strategy also varies in AJCC stage I patients. An important element of surveillance is education of the patient to allow them to identify any new lesions of concern or signs of recurrence. In one study by Hofmann *et al.*,[24] 30 (24%) of the 127 patients who had a first relapse were not being formally followed up at the time that the relapse was detected. This is because they had never been in a follow-up programme, had dropped out of follow-up, or had completed the formal follow-up process. These data demonstrate the often erratic and unpredictable course of the disease. In the same study, 68% of first relapses were detected by follow-up activity.

Nevertheless, regular clinical history and examination is the mainstay of most surveillance guidelines. Again, which type of health-care practitioner (e.g. nurse, surgeon, dermatologist) undertakes the examinations varies, as does the setting of these reviews, with recommendations for either primary care-based or secondary care-based (in-hospital) appointments. Specific radiological examination of patients may also be recommended for follow-up of stage I melanoma. Routine use of imaging modalities aims to detect the development of regional and distant metastases as early as possible, even before these become clinically apparent. However, if a patient is found to have clinical evidence of metastases, a further set of imaging modalities such as ultrasonography, computerised tomography (CT) and positron emission tomography-computerised tomography (PET-CT) may be used. These methods allow targeted biopsy, when possible, of the relevant melanoma deposits to allow histopathological assessment of the tissue. In the UK, modalities such as CT and PET-CT are not currently advocated as part of the routine management and follow-up of AJCC stage I disease, and it is felt to be unlikely that this situation will change in the next 5–7 years. As others have noted,[25] and described in more detail in *Chapter 2*, there is considerable variability of surveillance practices worldwide.

The British Association of Dermatologists (BAD) revised UK guidelines for the management of cutaneous melanoma[26,27] and the more recent NICE *Melanoma: Assessment and Management* guideline[16] advise that patients who have stage I melanoma are followed up to detect signs of recurrence after history and examination. This surveillance is undertaken as follows:

- Patients with stage IA melanoma should be seen two to four times over a period of up to 12 months, and then discharged.
- Patients with stage IB melanoma should be seen every 3 months for 3 years, and then every 6 months for a further 2 years.

There are no recommendations for the routine use of any radiological modality, only guidance that these can be implemented if required in symptomatic patients.

There are currently no biomarkers in routine use in any guidelines for stage I disease. Lactate dehydrogenase blood levels are validated for use in patients with evidence of metastases only.[28] There is also increased application of serum S100B, but, once again, in patients with evidence of metastatic disease only.[29]

Any changes to current recommended practice would need to consider multiple components, each of which would determine the costs, effectiveness, feasibility and acceptability of an alternative strategy (*Box 1*).

Current treatment options

The NICE guideline for melanoma,[16] published in July 2015, recommends excision of melanomas of stages 0–II with a 0.5- to 2-cm total margin, depending on stage and histopathological assessment of the biopsy (*Figure 1*). Risk of disease progression is estimated based on the AJCC eighth edition staging criteria.[19] As described in the previous subsection, patients deemed to be at a low risk of disease progression are followed up at regular intervals for a period of 5 years following diagnosis, undergoing visual and physical examinations.

Current draft NICE guidance does not suggest using SNLB for AJCC stage IA or IB melanomas with a Breslow thickness of < 1 mm, and acknowledges that a proportion of those patients with a negative SLNB will still experience melanoma recurrence.[16] This stance is further supported by a UK consensus statement made through a multidisciplinary meeting held by the Melanoma Focus group in 2018.[30]

Since 2011, there have been rapid developments in the therapeutic options available for metastatic disease, with accompanying improvements in patient-related outcomes. These developments have been so rapid that we are currently in the follow-up period for many drug trials.

BOX 1 Components that need to be assessed for any potential surveillance strategy

- Person(s) undertaking surveillance:
  - Patient, dermatologist, surgeon, primary care physician, specialist nurse, combination of practitioners.
- Site of surveillance:
  - Patient's home, primary/community care, secondary care setting.
- Availability/clinical utility of prognostic risk prediction tools for further disease stratification.
- Interval timing of review appointments.
- Duration of overall surveillance:
  - Immediate discharge, 1 year, 5 years, 10 years, life.
- Routine imaging interventions:
  - Which modalities, how often.
- Assessment of clinical benefit from surveillance strategy.
- Acceptance of any model by melanoma patients and service providers.
- Value for money.

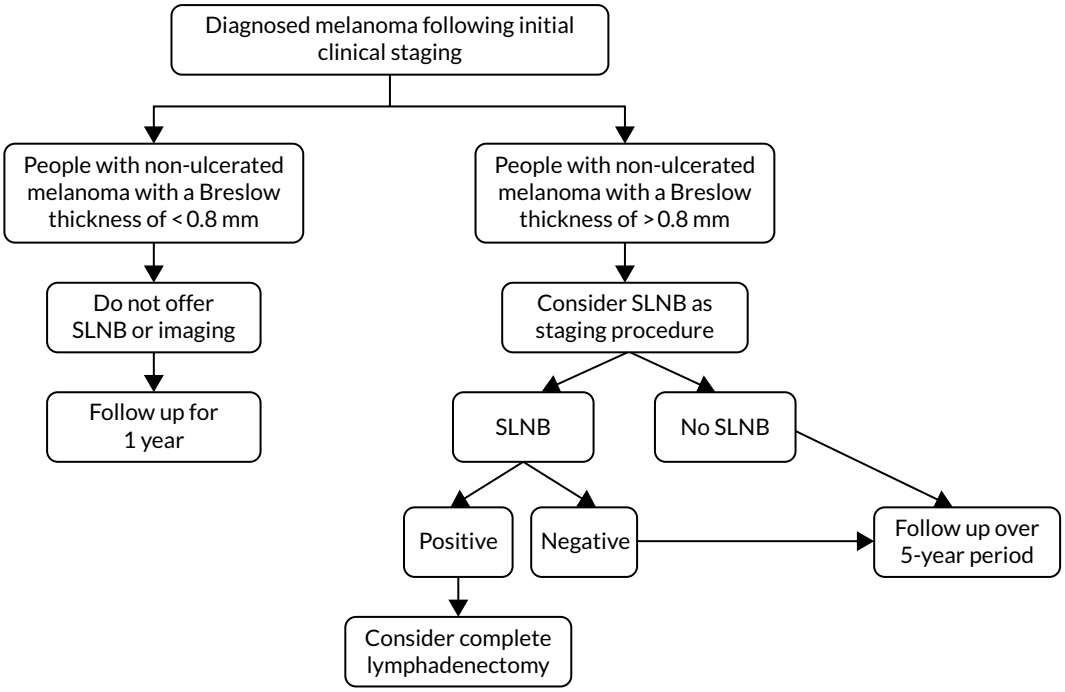


FIGURE 1 The National Collaborating Centre for Cancer's melanoma staging algorithm, including recommended use of SNLB.

Metastatic disease encompasses the following:

- Satellite lesions – skin or subcutaneous deposits within 2 cm of the primary tumour.
- In-transit metastases – these occur > 2 cm from the primary tumour, but before the regional lymph node.
- Nodal micrometastases – metastatic deposits evident only following histopathological analysis of SLNB tissue or regional lymph node dissection.
- Nodal macrometastases – metastatic deposits in regional lymph nodes that are either clinically apparent or found on histopathological assessment of regional lymph node dissection.
- Metastases to distant skin, subcutaneous tissue, lymph nodes or other visceral sites/organs.

Localised metastatic disease is broadly distinguished based on the distance of spread and the total metastatic tumour bulk (and is based on AJCC staging criteria):

- IIIA
  - one to three local lymph nodes with micrometastases (diagnosed on SLNB or node dissection)
- IIIB
  - one to three local lymph nodes with macrometastases (clinically palpable lymph node involvement or within-node dissection)
  - in-transit metastases/satellite lesions with no metastatic lymph node involvement
- IIIC
  - four or more local lymph nodes involved
  - in-transit metastases/satellite lesions with frank metastatic lymph node involvement.

'In-transit metastases' covers a wide range of clinical presentations, ranging from localised, small melanoma deposits that are easily amenable to further surgery to > 100 deposits of bulky melanoma tissue. In such cases, the clinical decisions are made based on the extent and technical feasibility of treatment.

Among the most established therapies for in-transit metastases are isolated limb perfusion (ILP) and isolated limb infusion (ILI). Both of these therapies involve the isolation of a limb's vasculature, with the addition of an anti-tumour agent into this closed system. The aim of therapy is to allow anti-tumour concentrations of the chemotherapeutic agent, without the associated systemic side effects. Traditionally, ILP and ILI have been carried out using melphalan, but, recently, they have been carried out with the addition of tumour necrosis factor. Overall, although tumour response rates range from 64% to 93%,[31,32] the median survival time post treatment is still only 2 years.[33] There is currently no suggestion that ILP/ILI can be used in localised melanomas without any evidence of frank metastatic disease.

For metastatic deposits in lymph nodes (following detection by either SLNB or nodal biopsy), the most common therapy is for a lymphadenectomy (with or without post-operative radiotherapy[34]) of the involved lymph node basin. This has significant morbidity attached to the procedure and it is debatable whether or not there is any benefit for patients in terms of overall melanoma survival; it is currently not recommended routinely in SLNB-positive patients.[30,35,36]

Distant metastases, encompassing stage IV disease, rely on systemic therapeutic options. In recent years, a raft of new therapeutic agents have been introduced. The current standard of care in the UK is in constant flux, but remains based on NICE guidance distilled from the continually changing evidence base for systemic therapies; however, there is still variation in local practice. It is generally accepted that adjuvant therapy with immune modulators should be made available to patients with frank stage II disease, or high-risk stage IIIA disease (a deposit of melanoma of > 1 mm^2 in the lymph node following SLNB), and that this is also a first-line treatment for stage IV disease. Combination therapies are also preferable for first-line use in stage IV or unresectable stage III disease.

The newer systemic agents can be categorised by their mode of action, either targeting the mitogen-activated protein kinase signalling pathway, or via immune checkpoint blockade. A multitude of clinical trials have been undertaken assessing the benefits of each group as first-line systemic therapy in patients with metastatic disease (usually AJCC IIIB and above), either as monotherapy or combined with another agent affecting the same pathway. *Table 2* outlines the most influential recent clinical trials.

The vast majority of systemic agents are aimed at patients with evidence of distant disease progression. However, with the long-standing hypothesis that earlier introduction of systemic therapies may result in better response outcomes, studies have shown a benefit in introducing systemic agents before there is clinical evidence of metastasis. The 2019 NICE guidelines[47] for treating stage III melanoma recommend that consideration be given to the use of two adjuvant therapies, nivolumab and pembrolizumab, in resected melanoma with evidence of lymph node involvement, including stage IIIA disease identified following SLNB. Similarly, dabrafenib and trametenib are also licensed and approved for resected stage III disease in *BRAF*-positive patients.

Such adjuvant regimes continue to be studied,[48] but there is a need for better risk prediction of the prognosis of people treated for melanoma, especially for those with an earlier-stage disease, a significant minority of whom will experience progression. However, even for later-stage disease, a considerable number of people may receive these newer systemic regimens, but with only limited prospect of any gain. Hence, they are potentially being unnecessarily exposed to the side effects of systemic therapy with little or no overall benefit.

TABLE 2 Overview of recent clinical trials of systemic medication

Drug	Trial name	Stages enrolled	Main outcomes
Nivolumab (Opdivo, Bristol Myers Squibb, New York City, NY, USA)	CheckMate 037[37]	Unresectable III, IV (second-line study in patients progressing following ipilimumab or targeted therapy)	• Median PFS 3.1 months • Median OS 16 months in favour of nivolumab
	CheckMate 066[38]	IV	• 12 month OS 72.9% • Median PFS 5.1 months
	CheckMate 238[39]	IIIB, IIIC, IV	• 12-month RFS 70.5%
Ipilimumab (Yervoy, Bristol Myers Squibb, New York City, NY, USA)	CheckMate 067[40]	Unresectable III, IV	• 5-year OS 26%
	CheckMate 238[39]	IIIB, IIIC, IV	• 1-year RFS 60.8%
Pembrolizumab (Keytruda, Merck Sharp Dohme, Kenilworth, NJ, USA)	KEYNOTE-002[41]	Advanced melanoma	• OS 14.7 months
Vemurafenib (Zelboraf, Hoffmann La Roche, Basel, Switzerland)	COLUMBUS[42]	IIIB, IIIC, IV	• Median PFS 7.3 months
	COMBI-v[43]	IIIC, IV	• Median PFS 7.3 months
	BRIM-3[44]	IIIC, IV	• Median OS 13.6 months • 4-year OS 19%
Encorafenib (Braftovi, Array Biopharma Inc, Boulder, CO, USA) + binimetinib (Mektovi, Array Biopharma Inc, Boulder, CO, USA)	COLUMBUS[42]	IIIB, IIIC, IV	• Median PFS 14.9 months
Nivolumab + ipilimumab	CheckMate 067[40]	Unresectable III, IV	• 5-year OS 52%
Trametinib (Mekinist, Novartis, Basel, Switzerland) + dabrafenib (Tafinlar, Novartis, Basel, Switzerland)	COMBI-d[45]	IIIC, IV	• 5-year PFS 19% • 5-year OS 34%
	COMBI-AD[46]	High-risk IIIA, IIIB, IIIC	• 4-year RFS 54%

OS, overall survival; PFS, progression-free survival; RFS, recurrence-free survival.

Description of the technologies under assessment

The technologies under assessment are alternative approaches to the surveillance of people who have been treated for AJCC stage IA or IB disease. Specifically, we will be considering strategies that vary in terms of one or more of the following, and which may include there being no organised surveillance in place:

- person(s) undertaking surveillance
- site of surveillance
- availability/clinical utlity of prognostic risk prediction tools for further disease stratification
- interval timing of review appointments
- duration of overall surveillance.

Summary of patient engagement

In addition to drawing on wider patient and public involvement activities undertaken by members of the research team, the study team included three people who have personal experience of melanoma and who are already engaged more broadly with members of the research team in improving the care for those with melanoma. These people were included as co-applicants on the original application to the National Institute for Health Research, and they commented and advised on that application.

They have been involved as the work progressed, particularly in helping shape detailed research plans during an advisory group meeting held in May 2018 and in discussions around whether or not the research is likely to meet service user needs, and how it could be best modified to do so. They and the rest of the team have also discussed, via e-mail and during an advisory group meeting held in July 2019, the results of the research. These discussions have been used to draw out key findings and the implications for patients, the public, practitioners, the NHS and further research.

Decision problem

Given that the incidence of cutaneous melanoma is increasing and the majority of people treated for melanoma have AJCC stage I disease that is seemingly low risk, there is an urgent, unmet need to identify those patients with the genuinely lowest-risk disease. Currently, a rigorous patient follow up is routinely carried out for all patients, perhaps unnecessarily straining health-care resources that are already stretched. Identifying those patients with genuinely low-risk disease and discharging them from follow-up earlier could save the NHS upwards of £22.5M over a 5-year period,[49] facilitating reallocation of these resources to the smaller group of high-risk patients.

There is little evidence-based guidance on how surveillance regimens should be organised, with considerable variability internationally. Before any changes in a surveillance strategy are introduced, it is essential that any alternative is evidence based. This means that it is essential to gather and synthesise what is already known in a transparent, concise manner to help guide judgements.

Specifically, for those who have been treated for AJCC stage I disease, we want to help reduce the anguish and distress felt genuinely by truly low-risk patients who unnecessarily fear that they are at risk of metastatic disease. A systematic review by Rychetnik *et al.*[50] in 2013 reported that around half of melanoma patients surveyed said that follow-up appointments made them anxious (with clinically significant levels of anxiety in approximately 20% of patients), sometimes accompanied by physical symptoms that can start weeks before the appointment. Should it be shown to be safe to follow up low-risk patients less intensively, then some of this distress and anxiety could be mitigated. Conversely, a less intensive follow-up may increase anxiety that a cancer could be missed, thereby running the very real risk that detection of metastasis may be delayed.

Should a viable alternative surveillance regimen be identified, in addition to the health impacts on patients, and the service implications to the NHS, there should be a decrease in the number of follow-up appointments. Thus, there could be reductions in the time and travel costs of attending visits incurred by patients and their families, wider system effects of less time away from usual activities (the majority of people treated for melanoma are aged $\leq$ 50 years and many have work and carer responsibilities) and an impact on traffic pollution caused by reduced patient travel.

To address the evidence gap, this assessment includes an evidence synthesis of relevant information needed to construct and evaluate alternative surveillance strategies. It includes a set of systematic reviews addressing different aspects of the decision problem, an economic decision model to determine the most effective and cost-effective strategy, and a value-of-information analysis to help inform the direction of future research.

Aims and objectives

The aim of this research was to evaluate the effectiveness and cost-effectiveness of different surveillance strategies for patients with AJCC stage I melanoma after surgical excision of a primary cutaneous tumour.

To meet this aim, the objectives were to:

1. identify different strategies for surveillance and follow-up after surgical excision of a primary cutaneous tumour and review the evidence on their effectiveness and cost-effectiveness
2. determine the prognostic performance of risk models used to determine the prognosis and risk stratification of patients with AJCC stage I melanoma after surgical excision of a primary cutaneous tumour
3. determine the diagnostic performance of tests used in surveillance and follow-up strategies in detecting recurrence and metastatic diseases in patients with AJCC stage I melanoma after surgical excision of a primary cutaneous tumour
4. develop a decision-analytic model to estimate the effectiveness and cost-effectiveness of the surveillance and follow-up strategies after surgical excision of a primary cutaneous tumour
5. undertake a value-of-information analysis to assess the need for further primary research.

Structure of the report

As with most health technology assessments, the work conducted has several related pieces of work. *Chapter 2* puts the research in context by presenting a summary of the existing guideline recommendations for the surveillence of stage I melanoma and considering the underlying quality of those recommendations. Each of the four subsequent chapters addresses one or more objective (see *Aims and objectives*), with the earlier pieces of work informing later pieces of work. Objectives 1–3 are addressed using systematic review methods that are appropriate to their objectives. These systematic reviews are reported in *Chapters 3–5*. These reviews are then used to inform and parameterise the economic evaluation decision model reported in *Chapter 6*. Each of *Chapters 3–6* ends with a discussion of the chapter's findings, and an overall summary of key findings is provided in the discussion (*Chapter 7*), along with strengths and limitations of the work and implications for practice and for future research. A schematic for how the different elements of reseach fit together is shown in *Figure 2*.

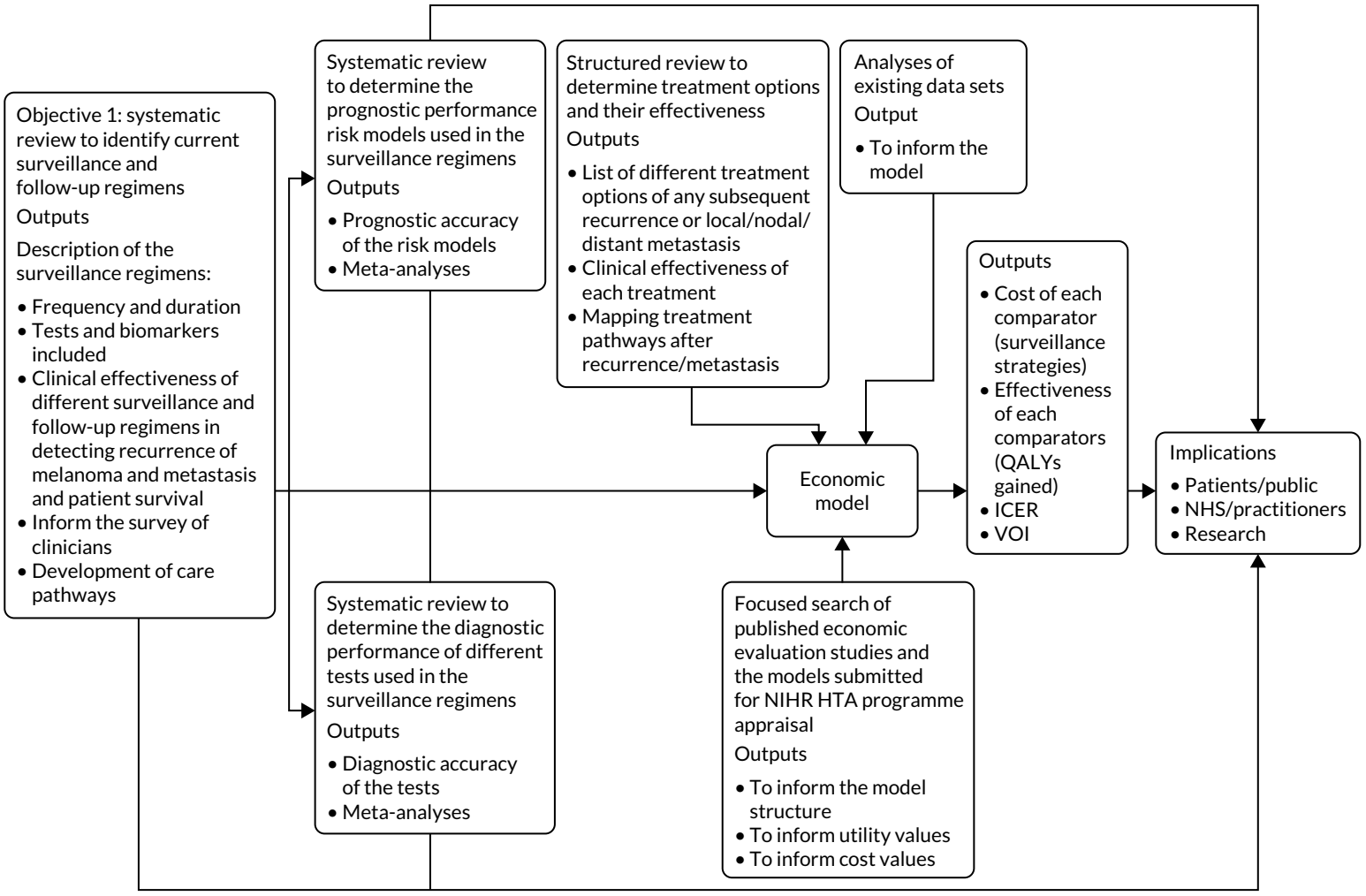


FIGURE 2 Schematic of the components of the health technology assessment. HTA, Health Technology Assessment; ICER, incremental cost-effectiveness ratio; NIHR, National Institute for Health Research; QALY, quality-adjusted life-year; VOI, value of information.

Chapter 2 Summary of existing guidelines for surveillance following treatment of stage I melanoma

Introduction

As described in *Chapter 1*, melanoma is a global health burden with a rising incidence. Given this, there is a need for guidelines focusing on prevention, diagnosis and further management. This chapter provides a narrative critique of the current melanoma guidelines available globally, with particular emphasis on the surveillance strategies for stage I melanoma.

The aims of this chapter are to (1) summarise the existing recommendations on surveillance for stage I melanoma and (2) consider if differences in recommendations can be explained in terms of the differences in the evidence base used, the interpretation of that evidence base or the methods adopted to develop the guideline. To provide a common basis of comparison, the Appraisal of Guidelines for Research & Evaluation II (AGREE II) was used.[51,52]

For the critique of how the surveillance recommendations were developed, three domains of the AGREE II were focused on. These were 'scope and purpose', 'rigour of development' and 'clarity of presentation', as these were most pertinent to our aims in reviewing these guidelines. These domains are briefly described in the following paragraphs.

The scope and purpose of a guideline includes its objectives and relevant health questions, as well as the population of interest. In this domain, the health intent, interventions, target population, outcomes/benefits, as well as context/setting, should be clearly stated in the guidelines. In addition, the disease stage, associated comorbidities and appropriate comparators should be included in the guidance, and appropriate health questions defined. This domain aims to clarify the potential impact of the guidance. For instance, the NICE melanoma guidance states its aim as:

> *. . . the assessment and management of melanoma . . . in children, young people and adults. It aims to reduce variation in practice and improve survival.*
>
> *Reproduced with permission from NICE.[16] © NICE 2015. Melanoma: Assessment and Management. Available from www.nice.org.uk/guidance/ng14. All rights reserved. Subject to Notice of rights (www.nice.org.uk/terms-and-conditions#notice-of-rights). NICE guidance is prepared for the National Health Service in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication*

This covers the expected outcomes (i.e. reducing variation in practice and improving survival), the target population and the health intent. To fully achieve these stated aims and support critical recommendations, there is a need for well-tailored questions to be included in the guideline.

The 'rigour of development' domain looks at the methodological thoroughness employed in producing a guideline. It covers the search process for supporting materials, selection criteria, description of the strengths and limitations of the body of evidence, clear description of evidence formulation process and explicit links between recommendations and supporting evidence. In addition, it appraises the review processes before publication and plans for updating the guidelines. For example, in the appendix section of its guidance, the Australian clinical practice guidelines for the diagnosis and management of melanoma[53] state the stepwise process used in developing its guidelines, including the role of systematic reviews in

providing key recommendations, dissemination to relevant stakeholders and plans for future updates. When available, it also offered explicit references to published literature for key recommendations.

Clarity of presentation examines how explicit the recommendations are, the provision of different management options and the ease of identifying key recommendations. Explicit guidance should be clear on the population that is affected by the recommendation, the intent of the recommendation, appropriate provisos and descriptions of alternatives, as well as being aesthetically accessible. For instance, flow charts, summary boxes or other forms of graphics may be employed in presenting the entire guideline or sets of recommendations, grouped according to relevance.

Ten melanoma guidelines were identified through grey literature searches carried out in May 2018 and updated in August 2019 (the guidelines were systematically identified as part of the searches conducted for the systematic review of surveillance strategies, reported in *Chapter 3*). *Table 3* summarises each guideline in terms of their recommendations for surveillance following treatment for stage I melanoma.

TABLE 3 Main features of national and regional melanoma guidelines around follow-up regimes for melanoma

Guideline	Duration of follow-up	Routine investigations	Clinician undertaking surveillance
• NICE 2015[16] • NICE 2019[54]	• AJCC stage IA: every 3–6 months for 1 year • AJCC stage IB: every 3 months for 3 years then every 6 months for a further 2 years	*Do not routinely offer screening investigations (including imaging and blood tests) as part of follow-up* *Reproduced with permission from NICE.[54] © NICE 2015. Surveillance Proposal Consultation Document: 2019 Surveillance of Melanoma (NICE Guidelines NG14 and CSG8). Available from www.nice.org.uk/guidance/ng14/documents/surveillance-review-proposal. All rights reserved. Subject to Notice of rights (www.nice.org.uk/terms-and-conditions#notice-of-rights). NICE guidance is prepared for the National Health Service in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication*	• Secondary care follow-up recommended • No specific clinician specialties identified
BAD 2010[27]	• AJCC stage IA: every 3–6 months for 1 year • AJCC stage IB: every 3 months for 3 years then every 6 months for a further 2 years	Nil	• Secondary care follow-up recommended, but can be undertaken in primary care if agreed • No specific clinician specialties identified
NCCN Guidelines® 2019[55]	AJCC stages IA and IB: every 6 months for 5 years, then yearly	Nil	Not specified

TABLE 3 Main features of national and regional melanoma guidelines around follow-up regimes for melanoma (*continued*)

Guideline	Duration of follow-up	Routine investigations	Clinician undertaking surveillance
ESMO 2015[56]	Defer to national guidelines	Nil	Not specified
American Academy of Dermatology 2019[57]	Every 6–12 months for 2–5 years. At least annually thereafter	Nil	Not specified
Cancer Council Australia Melanoma Guidelines Working Party[53]	AJCC stage I: annually up to year 10	Nil	Follow-up with a medical professional (GP, dermatologist, surgeon or medical oncologist)
Dutch Working Group on Melanoma 2013[58]	• AJCC stage IA: one-off visit 1 month after diagnosis • AJCC stage IB: visits every 3 months for first year, every 6 months for second year and every 12 months for years 3–5	Nil	Not specified
German Guideline Program in Oncology 2013[59]	• AJCC stage IA: every 6 months in years 1–3, then every 12 months up to year 10 • Stage IB: every 3 months in years 1–3, every 6 months in years 4 and 5, then yearly up to year 10	AJCC stage IB: • Lymph node ultrasonography every 6 months for 3 years • S100 serum analysis every 3 months for 3 years	Not specified
Swiss Cancer League 2016[60]	• AJCC stage I (≤ T1N0): every 6 months in years 1–3, then yearly up to year 10 • AJCC stage I (T2N0): visits every 3 months in years 1–3, every 6 months in years 4 and 5, then every 6–12 months up to year 10	AJCC stage I (T2N0): • S100 every 6–12 months for years 1–5 • Abdominal sonography and chest radiography in years 1–5 in individual cases	Not specified
Brazilian guidelines 2016[61]	No explicit recommendations for AJCC stage I melanoma	None routinely	Not specified

ESMO, European Society of Medical Oncologists; GP, general practitioner; N, node; NCCN®, National Comprehensive Cancer Network®; T, tumour.

As *Table 3* illustrates, there are variations between guidelines in their recommendations on surveillance. This variation exists not just for the intensity and duration of follow-up, but also over the tests that are recommended and who conducts the surveillance. In part, differences in who performs the surveillance may relate to differences in geography and national priorities, but other differences are less easily explained. By summarising the guidelines in terms of the selected AGREE II criteria, it may be possible to shed light on why guideline recommendations differ.

Summary of individual guidelines according to the selected AGREE II criteria

National Institute for Health and Care Excellence: Melanoma: Assessment and Management

This NICE guideline[16] was published in 2015 to guide clinical practice for the management of melanoma in England. The process of the guideline formulation began in 2013 and the full draft was

published in July 2015. There is an ongoing peer review of the guideline; the most recent report was in 2019,[54] and suggested significant section updates to the guidelines.

Scope and purpose

In the introduction to the NICE melanoma guideline,[16] the purpose was clearly stated as focusing on where there was differences in clinical practice. Recommendations were made on the staging and treatment of melanoma. This included the use of chemotherapy and immunotherapy to address more advanced disease.

The target populations were clearly defined and appropriate caveats were issued for specific population groups, such as children and adolescents. The health-care context was defined as appropriate for the various recommendations. As part of the guideline, tools to aid the implementation of the guideline in various settings were included in the additional tools and resources section of the guideline.

Rigour of development

The NICE guideline[16] contains a chapter detailing the methodology used in the design of the guideline. In developing the NICE guideline, a systematic review of the available evidence was undertaken using the population, intervention, comparator, outcome (PICO) question format and the evidence was graded using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) and Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tools. The grading of the evidence for the recommendations on follow-up for stage I melanoma was very low. The NICE guideline[16] was also balanced against economic evaluations on efficiency in making its recommendations. The Portable Document Format (PDF) file of the recommendation did not include explicit links to the sources of evidence, as seen in the web version. However, the PDF of the full guidance has a record of the evidence used and summarises this in the paragraphs preceding the recommendation summary box. The evidence used as the basis of recommendations around follow-up for stage I disease were mainly based on data from case series.

Clarity of presentation

The NICE guideline[16] includes flow charts, with headers based on the management options, and further classifications within each header into the various stages using the AJCC classification for melanoma. The web and PDF versions of the recommendations are easy to navigate, as they contain links to other areas of the guidelines. The recommendations on the follow-up of stage I melanoma were explicit and different from other stages. For instance, in addition to the general follow-up recommendations for all people who have had melanoma, those with stage IA were recommended to have follow-up for 1 year, whereas those with stage IB–IIB/IIC with negative sentinel nodes were to be followed up for 5 years. In both groups (i.e. stage IA and stage IB–IIC with negative SLNBs), routine screening investigations were not recommended.

A clear distinction was also made between patients with stage IIC melanoma who had negative sentinel nodes and those who had no lymph node biopsied. Those with no lymph node biopsy were followed up as a stage III melanoma, with recommendations to consider surveillance imaging as part of follow-up.

British Association of Dermatologists revised UK guidelines for the management of cutaneous melanoma

The 2010 revised UK guidelines for the management of cutaneous melanoma[27] are the current guidelines from the BAD, which have not been updated because of ongoing updates by NICE on melanoma management (see *National Institute for Health and Care Excellence:* Melanoma: Assessment and Management[16]). The BAD's methodology for guideline formulation was also accredited by NHS evidence.

Scope and purpose
A multidisciplinary group was employed to agree on the best management practice for cutaneous melanoma in the UK.

Rigour of development
Although searches were carried out, based on the available evidence, it was difficult to ascertain the rigour of such searches, as some were performed by individual authors in addition to group search findings. Full details of the searches were not readily available. However, in the introductory paragraphs, search terms were listed and the use of the PubMed database was alluded to, but the inclusion or exclusion criteria were not stated. The assessment of the BAD's guideline against the AGREE criteria was discussed in an article by Cox and Williams[62] in 2003, in which it was noted that the method of formulating evidence was not usually specified in the BAD guidelines. As detailed in appendix 1 of the guideline, it used both levels of evidence (I–IV) and a five-point grading system (A, meaning 'there is good evidence to support the use of the procedure', to E, meaning 'there is good evidence to support the rejection of the procedure') to assess the quality of the evidence. The source or the process of formulation of these levels or gradings were not explicitly referenced.

For stage IA melanoma, it was recommended that patients be followed up for 1 year after treatment. For stage IIB melanoma, follow-up was recommended for 5 years. These recommendations were assessed to be a level III (evidence obtained from well-designed non-experimental descriptive studies), grade B (fair evidence to support the use of the procedure) evidence, based on their predefined grading system. However, there was only one study explicitly linked to these recommendations.[63]
In terms of peer review during the guideline development phase, the pre-publication guideline was reviewed in a multidisciplinary meeting, which included lay representatives, and was further reviewed by the BAD executive before publication.[64]

Clarity of presentation
Summary recommendations were clearly set out in boxes for some of the areas covered, with the use of different coloured font for main headings. As this guideline was published in a journal, its presentation was constrained by journal style. The proposed recommendations for follow-up were clearly presented and grouped according to the various stages of disease. This reduced ambiguity in the guideline, as once the skin lesion was properly staged, it was easy to choose an appropriate follow-up plan.

National Comprehensive Cancer Network guidelines, version 2.2019: Cutaneous Melanoma
This is the most recent (2019) version of the cutaneous melanoma guideline produced by the National Comprehensive Cancer Network® (NCCN).[55] The NCCN is a not-for-profit alliance of 31 member cancer centres in the USA.

Scope and purpose
The aim of the NCCN Guidelines is to support the:

> *. . . sequential management decisions and interventions that currently apply to 97 per cent of cancers affecting patients in the United States.*
>
> *Referenced with permission from the NCCN Guidelines®[65] About the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®). National Comprehensive Cancer Network, Inc. 2020. All rights reserved. Accessed August 26, 2020. Available online at www.NCCN.org. NCCN makes no warranties of any kind whatsoever regarding their content, use or application and disclaims any responsibility for their application or use in any way*

The target audience of this review was clinicians, with a separate document available on the website for patients.[65] It aims to incorporate current evidence in response to gaps identified from annual institutional reviews and external requests. There is at least a yearly update of the guidelines to address these questions or gaps in knowledge.

Rigour of development

The development of the NCCN Guidelines was based on a critical assessment of research evidence, clinical expertise and consensus agreement. Similarly, the clinical evidence was categorised based on clinical evidence and consensus among panel members. The NCCN has a four-point category system for grading evidence. These are categories 1, 2A, 2B and 3. Category 1 is a high level of evidence with uniform NCCN consensus, 2A is a lower level of evidence with uniform NCCN consensus, 2B is lower level with some NCCN consensus and 3 is any level of evidence with no NCCN consensus.

The evidence for the follow-up of stage I disease was classified as category 2A:

> *Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.*
> *Referenced with permission from the NCCN Guidelines®[66] NCCN Categories of Evidence and Consensus. National Comprehensive Cancer Network, Inc. 2020. All rights reserved. Accessed August 26, 2020. Available online at www.NCCN.org. NCCN makes no warranties of any kind whatsoever regarding their content, use or application and disclaims any responsibility for their application or use in any way*

The guideline recommended that duration of follow-up should be individualised after the initial 5-year period. In the discussion section of the guideline, the clinical evidence that was considered when making the recommendations was referenced in the text, and mainly pertained to observational and retrospective data.

Clarity of presentation

The first section of the guideline contains the summary of recommendations in an interactive chart format, which facilitates navigation along a pathway for a specific stage of the disease. The ease of navigation facilitates the clarity of the guideline, given the depth of information on treatment and follow-up. However, the links in the interactive chart meant that a reader could go around in circles when trying to identify relevant information. For this reason, the recommendations on follow-up in the guidelines are best accessed by reading the text of the guideline.

European Society of Medical Oncologists clinical practice guidelines for cutaneous melanoma

The European Society of Medical Oncologists (ESMO) produced a guideline on cutaneous melanoma in 2015.[56] ESMO comprises expert members from all over the world, although they are predominantly European. Hence, this guideline is more clinician focused than the NICE guidelines, which consider other audiences.

Scope and purpose

The precise aim of the guideline is not clearly stated in the text. However, ESMO guidelines provide evidence-based recommendations on cancer care.[67] As with the other guidelines appraised, the ESMO guideline focuses on cutaneous melanoma. The ESMO guideline is a short guideline aimed at health professionals.

Rigour of development

There is a very brief methods section describing the ESMO standard operating procedure, which allows for relevant literature to be selected by expert authors, as well as by external expert review. The literature reviews on which the guideline recommendations were based were not extensive. However, there were discussions on the sources of recommendations, with the actual recommendations summarised in a box and graded using an adaptation of the Infectious Diseases Society of America's grading system.[68] This grading system used both levels and grades of evidence. The levels of evidence were from I to V. Evidence from at least one large randomised controlled trial (RCT) was at level I, whereas case reports and expert opinions were level V evidence. The grades of evidence ranged from A ('Strong evidence for efficacy with a substantial clinical benefit, strongly recommended') to E ('Strong evidence against efficacy or for adverse outcome, never recommended') (reproduced with permission from ESMO Guidelines Committee).[69]

Unlike the previously reviewed guidelines, this guideline does not grade the evidence on follow-up for all stages of melanoma. Instead, there is a comment on the variations in the recommendation. There is no specific recommendation for the follow-up of stage I and II melanomas. Instead, the main point raised was the lack of consensus on the duration and frequency of follow-up, as well as the use of imaging in follow-up. Like the other guidelines reviewed in this section, the recommendations on follow-up are based on evidence from cohort studies (judged by the guideline developers to be based on level III evidence). The recommendations derived from this level III evidence are general to all melanoma patients, regardless of the stage of the disease. These recommendations are protection from extended UV exposure, avoidance of artificial UV radiation, and self-examination.

Clarity of presentation

This guideline is considered to be a quick and easy guide to use. Recommendations are shown in tables supported by short sections of text summarising the literature supporting the recommendations. These recommendations are, as previously mentioned, based on summaries of the evidence. In addition, an updated flow chart covering diagnosis and treatment was added in 2016. This update was produced following an electronic update procedure in place for rapid dissemination of significant breakthroughs. There are no explicit target population-based recommendations on follow-up.

American Academy of Dermatology: guidelines of care for the management of primary cutaneous melanoma

Scope and purpose

The 2019 American Academy of Dermatology (AAD) guideline is the most recent version of the 2018 guideline by the AAD for cutaneous melanoma.[57] This guideline addresses the treatment of cutaneous melanoma in children, adolescent and adult populations. As with the other guidelines considered in this chapter, it explores the role of laboratory tests and radiological tests in surveillance, as well as the appropriate duration of follow-up. The mechanism for the ongoing update is unclear, but, on the guideline scripts available on the *Journal of the American Academy of Dermatology* website (www.jaad.org/; accessed 3 October 2019) (published in January 2019), there was a button to click on to check for the latest updates underneath the list of authors.

Rigour of development

A systematic review of available evidence was undertaken and the evidence was graded using the three-point scale unified grading of the Strength of Recommendation Taxonomy.[70] The evidence used for each set of recommendations was discussed and referenced in supporting text. A boxed summary of the evidence was then included. However, for the surveillance of stage I melanoma, expert opinion was used in the guidance, with no explicit link to any studies.

Clarity of presentation

The guideline had few summary boxes, which affected its ease of access. The recommendations that were made, however, were clear and easy to read. The full text of each recommendation was contained in a single box and there was a link to a further box that contained the sources on which the recommendation was based, along with the level of evidence for each recommendation.

Cancer Council Australia Melanoma Guidelines Working Party

As of September 2019, the clinical practice guidance for melanoma from the Cancer Council Australia Melanoma Guidelines Working Party[53] is undergoing revision, with sections being released as they are completed. The current revision began in 2014, using a 'wiki' web-based platform to enable rapid updates, necessitated by the rapid turnover of new evidence, and to allow for sharing of information and ease of contribution among panel/working group members. It also allowed for collaboration with the German Dermatologic Cooperative Oncology Group, which provided access to some of its systematic reviews. Most of the guideline has now been published, with only two questions under development: 'How should melanoma in children be managed?' and 'How should melanocytic tumour of unknown

malignant potential be managed?' Neither of these questions are directly relevant to the aims of the research reported in this health technology assessment.

Scope and purpose

The foreword section of the guideline described a generic intent, covering the target population and benefits:

> *The purpose of evidence-based clinical guidelines for the management of any medical condition is to achieve early diagnosis whenever possible . . .*
>
> *Cancer Council Australia (reproduced with permission from the Cancer Council Australia Melanoma Guidelines Working Party[53])*

Although the relevant criteria stated in the AGREE II tool were present,[52] these were not applied to this guideline directly, but stated as a purpose for any guideline formation.

The questions for each portion of the guideline are clearly stated. When relevant, the target population is identified and the intervention/exposure stated, as well as the expected outcomes. These questions are delineated in boxes for each section and were clear and concise. Similarly, when appropriate, the population of interest in each section is defined using either systematic or non-systematic review evidence.

Rigour of development

The guideline includes a dedicated chapter to the guideline development process, which includes a flow chart detailing the steps used in arriving at the recommendations. This chapter makes references to protocols used for the systematic review of the evidence, the grading of evidence, formulation of final drafts and content review. The plans for continued updates are incorporated into the design of the guideline.

For individual recommendations, there are explicit links to the sources of evidence, as well as due considerations of benefits, side effects and risks. It is worth noting that, in including systematic reviews from the German Dermatologic Cooperative Oncology Group, these guidelines were initially assessed for quality using the AGREE II checklist by the Australian working group.

Specific recommendations for surveillance of stage I melanoma were all grade C evidence, meaning that the 'Body of evidence provides some support for the recommendation(s), but care should be taken in its application'[53] (reproduced with permission from the Cancer Council Australia Melanoma Guidelines Working Party[53]). The evidence used in these recommendations was mostly derived from case series articles.

Clarity of presentation

The description of the different management options is well presented for population and practice contexts. These are grouped in header boxes for each set of recommendations. Although recommendations for each stage of melanoma are reported, this is not clearly flagged in the headers for each section. For instance, self-examination is a boxed recommendation for all stages. The next boxed recommendation is history and physical examination by a physician for stages I–III during follow-up. As with the other guidelines reviewed, radiological follow-up is not recommended for stages I–IIB.

Dutch Working Group on Melanoma

This is the 2012 national guideline produced by the Dutch Working Group on Melanoma.[58] The guideline addresses 19 questions. The appendices showing the list of questions covered are not available in English. However, a systematic review was conducted for three of the questions in the entire guideline; the remaining 16 questions had their recommendations based on studies put forward by guideline committee members.

A further revision was completed between 2014 and 2016 to answer three key questions: 'The role of ^{18}F-FDG-PET[fluorodeoxyglucose PET]/CT at diagnosis',[58] 'the role of ^{18}F-FDG-PET/CT in follow-up'[58]

and 'the role of the sentinel node biopsy'[58] (reproduced with permission from the Dutch Working Group on Melanoma[58]). This additional review was carried out by the Nederlandse Vereniging voor Nucleaire Geneeskunde (Dutch Society for Nuclear Medicine) and the Nederlandse Vereniging voor Pathologie (Dutch Association for Pathology).

Scope and purpose

The objectives of the guideline are clearly stated: 'The intention of the document is to be a guideline for daily practice in prevention, diagnosis, treatment and follow-up of patients with a skin melanoma'[58] (reproduced with permission from the Dutch Working Group on Melanoma[58]). The guideline is intended to cover all stages of the disease and is targeted at a health-based audience ranging from clinical staff to social workers. As with the other guidelines, this guideline covers all patient groups with melanoma.

Rigour of development

Akin to the previously discussed guidelines, the composition of the group, the methodology and the grading of evidence are all contained in the appendices of the guideline. It describes how guideline recommendations were arrived at using evidence-based and consensus agreements. This process of evidence synthesis is similar to that used by other guidelines considered in this chapter, in that expert opinion, or consensus agreements, were also incorporated. The recommendations for the follow-up of stage I melanoma are limited and are based on low-quality evidence.

Most of the recommendations on follow-up were based on other Dutch guidelines ('Cancer Survivorship Care' and cancer rehabilitation,[71] and detection of psychosocial distress[72]). These earlier guidelines have not been assessed for the rigour of the development. As the appendices available online are in Dutch only, critical assessment of the grading process and the entire guideline formulation process has not been performed. Nonetheless, the follow-up recommendations are linked to some evidence, although, in some cases, it is just one source of evidence.

Clarity of presentation

The English-language version of the guideline is available as a PDF only,[58] but there is a comprehensive index showing the various topics and subtopics covered. In-text, summary recommendations were tabulated following discussions on the evidence base and consensus agreements. The grading of these recommendations is included in the summary box. The language barrier made it difficult to assess the functionality of the website and additional components of the guideline. However, the English-language version makes a distinction between follow-up and aftercare in its recommendations. These are explained in simple terms and were judged to be easy to follow for all types of clinical or social care workers.

German Guideline Program in Oncology

The *S3-Guideline 'Diagnosis, Therapy and Follow-up of Melanoma' – Short Version* is the most recent (2013) English-language version of the short version of the German guideline for cutaneous melanoma.[59,73] However, there have been updates in 2015 and in 2016/17, owing to the rapid developments in the field. These updated versions are available online in German only.[73] The ongoing stated plan will be to have a live system that allows for regular updates. A new version of this guideline is planned by the end of 2019.

Scope and purpose

The focus of the guideline was on cutaneous melanoma diagnosis, management and follow-up. The guideline was aimed at clinical practitioners in the field of medical oncology, providing 'an accepted, evidence-based decision-making aid for the selection and performance of suitable measures for diagnostics, therapy and follow-up of cutaneous melanoma.'[59] Although the guideline has a clear clinician focus, there is an extended version of this guideline, as well as guidelines for patients.[73]

Rigour of development

Although there is a link to the methodology used, it was difficult to assess the rigour of development because this material is not presented in English. The modified Scottish Intercollegiate Guidelines Network[74] classification of evidence was combined with an agreed grading system developed by the guideline authors to assess the evidence used in the development of the guideline. A small portion of the recommendations are based on what was considered by the guideline developers to be sound clinical practice, rather than based on scientific evidence; this includes recommendations on the frequency of follow-up for stage I patients. It is recommended that, for all stages of melanoma, there should be a 10-year risk-adapted follow-up. This recommendation is based on evidence rated to be of level 1b, with a grade of B (which means a recommendation of 'should' be followed). For each recommendation, there are explicit links to the sources of evidence used to develop that recommendation.

Clarity of presentation

This guide is well presented, with tables and charts delineating the guideline recommendations. There is limited text in between recommendations, making it less cumbersome to locate the relevant sections. The recommendations are phrased in an unambigiuous way, making it easy to read and follow. It is also easy to understand the rational behind the recommendations by reading the text accompanying each recommendation.

Swiss Cancer League

The updated Swiss guidelines 2016 for the treatment and follow-up of cutaneous melanoma[60] are an update of a 2006 guideline and were initiated as a result of advances in diagnostic capabilities and treatment options.[60]

Scope and purpose

The guideline aims to provide 'a reasonably practical guide for all physicians (general practitioners, dermatologists, surgeons, oncologists, and others) who encounter cutaneous melanoma in their daily work'.[60] Hence, the target audience is clinicians. This guideline also focuses on cutaneous melanoma. The desired outcome of the update was to ensure adequate treatment of this condition among Swiss patients.

Rigour of development

There was no mention of the methods used in gathering the evidence. Available evidence was graded using the 'level of evidence' classification.[75] Using this classification system, the evidence level of IV (historical cohort or case–control studies) was given to the available evidence for follow-up length of stage I melanoma. This is because these data were reported to be 'historical' and 'dated'. Nonetheless, a 10-year follow-up is recommended. Links to supporting studies are placed in the body of the discussions, but are not explicitly linked to each recommendation.

Clarity of presentation

This guideline is a short review article. As with the other guidelines, the recommendations for follow-up are placed in a table and are easy to understand. The follow-up guidance is clearly differentiated by tumour stage.

Brazilian guidelines for diagnosis, treatment and follow-up of primary cutaneous melanoma – Part II

This guideline, from 2016, is a follow-up to the 2002 Brazilian guideline and is the second part of the full guideline.[61] There was a need for the update as a result of recent developments in diagnosis and treatment. This update covered 10 questions; five of these questions are covered in this second part of the guideline. The five questions were on follow-up for stage 0 and I melanoma, the role of body mapping in follow-up, the benefit of sentinel lymph node in primary melanoma and the benefits of preventative excisions of acral naevi and giant congenital naevi.

Scope and purpose

The aim of this part of the guideline is:

> *intended for diagnostic and therapeutic approach and follow-up of patients with suspected or confirmed diagnoses of primary [cutaneous melanoma] (PCM) with no clinical or histological evidence of metastatic disease (stages 0, I and II).*
>
> *Reproduced with permission from Castro* et al.[61] *© 2016 by Anais Brasileiros de Dermatologia. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited*

Rigour of development

A systematic review was employed to synthesise the evidence and the selected studies were graded based on the level of evidence on a four-point scale from A to D:

- *A: experimental or observational studies of higher consistency*
- *B: experimental or observational studies of lower consistency*
- *C: case reports/uncontrolled studies*
- *D: opinion without critical evaluation, based on consensus, physiological studies or animal models.*

Reproduced with permission from Castro et al.[61] *© 2016 by Anais Brasileiros de Dermatologia. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited*

Recommendations, other than those relating to duration and frequency of follow-up, are not grouped according to disease stage. A 10-year follow-up period is recommend based on grade D level of evidence (i.e. opinion without critical evaluation, based on consensus, physiological studies or animal models). There is no explicit link to the source of this recommendation.

Clarity of presentation

The recommendations for initial diagnosis and those for follow-up are lumped together. However, the guideline is strictly for non-metastatic cutaneous cancer, stages 0–II, and uses a question-and-answer format for each section. For instance, one of the questions in the article is 'How should stages 0 and I primary cutaneous melanoma patients be followed?'[61] [reproduced with permission from Castro *et al.*[61] © 2016 by Anais Brasileiros de Dermatologia. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and production in any medium provided the original work is properly cited], and this question makes up the section on the follow-up recommendation for stages 0–II cutaneous melanoma. Hence a thorough read of each section would be required to clearly unpick the recommendations. This approach also creates some ambiguity in interpretation of the guideline from a clinical perspective, and there is not a clear link between some recommendations and the various stages of cutaneous melanoma. This is examplified by the recommendations on the role of cutaneous mapping, which are based on risk stratification only.

Summary

This chapter has considered the recommendations on surveillance and follow-up of patients made in clinical guidelines throughout the world. In total, 10 guidelines were considered from eight countries, which were published between 2011 and 2019. The follow-up recommendations for stage I melanoma varied between the guidelines in terms of intensity and duration of follow-up, as well as what tests

were recommended and who should perform the follow-up. Comparing recommendations form national bodies, some countries' recommendations are less intensive for stage IA (e.g. Dutch Working Group on Melanoma 2013[58]) than the NICE recommendations[16,54] for the same stage, whereas other countries' recommendations are more intensive (e.g. German Guideline Program in Oncology 2013[59]) than the NICE recommendation for the same stage.[16] In many respects, these differences cannot be explained by differences in underlying risks. Nor can they be fully explained by differences in the methodologies that the different guidelines adopted, although there were some differences in methodology between guidelines. There were also some variations in evidence on which the recommendations were based and in the grading of available evidence. However, clearly defined questions and robust methodologies were employed for guideline development for most guidelines. A common thread for the recommendations made is that the value and strength of the recommendations were low. The limited data on which recommendations were based may well have contributed to the variations in guideline recommendations.

Arguably, even though the methdologies adopted by the different guideline developers were generally strong, recommendations have had to be made on very limited evidence. It is this unexplained variation in recommendations and the underlying evidence gap that guideline developers have faced that motivate the work in the following chapters.

Chapter 3 Systematic review to identify different surveillance and follow-up strategies for stage I melanoma patients following surgical excision

Brief overview

Surveillance strategies vary in a number of ways: by duration and frequency of contact with patients, and in terms of which practitioner sees patients, and in what type of diagnostic and prognostic tools are used. In a systematic review of surveillance strategies, all the countries that provided data on surveillance had programmes that followed patients for 5 years after treatment and recommended between one and six visits per year, in addition to recommended self-examination.[25] Self-examination is important because many (if not most) melanoma recurrences are detected by patients themselves.[76] As outlined in *Chapter 2*, not all countries use diagnostic imaging in surveillance visits, but many use sonography; radiography of the regional nodal basin, chest or abdomen; clinical photography; or positron emission tomography (PET), CT or magnetic resonance imaging (MRI). Some also assess a patient's blood count and liver function.

The NICE guideline for melanoma[77] recommends that, after stage IA, patients are seen by a clinician between two and four times in the first year after completion of treatment, and then discharged. After stages IB to IIB melanoma, or stage IIC melanoma with a negative SLNB, the guideline recommends that patients are followed up every 3 months for 3 years and then every 6 months for the next 2 years, after which they can be discharged. No imaging or blood tests are recommended during follow-up for either of these groups.

As described in *Chapter 2*, there is little consensus about the most effective and cost-effective way to follow up patients who have been treated for melanoma. Furthermore, the evidence base for the different strategies adopted is unclear. Previous studies suggest that existing guidance, which includes variation in frequency and duration of patient contact, as well as in recommended diagnostic and prognostic tools, is based on anecdotal evidence or retrospective assessment of historical cohorts.[25,76]

To clarify what evidence there is to support any surveillance for stage I melanoma, a high-quality systematic review is needed. This systematic review would be used to gather and synthesise the most robust evidence about all elements of surveillance strategies for melanoma.

Research aim

The aim of this systematic review was to identify variations in strategies for surveillance and follow-up after surgical excision of AJCC stage I primary cutaneous melanomas in adults and to assess the relative effectiveness on clinical and oncological outcomes, including recurrences, metastases and survival.

Methods

This review adheres to the guidelines for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement to ensure transparency of the process.[78] A protocol for the whole project of which this review is part is published on PROSPERO (CRD42018086784).[79]

Search strategy

The search strategy was designed by an experienced information specialist in collaboration with the project team. The search was designed in MEDLINE [via Ovid® (Wolters Kluwer, Alphen aan den Rijn, the Netherlands)] according to the following main concepts: [melanoma] AND [surveillance OR screening]. Database-specific thesaurus headings were used, together with title and abstract keywords. The strategy was translated to other databases (*Box 2*), altering thesaurus headings and search syntax as appropriate. The databases listed in *Box 2* were searched during the first week of May 2018 and the search was updated on 2 July 2019.

The search was then limited to studies published from 2011 onwards, the search date of a previously published systematic review of surveillance strategies for melanoma.[25] There were no restrictions according to language or publication status. The search strategy used in MEDLINE can be found in *Appendix 1*.

We did not update the systematic review authored by Cromwell *et al.*;[25] instead, we used it to identify publications prior to 2011. To reduce the screening burden of systematically searching from database inception, we screened all studies included in this review.[25] Full references of the review were also screened, in addition to the results of the systematic search limited to studies published from 2011 onwards.

A grey literature search plan was developed to complement this search by exploring (1) grey literature databases, (2) targeted websites and (3) reference leads from (1) and (2), with a focused attempt to locate international and national guidelines. By checking the references of international guidelines and the studies cited in a review by Cromwell *et al.*[25] extended out search to before the 2011 limit described above. The sources described in *Box 3* were searched between 20 July 2018 and 10 September 2018.

Titles and abstracts of search results were imported into EndNote [Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA] and deduplicated.

BOX 2 Databases searched for surveillance review

- MEDLINE (via Ovid), 1946 to June week 3 2019.
- EMBASE™ (Elsevier, Amsterdam, the Netherlands; via Ovid), 1980 to week 26 2019.
- CENTRAL [the Cochrane Library via Wiley Online Library (John Wiley & Sons, Inc., Hoboken, NJ, USA)] issue 6, 2019.
- Cochrane Database of Systematic Reviews (the Cochrane Library via Wiley Online Library), issue 6, 2019.
- Database of Abstracts of Reviews of Effects (the Cochrane Library via Wiley Online Library), issue 2, 2015.
- Health Technology Assessment Database (the Cochrane Library via Wiley Online Library), issue 2, 2018.
- NHS-EED (the Cochrane Library via Wiley Online Library) issue 2, 2015.
- CINAHL [via EBSCO*host* (EBSCO Information Services, Ipswich, MA, USA)], 1982 to June 2019.
- Science Citation Index [Clarivate Analytics; via the Web of Science™ (Clarivate Analytics)], 1970 to June 2019.
- Conference Proceedings Citation Index – Science (Clarivate Analytics; via the Web of Science), 1990 to June 2019.

CENTRAL, Cochrane Central Register of Controlled Trials; CINAHL, Cumulative Index to Nursing and Allied Health Literature; NHS EED, NHS Economic Evaluation Database.

BOX 3 Sources searched for grey literature in surveillance review

- OpenGrey (www.opengrey.eu/; accessed 17 April 2019, includes SIGLE, EAGLE, GreyNet).
- ClinicalTrials.gov (accessed 17 April 2019).
- Cancer Research UK (www.cancerresearchuk.org/about-cancer/find-a-clinical-trial; accessed 17 April 2019).
- BAD (http://www.bad.org.uk/; accessed 17 April 2019).
- British Skin Foundation (www.britishskinfoundation.org.uk/; accessed 17 April 2019).

EAGLE, European Association for Grey Literature Exploitation; SIGLE, System for Information on Grey Literature in Europe.

Inclusion and exclusion criteria

We based inclusion and exclusion criteria on the population, intervention, comparator, outcomes, timing and setting (PICOTS) formula, as outlined in the following sections.[80]

Population

Adults aged ≥ 18 years treated for AJCC (eighth edition) stage I cutaneous melanoma [stage IA (≤ 0.8 mm thick without ulceration) or stage IB (< 0.8 mm thick, or < 1 mm thick and ulcerated skin)].[19]

Or:

Adults aged ≥ 18 years treated for AJCC (seventh edition) stage I cutaneous melanoma [stage IA (T1a ≤ 1 mm thick) or stage IB (T1b with ulceration or mitoses ≤ 1 mm thick, or T2a 1.01 to 2.00 mm thick and no ulceration)].[20]

Non-randomised studies reporting patients with varying stages of cutaneous melanoma were included if ≥ 80% cases were at stage I, as we expected data to be sparse and did not want to omit potentially relevant evidence. Studies that did not specify a patient population were initially included pending confirmation from study authors. However, none of these met the inclusion criteria; therefore all were excluded. Studies reporting the Breslow depth, for patients with tumours of ≤ 2 mm, were included if there were no data on AJCC stage. Studies that included only patients with stage II–IV melanoma were excluded.

Intervention

We included studies that had any surveillance or follow-up strategies aiming to identify further primary melanoma, local recurrence or in-transit, regional or distant metastases. These were not limited by setting or by the type of clinician undertaking the follow-up. They could include clinical evaluation, patient education, skin self-examination (SSE) or radiological examination at any frequency. We excluded studies that focused on treatment of melanoma rather than surveillance.

Comparator

Studies with any comparator that allowed for the assessment of relative clinical effectiveness were eligible for inclusion (i.e. no surveillance or an alternative strategy).

Outcomes

The following were the outcomes of interest: overall survival, progression-free or recurrence-free survival, melanoma-specific survival, detection of recurrence as a new primary tumour, in-transit metastases and locoregional metastases. This could be presented as dichotomous or time-to-event data, such as percentages, hazard ratios (HRs), risk ratios (RRs) or Kaplan–Meier plots. No restrictions were placed on how outcomes were determined or confirmed (e.g. through biopsy, histology or imaging); all study-defined outcomes were allowed.

Timing
The timing for onset of surveillance strategies was restricted to patients who were post resection of any primary cutaneous melanoma tumours. The duration of surveillance (follow-up) was determined by individual studies and interpreted accordingly.

Setting
All studies were eligible for inclusion, regardless of whether the study was conducted in primary, secondary or tertiary care. No restrictions were applied to countries of origin conducting the primary research, although the relevance to current or future UK practice was assessed.

Study designs
We included RCTs and non-randomised comparative studies, for example quasi-experimental and comparative retrospective or prospective observational studies. We also looked at guidelines that recommended strategies for surveillance of stage I melanoma so that we could search their references for eligible studies. We excluded potentially underpowered non-RCT studies (arbitrarily defined as having < 100 patients) because they are at risk of selective reporting bias and publication bias, and they lead to small-study effects with imprecision.

To minimise selection bias in non-randomised study designs, we included studies that used statistical adjustment for baseline case mix using multivariate analyses, provided that the study had at least 80% stage I patients. We expected to see variables such as age, sex, ethnicity, tumour stage and grade, histology or performance status as adjustment variables; we excluded comparative observational studies if they did not adjust for at least two of these variables.

Data collection

Selection of studies
Selection of studies that met the inclusion criteria was conducted in two stages. In the first instance, studies were exported from the EndNote library and into Rayyan (Qatar Computing Research Institute, Doha, Qatar), a web-based tool designed to aid screening and selection of studies for systematic reviews.[81] For consistency and accuracy, two sets of two reviewers initially piloted the screening process. This was done by assessing 10% of the titles and abstracts, along with some full text studies, against the prespecified inclusion and exclusion criteria. Disagreements at this stage were resolved by either discussion between the reviewers or arbitration with another member of the study team. In the second stage, studies that appeared to meet the inclusion and exclusion criteria were imported into EndNote and full-text papers obtained. When full texts were not readily available, we accessed articles via interlibrary loans. Two reviewers independently evaluated these articles and made their selection in accordance with the eligibility criteria.

Data extraction
A data extraction form was created in Microsoft Word (Microsoft Corporation, Redmond, WA, USA) in accordance with Cochrane guidelines[82] and piloted on one study prior to use. After necessary adjustments, one reviewer undertook the data extraction of the included articles. The completed extraction form was checked for accuracy, completeness and consistency by a second reviewer. When stage I data were grouped with other stages of disease in an included study, we contacted the study authors, which led to us obtaining the relevant data for patients with stage I disease. We also contacted authors to obtain missing data or to clarify uncertainties. The following domains were extracted: country of origin, patient characteristics, study objectives, study design, tumour characteristics, follow-up regimens, analysis methods, risk of bias, outcomes and conclusions. An example of the data extraction form can be found in *Report Supplementary Material 1*.

When possible, all data extracted were those relevant to an intention-to-treat analysis, in which participants were analysed in the groups to which they were assigned. The time points at which outcomes were collected and reported were recorded.

Guidelines that recommended strategies for surveillance of stage I melanoma did not form part of the systematic review. However, to provide context, they were summarised and their conclusions are presented in *Summary of review of different surveillance strategies*.

Risk-of-bias assessment in included studies

We used the Cochrane Collaboration's 'Risk-of-bias' tool, RoB 2.0, which uses signalling questions to assess risk-of-bias judgements.[83] The tool examines five domains graded as being at a low risk-of-bias, of some concern or at a high risk of bias, from which an overall risk-of-bias judgement can be made. The domains considered are biases resulting from the randomisation process, deviations from the intended intervention, missing outcome data, measurement of outcomes, and selection of reported results.

Measures of effect

We planned to extract the following reported measures of effects of surveillance:

- dichotomous or binary data, odds ratios (ORs) or RRs or percentages
- time-to-event data, HRs or Kaplan–Meier plots.

Confidence intervals for all estimates missing data

We set out to report the number (per cent) of missing data for all variables/outcomes. We did not impute missing outcome data for any of our specified outcomes.

Data analysis

We summarised the characteristics of the surveillance strategies from included studies and guidelines in a table and provided a narrative summary of these.

We planned to conduct random-effects meta-analyses to pool data for each outcome in the review. In the absence of sufficiently robust or similar studies for a meta-analysis, we carried out a summary of studies, rather than a more formal narrative synthesis, owing to a lack of evidence.

Quality of the evidence using the GRADE approach

The GRADEpro tool was used to assess the overall certainty in the body of evidence for key outcomes.[84] The GRADE approach uses the risk of bias of individual studies, along with characteristics such as the imprecision and inconsistency of their results, to produce an overall estimate in terms of whether there is high, moderate, low or very low confidence that the systematic review estimates the true effect. These data are presented as a summary of findings table (see *Table 6*).

Results

Number of studies identified

The searches retrieved 10,723 citations in total; 10,592 were retrieved from the electronic databases, 104 from the published systematic review[25] and 27 from the grey literature and guidelines search (*Figure 3*). After deduplication, 6205 references remained. Following a title and abstract sift by one reviewer and two clinicians, 6134 references were excluded, resulting in 33 citations of articles and conference abstracts for full-text assessment.

Following this, we excluded 31 articles. The reasons for excluding the full-text papers were as follows: < 80% of participants at stage I or wrong stage (58%), and studies identified as prognostic studies (26%) diagnostic studies (10%) or prevention studies (6%). In addition, all included studies from the Cromwell *et al.*[25] systematic review were excluded because they were not surveillance strategies among individuals post resection of a stage I melanoma, and most included studies had a non-comparative design.[25] A list of excluded full-text articles retrieved from the literature search, with reasons for exclusion, is provided in *Appendix 2*.

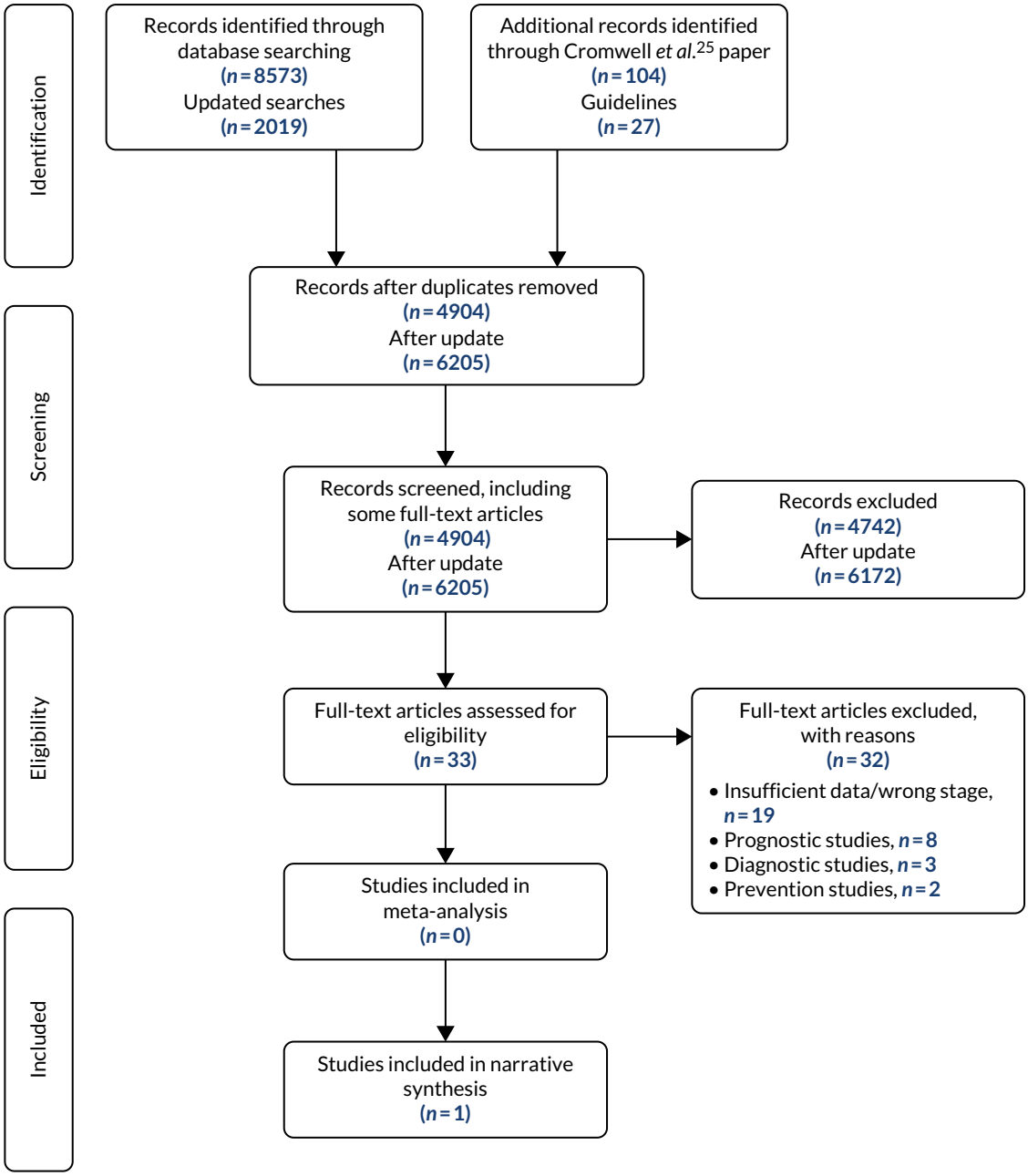


FIGURE 3 The PRISMA flow diagram for the review of surveillance strategies' clinical effectiveness.

Authors of relevant studies presenting aggregated data were contacted to provide data stratified by stage. Correspondence was received from Robinson *et al.*[85] and Damude *et al.*;[23] however, only Robinson *et al.*[85] provided data that fulfilled our inclusion criteria.

Characteristics of included studies

One RCT met our inclusion criteria[85,86] after provision of further data by the authors. Robinson *et al.*[85] assessed the frequency of SSE by patient–partner dyads. The study was conducted in the USA.[85] The study included a total of 494 participants with a mean age of 55 years [standard deviation (SD) ±10 years, range 18 to > 70 years). Descriptive information of the study is presented in *Table 4*.

Patients with stage 0–IIB melanoma participated in the trial from June 2011 to April 2015. This was a continuation of the trial initially reported by Turrisi *et al.*[87] Patients in the intervention arm received a structured skills training intervention, whereas patients in the control arm received customary care.

TABLE 4 Description of included studies for the review of guideline surveillance strategies

Study (first author and year of publication)	Design	Country, setting	Intervention strategy	Control strategy	Number of patients at baseline and stage	Patient median age in years) (range)	Patients at stage I, *n* (%)	Duration of follow-up	Outcomes
Robinson *et al.* 2016[85]	RCT, from June 2011 to April 2013	• USA • Northwestern University Feinberg School of Medicine, Chicago, IL, USA • Setting: hospital ambulatory care area and clinical offices	Skin self-examination with a partner after a structured skills training intervention (dyads)	Customary care of patients with partners (dyads)	• Interventions: ○ Workbook, *n* = 159 ○ In person, *n* = 165 ○ Tablet, *n* = 71 • Control: *n* = 99 • Stages 0–IIB	• Interventions: range 18 to > 70 • Control: range 18 to > 70	• Interventions: ○ [a]Stage IA, *n* = 114 (29%) ○ [a]Stage IB, *n* = 89 (23%) • Control: ○ [a]Stage IA, *n* = 31 (31%) ○ [a]Stage IB *n* = 24 (24%)	• Baseline • 4 months • 12 months • 24 months	• Primary: SSE frequency • Secondary: ○ New or recurrent melanoma, or ○ In-transit metastases detected by dyads or physician • Author provided: ○ New or recurrent melanoma, or ○ In-transit metastases detected by dyads or physician among those post resection of a stage I melanoma

a Data by stage provided by authors.

A total of 494 patients and their partners were randomised to one of four groups. Three of the dyad groups received a structured skills training intervention in SSE, either in person, from a written workbook or via a tablet. The fourth group served as control and received treatment as usual[87] and customary education.[85] Patients were seen by a dermatologist every 4 months. The primary outcome was frequency of SSE by patient–partner dyads, and the follow-up period and end point of the trial was 24 months. The secondary outcome was detection of a new or recurrent melanoma by the dyad or physician.

Patients at stages 0–IIB receiving the intervention had significantly increased SSEs with their partners at 4 months, compared with controls [mean difference 1.57, 95% confidence interval (CI) 1.29 to 1.85]. Mean differences at 12 and 24 months were lower (mean difference 0.72, 95% CI 0.39 to 1.06, and mean difference 0.94, 95% CI 0.58 to 1.30, respectively). Overall, data reported for stages 0–IIB showed that the intervention was successful in increasing SSE by patient–partner dyads, compared with controls at 24 months (mean difference 0.94, 95% CI 0.58 to 1.30; $p < 0.001$). We contacted the authors for data by disease stage, which were provided by them for our analyses.

The individuals undertaking surevillance (through SSE) of the outcome of interest to our review (detection of a new primary tumour, recurrent melanoma or metastases) were both the dyad and physician, with the site of surveillance being either the home or the care setting of the dermatologist (predominantly a secondary care setting). The interval timing of the review appointments with the dermatologist were 4 months; however, for surveillance by the dyad, the interval timing was dependent on their own timeline of use of SSE. The duration of follow-up for the surveillance in the trial, and thus by dermatologists, was 2 years. However, the SSE by the dyads was intended to last for life. There was no routine imaging involved in the surveillance strategy; rather, the strategy was based on a structured skills training intervention on how to self-identify plausible new or recurrent melanoma. Given that the study reported that SSE increased, it has been assumed that this surveillance stratgey is well accepted by melanoma patients. There is no large burden on health-care providers due to surveillance by SSE.

Risk-of-bias assessment of included studies

It was judged that there were 'some concerns' regarding the risk of bias in the study by Robinson *et al.*[85] Two allied papers were used to identify data pertinent to the risk-of-bias assessment.[86,87] The results are discussed in the following sections, and the judgements made using the Cochrane risk-of-bias tool, RoB 2.0,[83] are shown in *Table 5*.

TABLE 5 Risk of bias for the systematic review of surveillance strategies

	Risk of bias					
Study (first author and year of publication)	**Randomisation process[a]**	**Deviation from intended interventions[b]**	**Missing outcome data[c]**	**Measurement of outcomes[d]**	**Selection of reported results[e]**	**Overall bias**
Robinson *et al.*[85] 2016	Some concerns	Low	Some concerns	Low	Low	Some concerns

a Randomisation: rated as 'some concerns'. Two-stage randomisation at different time points, but baseline imbalances in patients assigned to each arm and additional intervention group added part-way through trial.[86,87]
b Deviations: patients were analysed in the groups to which they were randomised. Although some SSEs were conducted alone (not as a dyad), there were no statistically significant differences between intervention and controls on the mean number of SSEs performed alone.
c Missing outcome data: rated as 'some concerns' because of attrition. However, there were no statistically significant differences between patients completing the 24-month assessment and those lost to attrition, in demographics, original melanoma diagnosis, or time since diagnosis.
d Measurement of outcomes: rated as being low risk because assessors were blind to the intervention arm participants were in. Participants were informed at subsequent visits to prevent discussion of the intervention during skin examination by the dermatologist.
e Selection of reported results: rated as being low risk because, although there were multiple outcome measures over time, if a new lesion was identified, then a biopsy would be undertaken to confirm. The number of recurrences or new primaries was a secondary outcome.

Bias from randomisation process

Reviewers considered the randomisation process as giving rise to some concerns.[85] There were baseline imbalances in the number of participants assigned to each of the intervention arms. The first 150 pairs were randomised to one of the three groups (workbook, in person or control), and the remaining 344 pairs were randomised to one of four groups (workbook, in person, tablet or control).[85]

Bias from deviations of intended interventions

The study was judged to be at a low risk of bias. No deviations from the intended interventions were reported. However, an additional intervention (using a tablet computer) was added while recruitment was ongoing.

Bias due to missing outcome data

The study was judged to give rise to some concerns regarding incomplete outcome data.[85] This related to high and varying levels of attrition between trial arms. Reasons for non-participation were reported and there did not appear to be any notable differences between those completing the 24 months' assessment and those lost to attrition by demographic characteristics, initial melanoma diagnoses or time since diagnosis, as reported in the study results.[85] Reasons for not attending follow-up were reported as 'not learning anything new,' 'no change in pigmented lesion' and 'too far to travel'. Attrition reduces the ability of the study to detect a difference, should one exist.

Bias in selection of reported results

The study was assessed as having a low risk of selective reporting because the trial protocol was available as a trial registration on ClinicalTrials.gov and as a peer-reviewed manuscript. All prespecified outcomes were reported in the results.

Publication bias

We were unable to assess whether or not there was any publication bias because there was only a single study was included. However, we carried out comprehensive searches to reduce the risk of missing relevant studies.

Assessment of clinical effectiveness

Although three publications (including a conference abstract) reporting two RCTs were eligible for this systematic review,[23,85,88] they did not report data for stage I patients separately. Authors were contacted to provide further data, which we obtained for stage I melanoma patients from the authors of the study by Robinson *et al.*[85]

We were unable to conduct a meta-analysis as only one RCT met the inclusion criteria. We were unable to assess reporting biases using funnel plots or to conduct any subgroup or sensitivity analyses. Thus, data on effectiveness and safety from the included RCT were tabulated and presented in the summary of findings table (*Table 6*) and narratively summarised. For outcomes of interest, we have calculated and reported the magnitude of effect.

The primary outcome of the study by Robinson *et al.*[85] was the frequency of SSE by patient–partner dyads. The secondary outcome (among those post resection of stage 0–IIB melanoma) was detection of a new or recurrent melanoma by the dyad or physician. For those post resection of stage I melanoma, the population of interest in this review, data were provided by the study's author. New primaries, recurrences or metastases were detected in 49 out of 258 (19%) patients with stage IA or IB melanomas post resection of a primary melanoma followed up for up to 24 months. Data were not split by whether the disease was a new primary or recurrence, and recurrences could be at different stages from the original primaries. The types of melanomas identified were melanoma in situ, stage IA, superficial spreading, lentigo maligna and melanomas of ≥ 0.1 mm.

TABLE 6 Summary of findings: systematic review to identify different surveillance and follow-up strategies for stage I melanoma patients following surgical excision

Outcomes	Relative effect (95% CI)	Number of participants (*n* studies)	Quality of the evidence (GRADE)	Comments
• New primary, recurrence or in-transit metastases • Duration of follow-up: 24 months	RR 0.75 (0.43 to 1.31)	258 (1 study)	⊕⊕⊖⊖ Low[12]	• We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the study or from any reliable external source • Study included data aggregated by disease stage, but authors provided diaggregated stage I data (258/494) for analysis. New primaries, recurrences or in-transit metastases were detected in 49 out of 258 (19%) patients with stage IA or IB: 36 out of 203 (18%) in the intervention group, compared with 13 out of 55 (24%) in the control arm • Inclusion criteria included only studies that were randomised or that used statistical adjustment for baseline case mix using multivariate analyses, provided the study had at least 80% of stage I patients (to minimise selection bias in non-randomised study designs)

Overall survival, progression or recurrence-free survival and detection of new primary melanoma, recurrence or metatstases were not reported for stage I disease by any study

One ongoing multicentre trial also met the inclusion criteria.[23] The trial was conducted in six hospitals comparing a reduced follow-up schedule with the conventional schedule in 180 patients with stage IB to IIC cutaneous melanomas over a period of 1 year. Results at 3 years are available in abstract form only and we did not receive a breakdown by stage I on request. The primary end point of this trial was patient well-being; secondary end points were development of recurrence, second primary melanoma or metastases

Low quality:[a,b] Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

a Downgraded by one level for imprecision, sparse data and a low event rate in participants with stage I disease under the old AJCC seventh edition staging classification system

b Downgraded by one level for concerns regarding risk of bias.

Notes

Patient or population: adults (aged ≥ 18 years) treated:

- I for AJCC (eighth edition) stage I: cutaneous melanoma [stage IA (≤ 0.8 mm thick without ulceration) or stage IB (< 0.8 mm thick, or < 1mm thick and ulcerated skin)] or
- I for AJCC (seventh edition) stage I cutaneous melanoma [stage IA (T1a ≤ 1 mm thick) or stage IB (T1b with ulceration or mitoses ≤ 1 mm thick, or T2a 1.01 to 2.00 mm thick and no ulceration)].

Settings: any.
Surveillance strategy: SSE.
Comparison: any comparator that allowed for the assessment of relative clinical effectiveness (e.g. no surveillance or an alternative strategy).

There was no evidence of a difference between intervention and control arms in the proportion of patients with stage IA or IB melanomas in which a new primary or recurrence was detected in this subset (RR 0.75, 95% CI 0.43 to 1.31). However, imprecision affects our certainty of this finding and more evidence is needed to draw any conclusions.

The authors[85] concluded that patients with melanoma and their partners reliably performed SSE after participating in a structured skills training programme lasting approximately 30 minutes, with reinforcement every 4 months by the study dermatologist. No conclusions were drawn by the study authors[85] about how new primaries or recurrences were detected.

Discussion

Summary of review of different surveillance strategies

This review sought evidence about the relative effectiveness of surveillance and follow-up strategies to identify melanoma recurrence, new primary tumours and metastases in stage I cutaneous melanoma patients following surgical excision of the primary tumour. Only two RCTs (reported in three papers)[23,85,88] were eligible and we could obtain data on stage I patients from only one of them.[85] This study suggested that an educational intervention for patients and their partners improved self-identification of new primaries, regardless of whether it was delivered in person, through a workbook or via a tablet. However, among the subset of author-provided data on patients post resection of a stage IA or IB melanoma, there was no evidence of a difference in detection of a new primary tumour, recurrence or metastases between those undergoing SSE surveillance and those receiving usual follow-up.

This evidence is of low certainty according to GRADE because of the small number of studies and limited number of available relevant outcome data;[86] it is probable that the results of this review would change with the addition of new eligible studies. The certainty of the evidence was downgraded for imprecision, sparse data and a low event rate and for concerns regarding risk of bias. At present, evidence is based on just a single study[85] ($n = 258$ participants) and the evidence is incomplete and offers only internal validity as it was set in the USA. Only one of the prespecified outcomes (new primary or tumour recurrence) in our review was reported, meaning that there are complete gaps in the evidence in this area in terms of overall survival, progression-/recurrence-free survival and detection of recurrence (see *Table 6*).

As stage I is the most common stage at melanoma diagnosis, it is critical to understand the most effective method of surveillance following treatment. This review demonstrates that current evidence is insufficient and uncertain, so further robust RCTs are required, measuring recurrence and metastases, in addition to overall survival, as outcomes, to establish the most effective surveillance strategy. No assessment of surveillance strategies among those post resection of a stage I melanoma using clinical review, imaging, or diagnostic biopsy as the main component of the strategy were identified.

Strengths and limitations

To our knowledge, this is the first systematic review of surveillance or follow-up strategies for AJCC stage I melanoma. The review followed procedures set out by the Cochrane Collaboration for conducting systematic reviews of RCTs and non-randomised studies, and was robust.[82] We conducted comprehensive searches of bibliographic databases and grey literature. All stages of the review, involving screening, data extraction and assessment of risk of bias, were conducted by at least two researchers, either in duplicate or by one researcher with checks by a second researcher. We assessed the risk of bias using the Cochrane Collaborations's RoB 2.0 tool.[83] We contacted authors from both studies to provide further information on participants' staging of melanoma and obtained data for stage I patients from one study.[85] We excluded the majority of potentially relevant primary full-text articles because they were non-comparative or did not assess surveillance strategies.

Because the single included study[85] was conducted in a single country, the USA, the findings could be limited in applicability and generalisability. The assessment of risk of bias revealed an overall judgement of 'some concerns' due to attrition in the trial at 24 months.[85] Publication bias could not be investigated because of the number of studies identified, but the possibility should be considered as there may be studies that did not find positive results and remain unpublished.

Impact and implementations

Evidence for the effectiveness of surveillance and follow-up strategies for stage I melanoma is limited. A previous systematic review sought to identify the range of stage-specific surveillance practices for melanoma patients (any stage) and concluded that surveillance strategies vary around the world during the first 5 years post treatment.[25] Our review had narrower inclusion criteria with respect to study design and staging of melanoma. The paucity of this evidence in our review makes it difficult to make recommendations regarding the effectiveness of surveillance and follow-up strategies for stage I melanoma in the UK.

Conclusion

This review demonstrates that evidence for the effectiveness of surveillance strategies is poor for stage I melanoma patients. We were able to obtain data specific to stage I patients from only one of two included studies. This study suggested that an educational intervention encouraging SSE by patients and their partners might be promising and effective overall in increasing SSE and detection of new or recurrent disease by the patient–partner dyads. However, for patients with stage I disease, there was little evidence of benefit of the intervention, compared with control, for detecting new or recurrent disease.

The findings of this review are not wholly unexpected, given the assessment of the existing guidelines for surveillance of stage I disease presented in *Chapter 2*. What the work presented in both this chapter and *Chapter 2* illustrates is the paucity of data on which existing strategies are based on and it raises questions as to whether or not, or how, alternative strategies may be better than current practice. The following chapters go on to consider whether or not alternative strategies for surveillance could be developed. *Chapter 4* begins this process by considering the evidence base for approaches to identify those people with stage I disease who might be more at risk of recurrence and, consequently, where there may be more merit in adopting a more intensive surveillance strategy, rather than a less intensive strategy.

Chapter 4 A systematic review of the prognostic accuracy of risk models used for the prediction of recurrence, new primary tumours or metastasis for patients with American Joint Committee on Cancer stage I melanoma following surgical excision of a primary cutaneous tumour

Brief overview

A risk prediction model is a statistical tool that uses multiple predictors to estimate the absolute probability or risk that a certain outcome will occur in an individual with specific risk factors.[89] Great advances in the earlier detection of melanoma have been achieved through increased public awareness, the adoption of dermatoscopic examinations and a rapid '2-week wait' referral system in the UK.[18] There is also widespread belief that earlier detection of metastatic disease results in improved overall patient outcomes.[90] At present, however, there is no internationally accepted standardised model of follow-up of patients diagnosed with AJCC stage I cutaneous melanoma, with wide variations in care across North America, Australia, Europe and the UK.[25] Although the surgical excision of a primary melanoma is effective and long established, there has been a rapid pace of change recently with the addition of earlier investigatory techniques such as SLNB[91,92] and various radiological modalities,[93] and a raft of advances in the treatment of metastatic disease.[94–96] However, a structured, uniformly adopted, evidence-based model of patient follow-up after initial diagnosis is lacking. Current guidelines vary across the world, with most high-income countries using anecdotal evidence and expert opinion.[53] These are usually underpinned by the assumption that earlier detection of metastatic disease results in improved overall outcome. However, they often do not take a wider, holistic view of the patient pathway to identify a model that incorporates all of the elements used in the diagnosis and management of the condition. Thus, they may fail to adequately capture physical, psychological consequences and the costs of such strategies.

Although systematic reviews of predictive models for primary cutaneous melanoma have been conducted,[97,98] to date, to our knowledge, none has investigated the potential of models to predict recurrence, new primary tumours and metastases of AJCC stage I melanoma. This work seeks to systematically review and pool the evidence of the various elements that underpin an ideal model of follow-up, thus allowing others to make recommendations on future care models for AJCC stage I melanoma in the UK. With the rapid increase in melanoma rates, it is paramount that the UK develops a robust, evidence-based model of follow-up care for the majority of affected patients, namely patients with AJCC stage I disease.

Research aim

This review aimed to:

- identify all studies of prognostic risk models for recurrence of melanoma among AJCC stage I survivors
- determine the performance of prognostic risk models (external validation, including discrimination and calibration) used to determine the risk stratification of patients with AJCC stage I melanoma after surgical excision of primary cutaneous tumour.

Methods

The study adheres to the guidelines for the PRISMA statement to ensure transparency of the process.[78] As with the review reported in *Chapter 3*, details of this review are published in the study protocol published on PROSPERO (CRD42018086784).[79]

For this review, we planned to assess the prognostic accuracy of the biochemical and biophysical biomarkers and risk models used (alone or in combination) for the prediction of recurrence, new primary tumours or metastasis for patients with AJCC stage I melanoma following surgical excision of primary cutaneous tumour. However, following expert advice of methodological and clinical experts, we modified the objectives to focus on the accuracy of risk prediction models only. This change was necessary to make the review feasible, as a larger number of studies may have made completion of the study impossible.

Search strategy

The search strategy was designed by an experienced information specialist, in collaboration with the project team. The search was designed in MEDLINE (via Ovid) according to the following concepts: [melanoma] AND [risk models] AND [prognosis]. A published and validated prognostic study filter was used.[99] The strategy used database-specific thesaurus headings, along with title and abstract keywords, with appropriate use of stemming for alternative word endings, alternative spelling and plurals. The search strategy was translated to the databases listed in *Box 4*. No restrictions were applied according to language or country.

An example of the search strategy used in MEDLINE can be found in *Appendix 3*.

Supplementary searches were limited based on the development and uptake in practice of SLNB in melanoma as a new way of diagnosing and managing patients. The first evidence of its routine use in some of the larger melanoma centres in the UK, as well as in routine practice in the USA, was around 2000.[100] We, therefore, limited the subsidiary search criteria to this date.

In addition to the databases and resources reported in *Box 5*, guidelines and other subsidiary journal content were handsearched and references of relevant publications were searched. This supplemented the structured documented searches, aiming to ensure that relevant studies were not overlooked as a result of selective, poorly or inaccurately indexed, or unindexed content.

Inclusion and exclusion criteria

We based inclusion and exclusion criteria on the PICOTS formula, as outlined in the following sections.[80]

BOX 4 Databases searched for the review of risk prediction models

- MEDLINE (via Ovid).
- EMBASE (via Ovid).
- CENTRAL (the Cochrane Library via Wiley Online Library).
- Health Technology Assessment Database (the Cochrane Library via Wiley Online Library).
- CINAHL (via EBSCO*host*).
- Science Citation Index (via the Web of Science).
- Conference Proceedings Citation Index – Science (via the Web of Science).
- Cochrane Database of Systematic Reviews (the Cochrane Library via Wiley Online Library – to check included studies of relevant reviews).
- Grey literature was sought using similar keywords to search various resources including, but not limited to, the following:
  - OpenGrey (www.opengrey.eu/; accessed 17 April 2019, includes SIGLE, EAGLE, GreyNet).
  - Cancer Research UK (www.cancerresearchuk.org/about-cancer/find-a-clinical-trial; accessed 17 April 2019).
  - Melanoma UK (www.melanomauk.org.uk; accessed 17 April 2019).
  - National Guideline ClearingHouse (www.guideline.gov/; accessed 17 April 2019).
- Ongoing trials identified using the WHO's ICTRP platform of trials registries. (www.who.int/trialsearch; accessed 17 April 2019).

CENTRAL, Cochrane Central Register of Controlled Trials; CINAHL, Cumulative Index to Nursing and Allied Health Literature; EAGLE, European Association for Grey Literature Exploitation; ICTRP, International Clinical Trials Registry Platform; SIGLE, System for Information on Grey Literature in Europe; WHO, World Health Organization.

BOX 5 Databases searched for diagnostic accuracy review

- MEDLINE (via Ovid), 1946 to 2 July 2019.
- EMBASE (via Ovid), 1980 to 2 July 2019.
- CENTRAL (the Cochrane Library via Wiley Online Library), issue 6, 2019.
- CINAHL (via EBSCO*host*), 1981 to July 2019.
- Scopus® (Elsevier, Amsterdam, the Netherlands).
- Conference Proceedings Citation Index – Science (via the Web of Science), 1990 to July 2019.
- Cochrane Database of Systematic Reviews (the Cochrane Library via Wiley Online Library), issue 6, 2019.
- Science Citation Index (via Web of Science — to include Conference Proceedings), 1990 to July 2019.

CENTRAL, Cochrane Central Register of Controlled Trials; CINAHL, Cumulative Index to Nursing and Allied Health Literature.

Population

Adults aged ≥ 18 years treated for AJCC (eighth edition) stage I cutaneous melanoma [stage IA (≤ 0.8 mm thick without ulceration) or stage IB (< 0.8 mm thick, or < 1 mm thick and ulcerated skin)].[19]

Or:

Adults aged ≥ 18 years treated for AJCC (seventh edition) stage I cutaneous melanoma [stage IA (T1a ≤ 1 mm thick) or stage IB (T1b with ulceration or mitoses ≤ 1 mm thick, or T2a 1.01 to 2.00 mm thick and no ulceration)].[20]

Studies that combined patient populations (e.g. all stages of disease) were included only in cases for which it was specified that the test/data also applied to stage I cases. Studies that did not specify a patient population were included in the first instance, pending confirmation from study authors, when possible.

Types of prognostic models

We assessed all prognostic or predictive models used to predict the likelihood of recurrence (any site) or metastasis or survival in patients with stage I melanoma.[89] For a definition of a prognostic model, see *Brief overview*.

Outcomes

Studies were included that presented the predictive accuracy of the risk model in relation to recurrence, metastasis and survival using statistical measures:[101]

- discrimination: ability to differentiate between high and low risk
- calibration: agreement between observed and predicted risk
- overall performance: a combination of discrimination and calibration.

Timing

The application of the model had to have been post resection of the primary cutaneous tumour. The timing will be dictated by the duration of included studies and interpreted accordingly.

Setting

All studies were eligible for inclusion, regardless of whether the study was conducted in primary, secondary, or tertiary care.

Study design

Studies were included if they:

- used statistical methods to present or validate (external) models used to:
    - predict melanoma outcomes of interest [minimum of two predictors of outcomes, e.g. Breslow depths, location of tumour, type of recurrence (local, regional, distant), age, sex]
    - group patients based on their risk of developing such outcomes (risk prediction models).
- were validated – evaluated to determine the reproducibility of a developed prediction model for the derivative sample and prevent overinterpretation of current data either:[101]
    - internally – data for model development and evaluation are both random samples from the same underlying population
    - externally – predictions are calculated from the previously developed model and tested in new data that are different from the development population (e.g. from another hospital).

Data collection

Selection of studies

The selection of studies that met the inclusion criteria was conducted in two stages. Studies were exported from the EndNote library and into Rayyan.[81] For consistency and accuracy, two sets of two reviewers initially piloted the screening process by assessing 10% of the studies based on the titles and, when available, abstracts and some full texts against the prespecified inclusion and exclusion criteria. Disagreements at this stage were resolved either by discussion between the reviewers or with arbitration from another member of the study team. In the second stage, studies that appeared to meet the inclusion and exclusion criteria were imported into EndNote and full-text papers were obtained. When full texts were not readily available, we obtained articles via interlibrary loans.

Two reviewers independently screened these articles and made their selection in accordance with the eligibility criteria.

Data extraction

In pairs, four reviewers independently extracted data according to the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS).[102] The following data were extracted: source of data, participants, outcomes, predictors, model development methods, model performance and validation. The completed extraction forms were independently checked for accuracy and consistency, with any disagreements resolved through discussion or by arbitration from another member of the team. An example of this checklist is available in *Report Supplementary Material 1*.

Risk-of-bias assessment in included studies

Working in pairs, the risk of bias of each included paper was assessed independently by one reviewer and was checked by a second reviewer. Disagreements were resolved through discussion or with arbitration from a third member of the study team. Studies were assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST), which addresses four domains that may influence the applicability of the prediction models (participants, predictors, outcome and analysis).[103]

Missing data

We set out to report the number (per cent) of missing data for all variables/outcomes. We did not impute missing outcome data for any of our specified outcomes.

Data analysis and synthesis

We narratively synthesised the predictive performance of the prediction models by analysing the statistical measures of predictive performance. Models were assessed for discrimination, calibration and overall performance from studies providing sufficient data. The review aimed to pool the evidence of model performance by performing a meta-analysis when possible.

Quality of the evidence using GRADE approach

We had planned to assess the overall quality of evidence for key outcomes using the GRADEpro tool;[84] however, it was not developed for, and the evidence suggests it performs poorly for, prediction modelling studies.[104] Given this, we decided not to use the GRADE approach.

Results

Number of studies identified

Our search identified 25,251 records from the electronic databases. After deduplication, 20,878 records remained; following screening of the titles and abstracts, 112 full texts of potentially relevant articles were retrieved for examination. The PRISMA flow diagram outlines the study selection process and the reasons for exclusion (*Figure 4*). A total of 11 articles reporting 11 different risk prediction models met the full inclusion criteria of this review.[105–115] A total of 101 studies were reviewed fully and excluded for the following reasons: used single prognostic factors for model development (21%), combined stages of the disease (28%) or were not validated (51%). Details of the excluded studies are presented in *Appendix 4*.

Characteristics of included studies

A summary of the studies and patient characteristics is presented in *Table 7*. Eight of the studies were conducted in the USA,[106–109,111,113–115] one in the UK,[112] one in Australia[105] and one in Italy.[110] Seven studies used a retrospective cohort design,[105–109,112,115] three used a prospective design[111,113,114] and one used a retrospective cohort of prospectively collected data.[110] Patient data for development and validation of these models were taken from cancer registries, AJCC melanoma databases, clinical data from patients diagnosed and treated for melanoma or a combination of one of these sources.

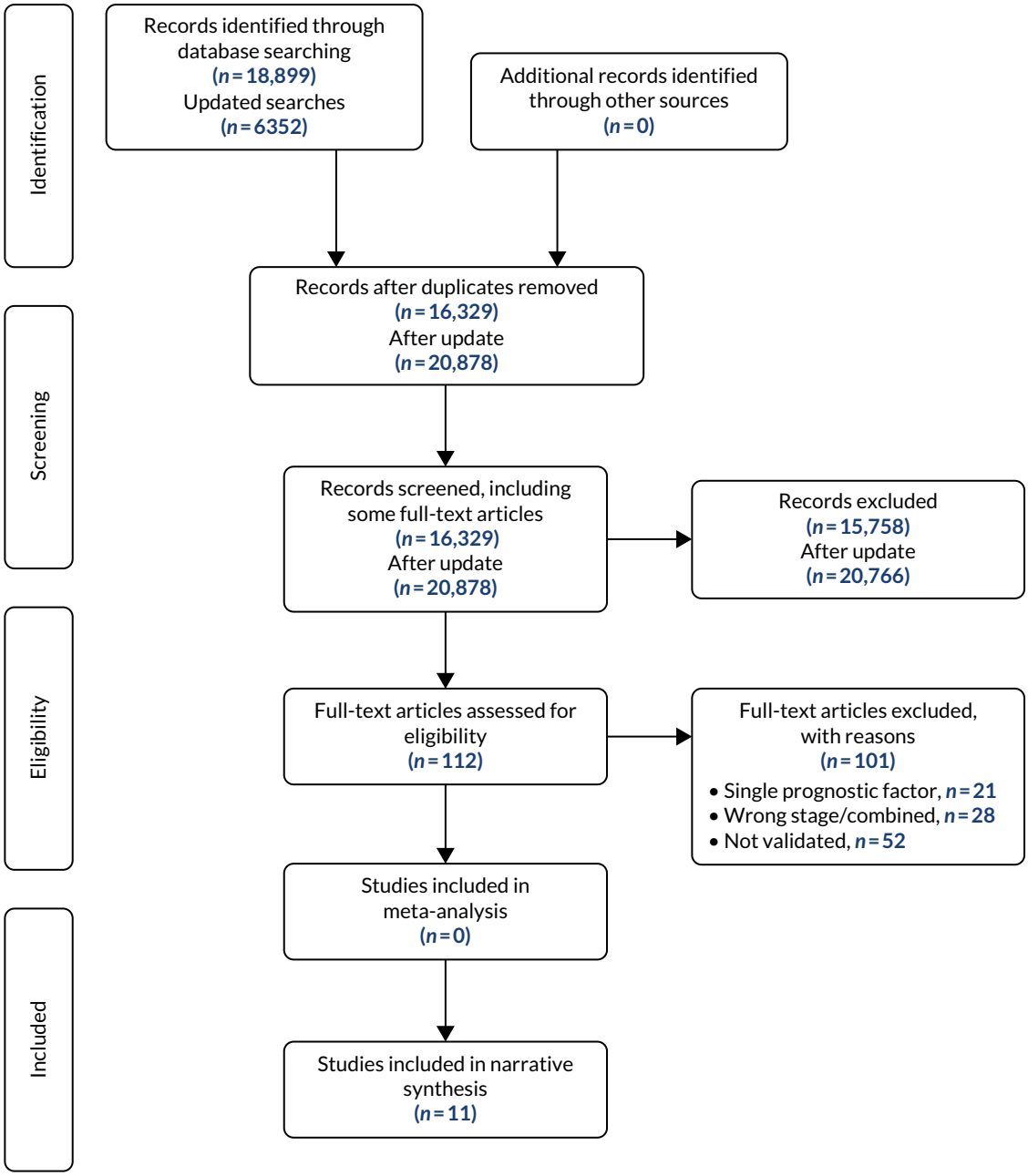


FIGURE 4 The PRISMA flow diagram for the risk prediction model review.

Each study presented a model that could be used to predict either recurrence, metastasis or survival. The studies concerned the development of a risk or prognostic score,[107,115] a nomogram,[110] the Melanoma Severity Index model,[105] a novel histopathological classifier,[111] an electronic prediction tool,[113] a classification tree,[114] prognostic trees,[108,109] a model focused on adding a new predictor to an established model[112] and a model used to validate AJCC staging.[106]

Outcome definitions and follow-up times varied across studies. The median time of follow-up ranged from 42.5 months[107] to 10.3 years.[110] All studies reported outcome measures of survival. Seven studies defined survival as patients who were alive at last follow-up or who died without evidence of melanoma.[105,106,108–110,112,113] Two studies defined survival as the number of patients who are alive after diagnosis.[107,111] The other studies did not provide a definition of survival.[114,115] Risk stratification for predicting overall survival was reported in nine studies.[106–114] Patients were reportedly grouped according to tumour thickness,[106,113] ulceration status,[114] melanoma-specific death or survival,[107,108,110,111]

TABLE 7 Characteristics of prediction model studies

Study (first author and year of publication)	Study design	Country	Statistical methods	Data source	Time period	Follow-up time	Participants
Baade *et al.*[105] 2015	Retrospective cohort	Australia	Multivariable probit regression model	Population-based Queensland cancer registry	1995–2008	Up to 16 years; median 7.2 years (86.4 months)	*n* = 28,654 • 44%, ≤ 0.50 mm • 28%, 0.51–1.00 mm • 9%, 1.01–1.50 mm • 5%, 1.50–2.00mm
Balch *et al.*[106] 2001	Retrospective cohort	USA	Cox proportional hazards regression model	Prospective population-based databases from 13 institutions merged to form the AJCC Melanoma Database	NR	• 12,837 (73%) for at least 5 years • 8633 (49%) for at least 10 years • 2485 (14%) for at least 20 years	• Total, *N* = 17,600 • *n* = 13,581 (stages I–II) ○ 39%, ≤ 1.00 mm ○ 29%, 1.01–2.00 mm ○ 22%, 2.01–4.00 mm ○ 10%, > 4.00 mm
Cochran *et al.*[107] 2000	Retrospective cohort (consecutive)	USA	Cox proportional hazards regression model	Subset of John Wayne Cancer Institute Melanoma clinical database, Division of Surgical Oncology, UCLA Random sampling of patients into two equal groups: 1. Development set: estimation set 2. Validation set: test set	1980–90	Total database: median 42.5 months (range 1–26.5 years)	*n* = 1042

continued

TABLE 7 Characteristics of prediction model studies (*continued*)

Study (first author and year of publication)	Study design	Country	Statistical methods	Data source	Time period	Follow-up time	Participants
Gimotty *et al.*[109] 2004	Retrospective cohort (consecutive)	USA	A recursive partitioning algorithm used in classification and regression tree	Development set: population-based SEER registry Validation set: new SEER patients seen 1991–5	• 1972–91 • 1991–5	At least 10 years	• SEER: $N = 884$ • Stage IA (T1a), $n = 759$ (86%) • Stage IB (T1b), $n = 123$ (14%) • New patients (validation): $n = 144$
Gimotty *et al.*[108] 2007	Retrospective cohort	USA	Recursive partitioning to develop classification trees in development and validation sets	• Development set: population-based SEER Registry • Validation set: clinical-based PLC registry	• 1998–2001 • 1972–2001	• SEER: the median time to last follow-up was 4.6 years • PLG: the median time to last follow-up was 8.1 years	• SEER: $n = 26{,}114$ • PLG: $n = 2389$
Maurichi *et al.*[110] 2014	Retrospective cohort of prospectively collected data	Italy: six European centres	Multivariable Cox regression	• Development set: patients diagnosed and treated for a single thin primary melanoma at one of six European clinical-based centres • Validation set: internal validation by calibration and computing the bootstrap-corrected Harrell's *c*-statistic	1996–2004	Median follow-up: • 124 months (IQR 106–157 months) • 10.3 years (IQR 8.8–13.1 years)	$N = 2243$ • Stage T1a, $n = 1128$ (50.3%); Breslow thickness (mm), median 0.52 (IQR 0.33–0.72) • Stage T1b, $n = 1115$ (49.7%); Breslow thickness (mm), median 0.71 (IQR 0.50–0.79)

Study (first author and year of publication)	Study design	Country	Statistical methods	Data source	Time period	Follow-up time	Participants
Rosenbaum *et al.*[111] 2017	Prospective cohort. Note that data were not split into discovery and test cohorts, because disease recurred in only 63 stage IB patients	USA	Linear regression analysis, Cox proportional hazards regression model, area under the receiver operating characteristic curve	• Development set: patients presenting to NYU with AJCC stage IB melanoma and enrolled in a prospective clinicopathological biospecimen database • Validation set: subset of patients with a recurrence or no recurrence, matched for age, sex, histopthological subtype, thickness, ulceration and mitotic rate	August 2002–May 2014	Median 4.4 years	*N* = 655 • Development set: *n* = 506 • Validation set: *n* = 149 • Breslow thickness (mm), median 0.98 mm (SD 0.44)
Saldanha *et al.*[112] 2018	Retrospective cohort (consecutive)	UK	Cox proportional hazards regression model and Kaplan–Meier survival plots. Models compared using Akaike information criterion	• Development set: patients presenting to University Hospitals, Leicester with primary invasive disease • Validation set: patients presenting to Nottingham University Hospitals NHS Trust with primary invasive disease	• Development set: 2004–11 • Validation set: 2003–5 and 2008–10	Median 71 months (5.9 years)	*N* = 1329 • Development set: *n* = 970 • Validation set: *n* = 359 • Breslow thickness (mm), median 0.90 (IQR 0.50–2.00) AJCC (seventh edition): • Stage IA, *n* = 376 (38.8%) • Stage IB, *n* = 326 (33.6%) • Stage IIA–III, *n* = 173 (17.8%)

continued

TABLE 7 Characteristics of prediction model studies (*continued*)

Study (first author and year of publication)	Study design	Country	Statistical methods	Data source	Time period	Follow-up time	Participants

Soong *et al.*[113] 2010	Prospectively observed cohort	USA	Multivariate analysis based on the Cox regression model	• Development set: AJCC population-based Melanoma Database (2008) – data from nine major institutions and co-operative study groups
				• Validation set: patients treated at Sydney Melanoma Unit, Australia
					• Development set: 26% diagnosed after 2002
					• Validation set: NR
						NR	*N* = 25,734
							• Development set: 14,760
							• Validation set: 10,974
							Breslow thickness (mm):
							• 0–0.50
							  ○ Development set: 22.3%
							  ○ Validation set: 18.1%
							• 0.51–1.00
							  ○ Development set: 23.3%
							  ○ Validation set 28.1%
							• 1.01–2.00
							  ○ Development set: 29.6%
							  ○ Validation set 26.4%
							• 2.01–6.00
							  ○ Development set: 25.0%
							  ○ Validation set: 27.3%

Study (first author and year of publication)	Study design	Country	Statistical methods	Data source	Time period	Follow-up time	Participants

							• Stage IA:
							  ○ Development set: 36.0%
							  ○ Validation set: 37.4%
							• Stage IB:
							  ○ Development set: 33.4%
							  ○ Validation set: 30.1%
							• Stage IIA–IIC:
							  ○ Development set: 30.6%
							  ○ Validation set: 32.5%

Tsai *et al.*[114] 2007	Prospective cohort study	USA	Survival tree	Registry data: AJCC population-based Melanoma Database	NR	NR	*N* = 13,268 • Tumour thickness 0.10–1.00 mm: *n* = 5299 (39.94%) • Tumour thickness 1.01–2.00 mm: *n* = 3753 (28.29%) • Tumour thickness 2.01–4.00 mm: *n* = 2836 (21.37%) • Tumour thickness > 4.00 mm: *n* = 1380 (10.40%)
Vollmer and Seigler[115] 2001	Retrospective cohort	USA	Cox proportional hazards regression model	University Melanoma Clinic	Late 1980 to early 1990s	Median follow-up 7.6 years	NR

IQR, interquartile range; NR, not reported; NYU, New York University; PLC, Pigmented Lesion Clinic; PLG, Pigmented Lesion Group; SEER, Surveillance, Epidemiology, and End Results; UCLA, University of California, Los Angeles.

growth phase lesions[109] or Breslow depth.[112] Two studies measured the risk of recurrence, defined and stratified either as recurrence for individual melanoma patients at different points after initial treatment[107] or as local recurrence (a recurrence in the scar at the primary site).[109] Gimotty *et al.*[109] classified and stratified patients with local recurrences as:

- type 1 – with radial growth phase (RGP)
- type 2 – with RGP and vertical growth phase
- type 3 – without RGP.

Another study[111] measured recurrence-free survival, defined as the time from diagnosis to the first recorded date of regional or distant metastases. Two studies[109,110] provided outcome measures for metastasis. Gimotty *et al.*[109] defined metastasis as regional metastasis (in-transit dermal or subcutaneous metastases and/or nodal involvement). Maurichi *et al.*[110] did not provide a definition of metastasis.

Characteristics of included models

Most of the studies used regression methods for building the models. Characteristics of the models are presented in *Table 8*. Seven studies used the Cox proportional hazards method.[106,107,110–113,115] Two studies used a recursive partitioning algorithm used in classification and regression tree analysis.[108,109] Tsai *et al.*[114] used the survival tree analysis method. Baade *et al.*[105] used the probit regression method.

At model development stage, various methods were used across the studies to choose the variables used in the final model. Two studies used the backward procedure based on the Akaike information criterion, an estimate of the measure of the quality of available models as they relate to one another for a certain set of data.[110,112] The Akaike information criterion is used to determine what variables influence the prediction of an outcome of interest and how these variables influence the outcome by estimating several different regression models to balance the trade-offs between the complexity of a given model and its goodness of fit.[118] Two studies[108,109] selected predictors based on evidence of previous validation studies and two studies[107,111] performed a univariate analysis to select only the variables for which evidence of statistical significance was generated. The remaining studies did not report on the criteria used to select the candidate predictors.[105,106,113–115] The number of predictors in the studies ranged from 3[115] to 11 predictors.[109] Various possible risk factors were identified, the most common being age, tumour site, tumour thickness, sex and ulceration. Other predictors identified included metastasis, mitotic rate, positive lymph node, Clark's level, growth phase, RGP regression, microsatellites, anatomical level, presence or absence of lymphovascular invasion, tumour-infiltrating lymphocytes, histological subtype, conformation status and digital or manual area.

Risk-of-bias assessment of included studies

Table 9 summarises the risk of bias and concerns regarding applicability to the intended population and setting of the included studies. Assessments were conducted using the PROBAST. Overall, eight studies[105,107,109–112,114,115] were judged to be at high risk of bias and three[106,108,113] were rated as having an unclear risk of bias. Five studies[106,108–110,113] were deemed to have a low risk of bias regarding applicability, another five[105,111,112,114,115] were deemed to have unclear risk and one study[107] was deemed to have high risk. Bias was introduced by various methods.

Selection of participants

All studies were judged to be at a high risk of bias for this domain. All studies used existing data sources to develop or validate their models. Participant selection was based on retrospective[105–110,112,115] and prospective[111,113,114] cohort studies of cancer registries,[105,106,108,109] AJCC databases,[113,114] clinical databases,[107,110,111,115] and hospital records.[112] As data are taken from existing sources, there is no information on how patients were selected and for what purpose. In model development and validation, bias may be introduced when routinely collected data are used, as opposed to data obtained from primary research. Existing data may not provide the full or accurate clinical features under investigation.[119]

TABLE 8 Characteristics of models

Study (first author and year of publication)	Predictors in final model	Model performance			Validation	
		Discrimination	Calibration	Overall performance	Internal	External
Baade *et al.*[105] 2015	MSI: $n = 7$ • Age at diagnosis, thickness, SRT-thickness, body site, ulceration, positive lymph nodes, metastasis	• Discrimination: *D*-statistic 1.50 (95% CI 1.44 to 1.56) • Harrell's *c*-statistic 0.88 (95% CI 0.88 to 0.89)	Not reported	Explained variation: RD2 statistic: 0.47 (95% CI 0.45 to 0.49)	Internal–external cross-validation across geographically defined subset	None
Balch *et al.*[106] 2001	$n = 8$ • Thickness, ulceration, age, sex, site, anatomical level and sex, nodal status, number of metastatic nodes	Not reported	Not reported	Not reported	The melanoma patient data were used to validate the proposed AJCC staging system	None
Cochran *et al.*[107] 2000	$n = 5$ • Thickness, ulceration, age, sex, site	Not reported	Not reported	Not reported	Patients randomly sampled into two equal groups: an estimation set and a test set	None
Gimotty *et al.*[109] 2004	$n = 11$ • Age at diagnosis, sex, anatomical site, mitotic rate, TILs, thickness, Clark's level, growth phase, RGP regression, ulceration, microsatellites	• Discrimination in SEER programme sample for 10-year metastasis • Risk groups: AUC = 0.85 • Stage: AUC = 0.59	Not reported	Not reported	New patients meeting study eligibility criteria between 1991 and April 1995	None
Gimotty *et al.*[108] 2007	$n = 6$ • Thickness, anatomical level, ulceration, site, sex, age	• Discrimination: SEER programme: AUC = 0.76 • PLG: AUC = 0.83	Not reported	Not reported	None	New patients

continued

TABLE 8 Characteristics of models (*continued*)

Study (first author and year of publication)	Predictors in final model	Model performance			Validation	
		Discrimination	Calibration	Overall performance	Internal	External
Maurichi *et al.*[110] 2014	*n* = 8 • Age, sex, site (head and neck, trunk, limbs), Breslow depth, mitotic rate, ulceration, Clark's level, lymphovascular invasion, regression, and TILs	Discrimination by adjusted Harrell's *c*-statistic = 0.88	Nomogram performance was assessed by calibration plot as an indicator of internal calibration	Not reported	Internal validation of nomogram by calibration of nomogram	None
Rosenbaum *et al.*[111] 2017	*n* = 6 • Age, sex, histological subtype, thickness, mitoses, and ulceration status *n* = 3 • Novel predictors: histopathological width, conformation status (contiguous or non-contiguous), digital or manual area	Patients classified using Youden Index of the ROC curve using digital area, conformation and baseline variables: • AUC = 0.733, (95% CI 0.647 to 0.818) • Baseline classifier: AUC = 0.635 (95% CI 0.545 to 0.724)	Not reported	Not reported	10-fold cross-validation	None
Saldanha *et al.*[112] 2018	*n* = 2 • Breslow depth as a novel upstaging feature. Classified into low, mid and high categories (≤ 25%, 30–75%, ≥ 80%) AJCC eighth edition staging	Discrimination: • Gönen and Heller's[116] *k* – Leicester 0.78 (SE 0.01) vs. Nottingham 0.78 (SE 0.01) • Harrell's c-index – Leicester 0.84 (SE 0.03) vs. Nottingham 0.81 (SE 0.04) (slight loss of discrimination)	Calibration: perfect calibration in any validation set would be represented by a calibration slope of 1, and the slope in the validation cases was 0.88 (SE 0.12)	Not reported	None	Patients from Nottingham University Hospitals NHS Trust

Study (first author and year of publication)	Predictors in final model	Model performance			Validation	
		Discrimination	Calibration	Overall performance	Internal	External
Soong *et al.*[113] 2010	*n* = 6 • Age, sex, primary melanoma site, primary tumour thickness, level of invasion, primary tumour ulceration	Not reported	Calibration: concordance correlation coefficients of 0.90 and 0.93 for 5- and 10-year survival rates	Not reported	None	Patients from Sydney Melanoma Unit, Australia
Tsai *et al.*[114] 2007	*n* = 6 • Age, sex, primary melanoma site (extremity or axial), tumour thickness (mm) (0.10–1.00, 1.01–2.00, 2.01–4.00, > 4.00), Clark's level of invasion, and ulceration (absent/present)	Not reported	Not reported	Measure for overall performance, captures both discrimination and calibration (Brier score). Brier score range: 0.02 at year 1 to 0.20 at year 15 for the propsed model (intergrated tree-based scheme)	Fivefold cross-validation	None
Vollmer and Seigler[115] 2001	*n* = 3 • Age, sex, body site of the primary tumour [extremities (excluding acral) and all other sites (including acral)]	Not reported	Not reported	Not reported	Cross-validation using data from another set by Stadelmann[117]	None

AUC, area under the curve; MSI, Melanoma Severity Index; PLG, Pigmented Lesion Group; ROC, receiver operating charateristic; SE, standard error; SEER, Surveillance, Epidemiology, and End Results; SRT, smooth rank transform; TIL, tumour-infiltrating lymphocyte.

TABLE 9 Risk of bias (PROBAST)

	Risk of bias				Applicability			Overall	
Study	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of bias	Applicability
Baade *et al.*[105] 2015	–	?	–	?	–	+	–	–	–
Balch *et al.*[106] 2001	–	?	–	?	–	+	–	–	–
Cochran *et al.*[107] 2000	–	+	–	–	–	+	–	–	–
Gimotty *et al.*[109] 2004	–	+	–	?	?	+	+	–	?
Gimotty *et al.*[108] 2007	–	+	–	?	?	+	+	–	?
Maurichi *et al.*[110] 2014	–	+	–	+	?	+	+	–	?
Rosenbaum *et al.*[111] 2017	–	+	–	?	?	+	+	–	?
Saldanha *et al.*[112] 2018	–	+	–	?	–	+	–	–	–
Soong *et al.*[113] 2010	–	+	–	?	–	+	–	–	–
Tsai *et al.*[114] 2007	–	+	–	?	–	+	–	–	–
Vollmer and Seigler[115] 2001	–	+	–	?	–	+	–	–	–

+, low risk of bias/low concern regarding applicability; –, high risk of bias/high concern regarding applicability; ?, unclear risk of bias/unclear concern regarding applicability.

Two studies[107,115] were also rated as being at a high risk of bias because of the particpants selected. In neither study is it clear how many participants had stage I disease, as per our review question, because very little detail was given on the participants in the study, and there was no summary of patient characteristics to judge the severity of their melanoma.

Risk of bias introduced by predictors

Nine studies[107–115] were judged to have a low risk of bias for the selection of risk factors included in the final model. The risk factors considered for use in all the studies appear to be representative of the factors used primarily in melanoma. The study by Baade *et al.*[105] was rated as having an unclear risk of bias because there was an issue of missing data for 30% of the initial cohort. The study used a predictive mean matching multiple imputation approach for handling missing data to evaluate their effect on risk model estimation and the reliability of the predictions. This method assumes that the missing data can be replaced by the available data, as they are thought to be similar. In this case, as the data were reviewed retrospectively, it is unclear how similar the missing values and the replacement values are. Another study was also rated as having an unclear risk of bias because the information about some prognostic factors was not available consistently enough to include them in the prognostic factors analysis.[106]

Risk of bias introduced by outcomes

All the studies were rated as having a high risk of bias for this domain. Follow-up times introduced bias in some of the studies. Six studies[105,107,111,112,114,115] were rated as having a high risk of bias for outcome analysis because insufficient time was taken to follow up patients. The median follow-up times ranged from 3.5 years[107] to 7.6 years.[115] In localised melanoma, this is too short to detect recurrence and death from melanoma; 10 years' follow-up is generally considered sufficient for adequate patient evaluation.[113] Although the study by Soong *et al.*[113] reported a follow-up time ranging from 0 to 20 years, the risk of bias was minimised by identifying patients who were alive at the time of the last follow-up or who died without evidence of melanoma.

Although all model development or validation studies prespecified the outcome definition for overall survival, one of the studies[110] failed to provide the definition of metastsis as an outcome. Overall, all studies were rated as having a high risk of bias because the data used were based on routinely collected data, meaning it is possible that different outcome definitions were used.

Risk of bias introduced by the analysis

All studies were rated as having an unclear risk of bias due to the analysis methods used. Six of the studies[105,106,108,109,112,115] did not include all the enrolled participants in their analyses. This was because data were missing for at least 30% of the initial cohort,[105] an unknown number of patients died[106] or there were unexplained missing data.[108,109,112,115] Excluding some participants would likely introduce differences in the performance of a prediction model for predicting outcomes. Moons *et al.*[120] also suggest that determining the extent of bias in prediction models developed based on routine data can be problematic, often because of the lack of clarity in the elibility criteria. One study was rated as having an unclear risk of bias as a result of the inclusion of variables identified as significant following a univariate analysis.[111] As this method excludes risk factors that are considered non-significant for the developed model, it is likely to reduce the performance of the model and the model may not perform well in different populations.[121] The other two studies were rated as having an unclear risk of bias because they provided no information regarding the enrolment of participants, handling of missing data and whether or not complexities in the data were accounted for appropriately.[113,114]

In addition to this, it was difficult to decide whether or not an appropriate sample size was used for model development and internal validation. The events per variable (EPVs) were not reported in each study and none of the studies described which method it used to determine the appropriate sample size.

Concerns regarding applicability

Concern regarding selection of participants

Four studies[108–111] were rated as having low overall concern regarding applicability, five were rated as having unclear concern[105,106,112–114] and two were rated as having high concern.[107,115] The four studies were rated as unclear because, although participant data were taken from routine care or cancer registries, it was unclear whether or not the participants included in the studies matched the participants in the review question, as inclusion criteria were not provided. There was high concern for the participants selected not matching the review question for two studies.[107,115] Both studies provided no information regarding the characteristics of patients or the severity of their melanoma. The five studies that were rated as having an unclear concern regarding applicabilty included participants with different stages of melanoma. Therefore, the population was not confined to patients with stage I melanoma. The proportions of stage I patients included were 85.9%,[105] 75.2%,[113] 68.2%,[114] 68.1%[106] and 66.4%.[112] Including melanoma patients of varying stages is likely to introduce differences in the performance of a prediction model[122] for predicting outcomes in early-stage versus advanced-stage melanoma.

Concern regarding assessment and timing of predictors

All studies were rated as having low concern regarding the definition, assessment and timing of the predictors. All predictors were measured using methods potentially applicable to the daily practice that is addressed by the review.

Concern regarding the applicability of the outcome determined

The applicability of seven studies[105–107,112–115] was rated as being of high concern. This was due to the inclusion of patients with advanced-stage melanoma, which meant that an accurate prediction estimate for early-stage melanoma patients only was unlikely. The rest of the studies were judged to have an unclear concern in terms of outcome definition, timing and method of determination defining the outcome, as intended by the review question. This was due to the inclusion of patient data from existing records. Not enough information was provided regarding the inclusion and exclusion of patients in the registries.

Performance of prediction models

Discrimination

Six studies[105,108–112] reported the discriminatory ability of the models to distinguish between patient survival, or those who have either recurrence or metastases of melanoma, and those who have not. There are several measures that can be used to quantify how well a test can accurately distinguish between patients, from low to high risk.

Three studies reported the estimates for the area under the receiver operating characteristic (AUROC), which plots sensitivity against (1 – specificity).[108,109,111] Values range from 0.5 (no discriminative ability) to 1 (perfect discriminative ability).[123] Gimotty *et al.*[109] reported discrimination for metastasis using the AUROC for risk groups as 0.85; a lower value for AJCC stages Ia and Ib was reported: 0.59. Gimotty *et al.*[108] reported the AUROC for survival from a sample of a US cancer registry, the original data set, as 0.76, and for patients seen at a hospital, the validation sample, as 0.83. Neither study reported the variability statistics for the AUROC. Rosenbaum *et al.*[111] reported discrimination for recurrence using the AUROC for a novel histopathological classifier as 0.733 (95% CI 0.647 to 0.818), compared with the baseline classifier alone, which was 0.635 (95% CI 0.545 to 0.724).

Three studies reported discrimination for survival data using the Harrell's *c*-statistic, also known as the concordance statistic/probability.[105,110,112] This is an equivalent of and interpreted in the same way as the AUROC that is used to measure the discriminative ability of linear regression models for binary outcomes. The concordance probability is the probability that, of a randomly selected pair of patients, the patient with the shorter survival time has the higher predicted risk.[124] Baade *et al.*[105] reported the

c-index as 0.88 (95% CI 0.88 to 0.89), Maurichi *et al.*[110] reported it as 0.88 and Saldanha *et al.*[112] reported the c-index for Leicester cases, the original data set, as 0.84, compared with 0.81 for Nottingham cases, the validation sample. Again, no variability statistics for the c-index were reported.

Baade *et al.*[105] also assessed the discriminatory ability of the model to predict overall survival by calculating Royston and Sauerbrei's[125] *D*-statistic. This statistic quantifies the observed separation between patients with low and high predicted risk, and was reported as 1.50 (95% CI 1.44 to 1.56).

Saldanha *et al.*[112] also reported the discrimination ability of the model using the Gönen and Heller[116] *k*-statistic, used to evaluate the discriminatory power and predictive accuracy of non-linear statistical models. The *k*-statistic is an extension to time-to-event data of the AUROC, which is used to assess the discrimination of logistic regression models. It involves only the regression parameters and the covariate distribution, and is, therefore, asymptotically unbiased.[126] This is based on the reverse definition of concordance, which is the probability that, of a randomly selected pair of patients, the patient with the higher predicted risk has the shorter survival time, and has the same interpretation as the c-index.[116] Saldanha *et al.*[112] reported the *k*-statistic for Leicester cases, the original data set, as 0.78 (no variability statistics provided), compared with Nottingham cases, the validation sample, which was 0.78, demonstrating evidence of discriminatory ability and that the model fit was retained between training and validation sets.

Calibration

Calibration measures were reported in three studies.[110,112,113] This refers to a model's accuracy of predicted risk probabilities and indicates the extent to which expected outcomes (predicted from the model) and observed outcomes agree. It is often assessed graphically by a calibration plot, with predictions on the *x*-axis and the actual outcome on the *y*-axis, whereby a perfect calibration is represented by a diagonal line on the graph with numerical values between 0 (no agreement) and 1 (perfect agreement).[101] Visual representation by Maurichi *et al.*[110] shows that most of the predicted values for 12-year overall survival were ≥ 0.7, suggesting that the model was well calibrated. Another study reported a calibration slope of 0.88 ($p = 0.5$)[112] in the validation cases for predicting overall survival. The other study[113] reported concordance correlation coefficients of 0.9 and 0.93 for 5- and 10-year survival rates, demonstrating high accuracy of the prediction model. Both results indicate high accuracy of the prediction models, as the predicted and actual observed survival probabilities are close to each other, given that the values of the slopes reported are closer to 1 (perfect calibration).[127]

Overall performance

Two studies measured the overall performance of the developed models.[105,114] Baade *et al.*[105] used the the R^2 measure, a statistic that indicates the percentage of the variance to measure overall model performance. The statistic ranges from 0%, when no variation is accounted for, to 100%, when all variation is accounted for. The measure of explained variation measure in the model was reported as 0.47 (95% CI 0.45 to 0.49).

Tsai *et al.*[114] assessed the Brier score, the result of a statistical test used to examine the accuracy of goodness of fit.[114] The score ranges from 0% for a perfect model to 0.25% for a non-informative model and a higher score means a higher inaccuracy of a prognostic classification scheme.[101] The model proposed by Tsai *et al.*[114] included an integrated tree-based approach for prognostic grouping of localised melanoma patients, compared with the existing AJCC melanoma staging system that used the tumour, regional lymph nodes and distant metastasis [i.e. tumour–node–metastasis (TNM)] system to classify patients. The study calculated the Brier score as a function of time for three classification schemes (all patients pooled in one group, the proposed integrated approach and the AJCC schemes) for up to 15 years; this was represented graphically. The score for all three schemes increased with time from 0% at year 1 to approximately 0.25% at year 15. However, both the AJCC and integrated

schemes were shown to yield superior performance, compared with the pooled Kaplan–Meier estimates, with the integrated approach having a slightly better improvement in Brier score than the AJCC scheme.[114] These results show that the proposed integrated approach is preferred over the AJCC scheme.

Model evaluation methods

All studies in the review were validated either internally or externally. Internal validation refers to the efficiency of a model developed and evaluated from the same underlying population.[128] External validation refers to how well a model predicts an outcome in a new data set, different from the development population.[129]

Eight studies[105–107,109–111,114,115] reported that they internally validated their models. Four studies[105,111,114,115] used four different cross-validation methods. Baade *et al.*[105] reported that they used the internal–external method (assessing consistency across a variety of different geographical areas in the state). However, the model was not validated against an external, independent data set. Rosenbaum *et al.*[111] used the 10-fold method (whereby the original sample is randomly partitioned into 10 equal-sized subsamples). Tsai *et al.*[114] used the fivefold method (whereby the original sample is randomly partitioned into five equal-sized subsamples). Vollmer and Seigler[115] used simple cross-validation methods (splitting the data sets into training/development samples and validation samples). Cochran *et al.*[107] used the random split-sample method, whereas Maurichi *et al.*[110] validated their model based on a calibration plot by assessing the congruence of expected outcomes (predicted from the model) and observed outcomes. Gimotty *et al.*[109] validated their model on new patients meeting study eligibility and Balch *et al.*[106] validated the model using the melanoma patient data used to validate the proposed AJCC staging system.

The rest of the models[108,109,112] were externally validated using the geographical validation method. Predictions were calculated from the previously developed model using the training/development population and tested in new data sets different from this population in a different geographical area. Soong *et al.*[113] developed their model in the USA and validated it by testing it on a data set comprising patients treated at a Sydney Melanoma Unit in Australia. Gimotty *et al.*[108] developed their model using data from a US population-based Surveillance, Epidemiology, and End Results cancer registry (1988–2001) and validated it using patients seen by the University of Pennsylvania's Pigmented Lesion Group (1972–2001). Saldanha *et al.*[112] developed their model using data from Leicester cases diagnosed between 2004 and 2011 and validated in set of cases from Nottingham University diagnosed from 2003 to 2005 and from 2008 to 2010. All three models reported measures of model performance in both the original and validation samples.

Discussion

Summary of model performance assessment

This systematic review identified studies describing 11 different models developed for the prediction of recurrence, new primary tumours or metastasis in patients with AJCC stage I cutaneous melanoma following excision. The models differed in the predictors used depending on the outcome of interest and statistical measures used to assess model performances; therefore, it was inappropriate to quantitatively synthesise their results. The lack of consensus in the approach used to select predictors is reflected in the model development methods. Only six studies reported the criteria used. Two studies used the backward procedure, which starts by including all predictors at the beginning and then subsequently removing predictors based on predefined criteria.[110,112] This is a preferred method because it has the ability to eliminate redundant predictors.[130] Two studies used the univariate analysis method to screen variables.[107,111] This is the simplest method as it analyses one predictor at a time. However, this is likely to overestimate regression coefficients and overfit models.[131] Two studies[108,109] used evidence from previous study reports to identify predictors, and the rest of the studies did not report on how predictors were selected; therefore, it is difficult to determine whether or not the methods used were appropriate.

Model performance measures were available for assessing discrimination in six studies,[105,108–112] assessing calibration in three studies[110,112,113] and assessing overall performance in two studies.[105,114] The area under the curve (AUC) of the six studies ranged from 0.59 to 0.88. The discriminative performance of the models is considered acceptable when the AUROC statistics and their equivalent are ≥ 0.7.[123] Not all studies reported the variability statistics for the AUC. However, those that did report variability statistics reported a value below this estimate; therefore, it is unclear if all the models could accurately discriminate between those with defined outcomes and those without. The three studies that assessed calibration measures all reported values ≥ 0.7, which is closer to 1 (perfect calibration).[110,112,113] This suggests that all three models have the ability to accurately generate predictions that are close to the observed outcomes. Two studies[105,114] measured overall performance. Baade *et al.*[105] assessed this by assessing how well the model fits the data using the R^2 statistic. Higher R^2 values represent smaller differences between the observed data and the fitted values; the model by Baade *et al.*[105] reported an R^2 of 0.47 (95% CI 0.45 to 0.49), indicating that the model explains an estimated 47% of the variation. Another study[114] measured the overall performance by assessing the Brier score, a statistical test used to examine the goodness of fit. This score is useful because it simultaneously captures discrimination and calibration, and summarises the magnitude of error in the probability forecast; values range from 0.0 (total accuracy) to 1.0 (total inaccuracy).[132] The score tested the performance of an integrated tree approach, as opposed to the AJCC scheme, over a period of 15 months. Although there was little difference between the scores (all were $< 0.25\%$), the new model presented slightly better scores, suggesting that it was the preferred model to accurately predict survival rates for individual patients with localised melanoma.

Most of the studies validated their prediction models internally using data from their development set.[105–107,109–111,114,115] Research shows that, although models validated internally may show acceptable performance, it is not guaranteed that they will produce the same results in a different group of participants.[133] A few of the models were validated using external populations from other institutes.[108,112,113]

All models were rated as having a high risk of bias. The main source of bias related to the inclusion of patient data from existing databases or cancer registries. Although using existing data can be beneficial in that they are cheaper and quicker to obtain, there is a risk that information may have been collected for a particular purpose, thereby including irrelevant items and perhaps not recording outcomes of interest. Some of the studies identified as such included patients with advanced stages of melanoma; this could have distorted the predictive ability of the models in their favour.

Another source of bias was the omission of statistical analysis in the estimated predictive performance of the models. There was often insufficient or no information given regarding sample size determination. Sample size is often based on the ratio of the number of individuals with the outcome event to the number of candidate predictors (EPV). Models developed from data sets with a low EPV are likely to produce biased estimates.[120] Three studies did not provide definitions for outcome measures.[110,114,115] Having a predefined outcome reflecting a clinically significant and relevant health state limits the potential of bias.[134]

Strengths and limitations

The main strength of this systematic review is that, to our knowledge, it is the first to summarise the evidence presented by prediction models for AJCC stage I melanoma. The review followed procedures documented by the Cochrane Collaboration for conducting systematic reviews, the CHARMS for extracting data, the PRISMA guidelines for reporting and the PROBAST for assessing risk of bias; therefore, it is robust. We had planned to assess the overall quality of evidence using GRADE. Although the tool has been adapted for assessing overall quality in prognostic studies, it was not found to be suitable for prediction models.[104] We conducted comprehensive searches of bibliographic databases and grey literature. All stages of the review, including screening, data extraction and the risk-of-bias assessment, were conducted by either two researchers in duplicate or by one researcher with checks by a second researcher.

As with most systematic reviews, the main limitation was the quality of the published studies. None of the studies reported having followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD).[135] First, however, eight of the studies in the review were published between 2001 and 2014, before TRIPOD was published; therefore, they could not follow these guidelines. Second, only a small number of models were externally validated, making it difficult to determine external validity. Third, model comparisons and meta-analyses were problematic because of the variety in the predictors and statistical measures used for model performances.

Impact and implementation

The results of this systematic review highlight the relative lack of appropriate evidence levels underpinning current melanoma follow-up practice in AJCC stage I disease. As all of the studies were judged to be at high risk of bias, they cannot be recommended for estimating the probability or risk of recurrence, new primary tumours or metastases in routine clinical practice. This review identifies the need for ongoing development of risk prediction models that encompass known patient and tumoral variables in conjunction with new and developing prognostic biomarkers. These models must be developed in a way that follows biomarker development and reporting guidelines, such as TRIPOD, thus enabling appropriate critical appraisal and further assessment in the general AJCC stage I melanoma population.

The results of this review clearly outline a need for ongoing biomarker and prognostication studies and, therefore, should act as an evidence base and catalyse project development and funding. The review also demonstrates the potential impact of such studies on future follow-up guidelines and management of patients, given the relative scarcity of evidence-based practice at present.

Conclusion

This review identifies prediction models to predict the recurrence, new primary tumours and metastases in early-stage melanoma. However, they were all rated as having a high risk of bias and, therefore, cannot be recommended for use in clinical practice. The data elements most commonly used in these tools are patient demographic information or histological features of the primary tumour.[136] However, these data do not offer a wide enough scope of information to allow accurate prognostication of melanoma, which is heterogeneous in its biology and progression. Numerous biomarkers have been identified in recent years, with varying degrees of validation in clinical cohorts.[137] These offer greater potential to prognosticate at the individual patient and tumour levels, thus facilitating individualised follow-up and treatment regimens for patients.

Chapter 5 The diagnostic accuracy of high-resolution ultrasonography, fine-needle aspiration or core biopsy to detect recurrence and locoregional metastases during surveillance

Brief overview

After excision of a primary cutaneous melanoma, the disease may relapse, initially presenting as a local recurrence or satellite lesion (skin or subcutaneous lesion) within 2 cm of the primary tumour. In-transit metastasis may occur following progression through the lymphatic system with a lesion developing > 2 cm from the original site. Further progression of disease through the lymphatic system to the regional lymph nodes is known as regional recurrence. Progression to non-regional lymph nodes, then to organs and distant sites is known as metastatic melanoma. Lymph node metastases are believed to be an important prognostic factor for stage I and II melanoma patients.[138,139] Detection of metastases in the lymph nodes is initially by palpation of the affected nodes, followed by confirmation using sonography.

As described in *Chapter 2*, surveillance strategies for recurrence vary across countries by frequency and by the types of monitoring tests utilised. The range of approaches routinely used for detection of recurrences or new primary tumours and metastases may include clinical examinations, such as medical history, skin and lymph node examination and palpation. Further testing on suspicion of melanoma may involve imaging techniques or biopsy, such as ultrasonography of the abdomen, resected tumour scar, lymphatic drainage areas and regional lymph nodes; chest radiography; CT; MRI; PET; PET-CT; skin biopsies; and SLNB or blood tests. However, guidance published by the BAD advises that routine tests, including CT and blood investigations, are not recommended for staging of asymptomatic patients with stage I or II primary melanomas, because true-positive rates are low and false-positive rates are high.[27]

In practice, ultrasonography and fine-needle aspiration cytology (FNAC) (a minimally invasive procedure) or core biopsy are used when an enlarged lymph node is detected during follow-up visits. Ultrasonography may also be used for the detection of non-enlarged metastatic lymph nodes, which may not be palpable, and partially metastatic lymph nodes.[140] Importantly, ultrasonography may distinguish between benign and malignant palpable nodes, but is unable to detect micrometastases. If micrometastases are suspected in the regional nodes, then high-resolution ultrasonography and FNAC may be used preoperatively to replace SLNB.[141,142]

Sensitivities of between 5% and 89.4% have been reported for ultrasonography and/or FNAC in melanomas.[141,143–145] Specificities are higher, ranging from 84% to 100%.[146] This variability may be explained by heterogeneity between patients with melanomas of different stages in the populations of these studies. However, the findings are comparable to meta-analyses of ultrasonography and FNAC for surveillance of lymph nodes in melanoma, which report overall sensitivities and specificities of 96% and 99%, respectively,[147] for ultrasonography and of 97% and 98%, respectively, for FNAC.[148]

This systematic review supplements and updates these reviews by focusing on high-resolution ultrasonography for detecting satellite, in-transit and locoregional lymph node metastases during surveillance or in symptomatic patients with an initial diagnosis of stage I melanoma.

Research aim

The aim of this systematic review was to determine the diagnostic performance of high-resolution ultrasonography with or without FNAC for surveillance and follow-up to detect recurrence in patients who have had AJCC stage IA or IB melanomas surgically excised.

Specifically, the focus was on detection of:

- local recurrence or satellite metastases within 2 cm of the surgical scar of the primary tumour
- new primary melanomas – in-transit metastases occurring on the skin or subcutaneous layers that are > 2 cm from the primary lesion, but not beyond the regional nodal basin and lymph node metastases
- regional recurrence in local lymph nodes
- metastatic melanoma.

Methods

This review adheres to the guidelines for the PRISMA statement to ensure transparency of the process.[78] A protocol for the whole project, of which this review is part, is published on PROSPERO (CRD42018086784).[79]

Search strategy

The search strategy was designed by an experienced information specialist, in consultation with the project team, and was based on previous scoping of the literature. The search was designed on MEDLINE (via Ovid), using a combination of controlled medical subject heading thesaurus terms, relevant keywords and text words to identify studies of early-stage, stage I and stage II cutaneous melanoma patients and diagnostic tests. The search strategy contains appropriate use of stemming for alternative word endings, alternative spellings and plurals, and was translated to conform to other bibliographic databases using the following topic outline: [melanoma] AND [ultrasound OR biopsy] AND [surveillance]. Database-specific thesaurus headings, along with title and abstract keywords, were used and translated to other databases, altering the thesaurus headings and search syntax as appropriate. Terms relating to ocular melanoma were excluded. Identification of a suitable diagnostic filter on which to limit database results was researched seeking reliability and consistency in performance.[149] The diagnostic search filter chosen and used was published, validated and adapted for use in other databases as necessary [the study developed three search strategies (A–C), from these strategy A was selected as this was the most sensitive strategy identified].[150] The searches of the databases (see *Box 5*) were run on 4 April 2019 and updated on 3 July 2019. The search strategy used in MEDLINE can be found in *Appendix 5*. Following deduplication, publications were limited from 1998 to July 2019, as this is when SLNB became clinically utilised as a prognostic indicator.[100]

A grey literature and guidelines search was conducted to identify further relevant material not retrieved by the database searches (*Box 6*).

BOX 6 Sources searched for grey literature for diagnostic accuracy review

- Open Grey (www.opengrey.eu/; accessed 17 April 2019).
- Cancer Research UK (www.cancerresearchuk.org/about-cancer/find-a-clinical-trial; accessed 17 April 2019).
- Melanoma UK (www.melanomauk.org.uk; accessed 17 April 2019).
- National Guideline Clearinghouse (www.guideline.gov/; accessed 17 April 2019).
- Organisational websites: BAD, British Skin Foundation, The King's Fund, etc.
- Ongoing trials identified using the World Health Organization's International Clinical Trials Registry Platform (www.who.int/trialsearch; accessed 17 April 2019).

A range of guidelines on melanoma from a range of countries were also identified, and the evidence supporting relevant recommendations relating to the diagnosis of recurrent disease was checked.

Inclusion and exclusion criteria

Population characteristics

Studies were included if they considered one of the following:

- adults aged ≥ 18 years treated for AJCC (eighth edition) stage I cutaneous melanoma [stage IA (≤ 0.8 mm thick without ulceration) or stage IB (< 0.8 mm thick, or < 1 mm thick and ulcerated skin)][19]
- adults aged ≥ 18 years treated for AJCC (seventh edition) stage I cutaneous melanoma [stage IA (T1a ≤ 1 mm thick) or stage IB (T1b with ulceration or mitoses ≤ 1 mm thick, or T2a 1.01–2.00 mm thick and no ulceration)].[20]

If more than one study was reported by the same centre or institution, where participant populations could have been duplicated, we included either the most recent publication or the publication with the most complete participant data. Studies of patients with any stage of melanoma were included if the data for stage I disease were available independently. If the patients' disease stage was not clear, then authors were contacted for further information.

Target condition

The target conditions were local recurrence, satellite lesions, new primary lesions, in-transit metastases, locoregional lymph node metastases or metastatic melanoma after resection of a stage I melanoma.

Index/comparator tests

The detection of local melanoma recurrence, satellite lesions, new primary lesions, in-transit or locoregional lymph node metastases, or metastatic melanoma could include the following tests independently or in combination:

- ultrasonography of the resected tumour scar, lymphatic drainage area or regional lymph nodes
- FNAC/fine-needle biopsy (FNB).

All index tests should be confirmed by the use of an independent reference standard.

Reference standards

In patients testing positive for local recurrence, satellite lesions, new primary lesions, in-transit metastases and/or locoregional lymph node metastases, or metastatic melanoma, the reference standard was taken to be:

- histopathology results from excision biopsy, incision biopsy, wide local excision, punch biopsy, shave biopsy or core biopsy
- histopathology results from lymph nodes or distant secondary sites sampled by lymph node dissection, SLNB or core biopsy
- clinical follow-up when histopathology was not available.

Outcomes

Conventional outcomes for assessment of diagnostic test accuracy were extracted from each eligible study. These included the sensitivity, specificity, likelihood ratio, diagnostic odds ratio, false-positive rate and summary receiver operating curve (ROCs) of the index and reference tests. The data required to derive these parameters were the number of true-positive, true-negative, false-positive and false-negative cases reported in each included study for index and reference tests. If data were missing for a full 2 × 2 table, we contacted the authors of the article. Data were extracted from studies at the time point of diagnosis of lesions under consideration and before commencement of any treatments.

For indeterminate values (e.g. atypical, suspicious, probable or possible malignant lesions for which the test did not provide a clear negative or positive result, or values were missing in reported test results), the sensitivity and specificity were calculated for each scenario.[151]

Settings

Studies from any country and all settings were eligible, including primary, secondary and tertiary care. Non-English-language articles were translated when resources were available.

Study designs

Primary studies eligible for inclusion were randomised trials, prospective or retrospective cohort designs, cross-sectional studies, and diagnostic case–control studies with separate diseased and non-diseased groups. Studies had to report sufficient data to enable us to construct 2 × 2 tables of diagnostic metrics. Eligible studies had to have participants receive a single index test and a reference standard, or receive more than one index test and a reference standard.

Data collection

Study selection

Titles and abstracts of articles retrieved from the search strategies were screened by three reviewers using Rayyan.[81] Studies selected by two of the three reviewers were included for review. At this stage, non-English-language papers were included from titles and abstracts. Full-text articles were obtained for included citations. If translation of full-text non-English-language papers was not possible, these papers were coded as 'non-English studies awaiting assessment.'

Data extraction

Data were extracted from each study by one reviewer and independently verified by a second reviewer. Any disagreements were resolved by discussion between the reviewers or with arbitration from a third reviewer. Data were extracted from each included study based on the STAndards for the Reporting of Diagnostic accuracy studies (STARD) checklist.[152] The type of data extracted is outlined in *Report Supplementary Material 1*.

Risk of bias

The QUADAS-2 tool was used to assess the risk of bias of diagnostic accuracy studies.[153] This tool covered four domains: summarising the review question, tailoring the tool, constructing a flow diagram for the primary study, and judging bias and applicability.

Data analysis and synthesis

Data from the 2 × 2 tables were used to calculate sensitivity and specificity for each study. Individual study results were presented graphically by plotting estimates of sensitivities and specificities in forest plots. If more than one threshold was reported, data from one threshold were to be chosen to be incorporated into a meta-analysis. Meta-analysis of pairs of sensitivity and specificity values was planned using a bivariate random-effects approach. This approach would enable the calculation of summary estimates of sensitivity and specificity, while correctly dealing with the different sources of variation.[154] The planned meta-analysis was not conducted, as each study used different index tests and would be used at a different stage in the diagnostic work-up.

Sources of heterogeneity in studies

Risk factors for melanoma progression and prognosis obtained from demographic data in the studies were to be included as potential sources of patient heterogeneity between studies. A range of potential factors influencing heterogeneity may include sex, age of participants, tumour characteristics, presence of ulceration, stage of disease, site of primary tumour (trunk, lower limbs, upper limbs, head or neck, hand or foot), clinical node status at follow-up (post operative), sentinel lymph node status at follow-up (post operative)

and comorbidities. Potential sources of heterogeneity in reference tests may include the experience of surgeons, the method of sampling (either cytology or biopsy) and preparation techniques of the tissue sample [e.g. fresh-frozen section (cryosection) or paraffin section]. Heterogeneity of the index tests may arise from variations in the clinical pathway, including disease stage; professionals involved (e.g. radiographers, radiologists or clinicians); experience of operators performing index and reference tests; differing frequencies of ultrasonography instruments, manufacturers and models; and tumour depth and spread, ulceration, regions of interest for ultrasonography and anatomical sites of lesions.

Heterogeneity was to be investigated in the first instance through visual examination of forest plots of sensitivities and specificities and through visual examination of the ROC plot of the raw data. Heterogeneity was also to be assessed statistically using the I^2 statistic.[155] Had suitable data to meta-analyse been identified, sensitivity analyses exploring heterogeneity would have been conducted. Likewise, subgroup analyses by sample size of the study would have been conducted, as sample size is known to influence sensitivity and specificity.[156,157]

Results

Number of studies identified

Afer deduplication, the electronic database searches retrieved 2226 records and a further 24 from search updates. One additional study was included from reference lists of systematic reviews. From screening these citations, 106 primary studies, including conference abstracts, were identified as potentially relevant. Full-text articles were then retrieved for review. Five studies were non-English language (written in French, Hungarian, Japanese, Russian and Spanish). Seventeen citations were conference abstracts; each was checked for journal publication of the full reports. One foreign language paper was unable to be translated.[158] Study authors of five studies that reported combined data across stages were contacted for more information.[24,159–162]

Following this process, two English-language studies met the inclusion criteria and were assessed.[163,164] These reported the diagnostic accuracy of FNB[163] and the diagnostic accuracy of high-frequency ultrasonography.[164] We wrote to five study authors to seek further clarification and a breakdown by stage I disease to attempt to include further data.

In addition, other than the five studies awaiting classification and one needing translation, 98 full-text studies or conference abstracts that did not meet the inclusion criteria are recorded in *Appendix 6*, with reasons for exclusion. Studies were excluded based on the following: no analysis by disease stage or Breslow thickness ($n = 29$), stage not reported ($n = 22$), preoperative staging ($n = 13$), stages combined ($n = 10$), insufficient patients or data ($n = 5$), not diagnostic accuracy study ($n = 4$), no relevant outcomes ($n = 3$), advanced stage ($n = 3$), no diagnostic test ($n = 1$), letters ($n = 3$), review ($n = 3$), treatment ($n = 1$) and animal study ($n = 1$). Studies could meet one or more of these criteria. The PRISMA flow diagram in *Figure 5* outlines the study selection process and the reasons for exclusion.

Characteristics of included studies

One of the two included studies was conducted in Australia,[163] and the other in Germany.[164] The two studies considered different diagnostic tools used at different points in the care pathway. The characteristics of each study are presented in *Table 10*.

Doubrovsky *et al.*[163] evaluated the diagnostic accuracy of FNB, which is analagous to FNAC, to detect metastatic melanoma. They used a retrospective cohort study design, with data collected between January 1992 and December 2002 from the Royal Alfred Hospital in Sydney, NSW, Australia. The sample comprised 1582 confirmed FNBs at melanoma stages I–IV. A total of 323 confirmed cases (20%, 323/1582) were included in an analysis of stage I disease. Males accounted for 63% of

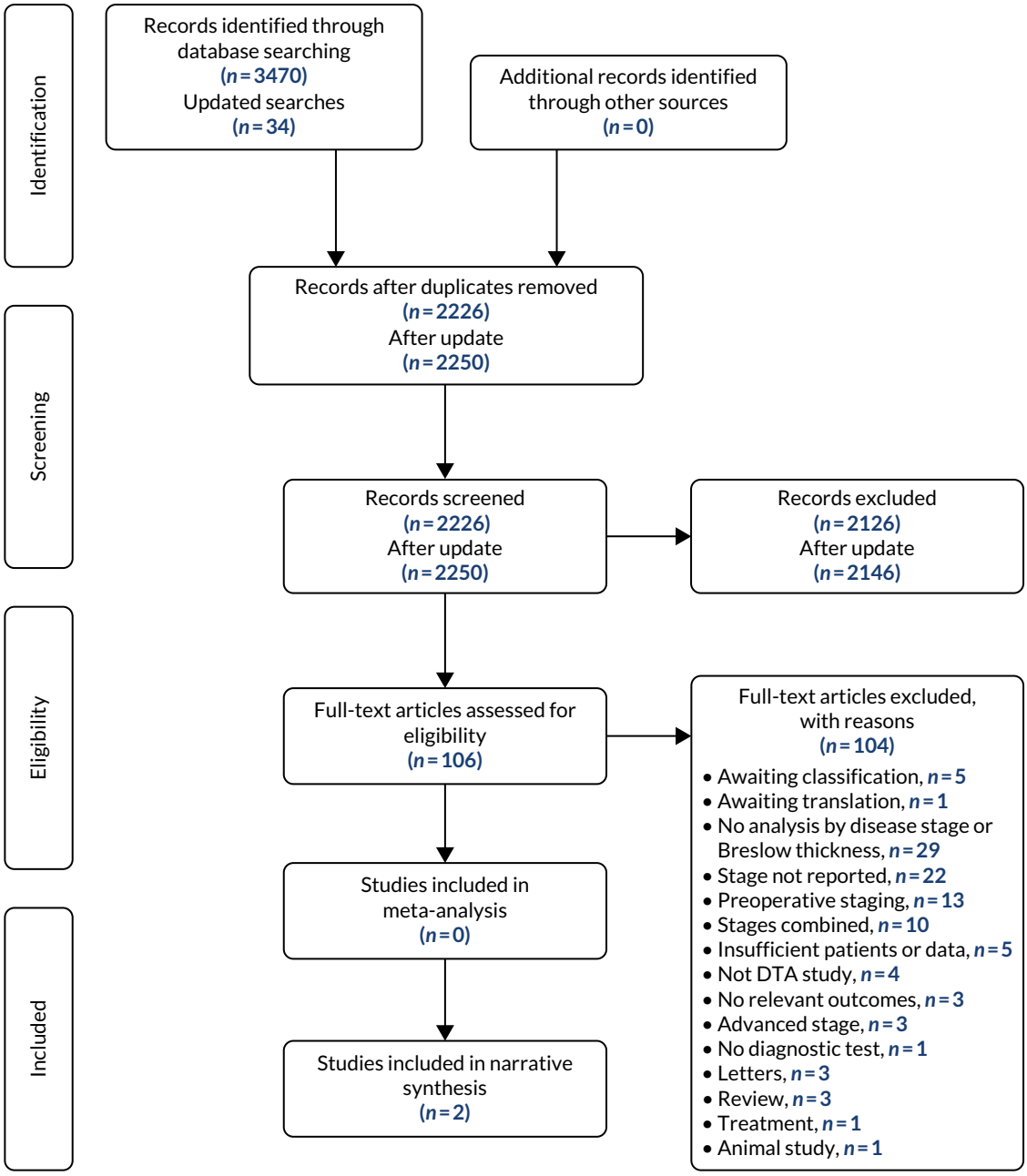


FIGURE 5 The PRISMA flow diagram for the diagnostic accuracy review. DTA, diagnostic test accuracy.

participants and ages ranged from 10 to ≥ 81 years. Diagnostic accuracy was evaluated by AJCC stage, the anatomical sites of lesions, use of immunostaining, year, sex, age, FNB attempts, needle size, presence of necrosis, pathologist case load, primary Breslow thickness, primary lesion ulceration status, primary lesion mitotic rate, histological subtype and predominant subtype.

Krüger *et al.*[164] evaluated the diagnostic accuracy of high-resolution ultrasonography of the lymph nodes to detect locoregional lymph node metastases. A prospective cohort design was used, with data collected prospectively between 2004 and 2008 from the Dermatology Department in Göttingen, Germany. Participants recruited had stage I–IV disease (AJCC 2002), and a median age of 58 years. Forty-eight per cent were male. The diagnostic accuracy of combined clinical and sonographic examinations was evaluated in 433 patients. The sample was composed of 1314 investigations in patients with melanoma stages I–IV. A total of 669 investigations (51%, 669/1314) were included in an analysis of stage I disease. An average of three paired investigations (clinical and sonographic) were

TABLE 10 Description of included studies for the diagnostic accuracy review

Study (first author and year of publication), country and design	Index test and definitions	Reference test and definitions	Number of patients at baseline	Age (years)	FNB or ultrasonography of patients at stage I	Duration of follow-up	Outcome
• Doubrovsky *et al.*[163] 2008 • Australia • Retrospective consecutive cohort	FNB with results read and reported by cytopathologists Test categories: positive, suspicious or negative for metastatic melanoma • Positive: specimens had sufficient, well-preserved malignant cells with typical cytological features and pigment, and/or were confirmed by immunochemistry • Suspicious: specimens with small numbers of atypical or poorly preserved cells, and cells lacking specific features of melanoma ○ Or insufficient sample • Negative: contained no material that was diagnostic of metastatic melanoma	a. Histopathology of excised lesion b. Clinical follow-up when lesion excision not warranted, and histopathology thus not available • Categories: positive, suspicious, or negative for metastatic melanoma	• $n = 1416$ • Average number of FNBs per patient = 1.56	Range: 10 to ≥ 81	• Number of patients NR • Number of FNBs at stage I = 400 • Confirmed FNBs, $n = 323$	≥ 6 months	TP, FN, TS, FS, FP, TN, sentinel node

continued

TABLE 10 Description of included studies for the diagnostic accuracy review (*continued*)

Study (first author and year of publication), country and design	Index test and definitions	Reference test and definitions	Number of patients at baseline	Age (years)	FNB or ultrasonography of patients at stage I	Duration of follow-up	Outcome

• Krüger *et al.*[164] 2011
• Germany
• Prospective consecutive cohort
	• Same-day paired clinical examination and ultrasonography
	• Presence of lymph node metastases by ultrasonography:
	  ○ TP: positive by ultrasonography, positive by histopathology
	  ○ FN: negative by ultrasonography, positive by histopathology or during clinical follow-up
	  ○ FP: clinically suspicious findings or positive ultrasonography result followed by negative histopathology result
	  ○ TN: no material classified as metastatic melanoma
	  ○ TS: test positive by histopathology
		a. Histopathology of lymph node biopsy
		b. Clinical follow-up when histopathology not available
			• $n = 433$
			• Mean number of investigations per patient = 3.08 ± 2.28
				• Mean 55.49 ± SD 17.21
				• Range: 13–95
					• 400 paired investigations (clinical and ultrasonography)
					• Stage IA–IB: 323 confirmed by histopathology or cytopathology
						• Mean follow-up (days): 363.0 ± SD 318.4
						• Intervals: mean 166 days
						• Equivocal or uncertain examination findings were re-examined by ultrasonography within 3 or 4 weeks
							TP, FN, FP, TN

FN, false negative; FP, false positive; FS, false suspicious; NR, not reported; TN, true negative; TP, true positive; TS, true suspicious.

performed for each participant on the same day. With respect to stage I disease, we estimated that 223 participants received 669 paired investigations. Diagnostic accuracy was evaluated by AJCC stage, melanoma subtype, lymph node dissection status and lymph node surgery before investigation.

Description of index tests

In the study by Doubrovsky *et al.*,[163] FNBs were performed by the reporting cytopathologist or a supervised trainee pathologist. A hollow bore needle was inserted directly into the lesion (once localised and stabilised) and the sample was retrieved. The specimen was then transferred directly onto glass slides. One slide was stained with Diff-Quik (MICROPTIC S.L, Barcelona, Spain) and the other stained using the Papanicolaou staining method. Residual material was retained for subsequent testing if necessary. Air-dried slides were assessed by a cytopathologist immediately after the sampling and staining procedure.

Krüger *et al.*[164] performed high-resolution ultrasonography using a 7.5- to 10.0-MHz real-time scanner. Paired, non-blinded clinical and ultrasonography investigations were performed for each patient. Each clinical examination included a medical history, physical examination, and inspection and palpation of the primary tumour scar. Scanning of the regional lymph nodes was standardised, with investigations performed in longitudinal and transverse sections. The morphological criteria of the lymph nodes (including size, shape, echogenicity of the centre and cortex) were evaluated. No information was reported on the expertise of staff who conducted these tests.

Interpretation of index tests

Doubrovsky *et al.*[163] looked at FNAC alone as an index test. FNB findings were categorised by Doubrovsky *et al.*[163] as positive, suspicious or negative for metastatic melanoma. Samples classified as suspicious for metastatic melanoma were composed of cells from unclassified/unspecified malignancies, or cases categorised as suspicious for melanoma that had a small number of atypical cells, poorly preserved cells or cells lacking specific features of melanoma. When verified, these were reclassified as either true suspicious or false suspicious.

Krüger *et al.*[164] looked at ultrasonography alone as an index test. Ultrasonography findings were considered suspicious for malignancy by Krüger *et al.*[164] when at least one of the following criteria applied: the Solbiati[165]/Vasallo[166] index was < 2, the whole lymph node structure had a predominance of low echogenicity, the lymph node centre had low echogenicity, or there were asymmetrical regions with low echogenicity in the lymph node margin.

Description of reference tests

In the study detailed by Doubrovsky *et al.*,[163] histopathological evaluation of the excised lesion was the reference test used to confirm metastasis. When histological material was not available, follow-up was used as the reference test.

In the study detailed by Krüger *et al.*,[164] lymph nodes were removed by excisional biopsy for histopathological assessment of lymph node sections.[164,167] The definitive reference test was histopathology of the excised lymph node, but, in some cases, fine-needle aspiration was also performed under sonographic guidance.[164] When clinical and ultrasonography tests were negative, follow-up was used as a further reference test.

Interpretation of reference tests

When evaluating the diagnostic performance of FNB, Doubrovsky *et al.*[163] used two reference standards. Histopathology was the reference standard in 1120 of the 1582 cases (71% of cases) and follow-up for ≥ 6 months was the reference standard in the remainder ($n = 462$, 29% of cases). Follow-up was relied on as the reference standard when histopathology was not appropriate.[163] Instances when histopathology was indicative of metastatic melanoma were not described in the study report.

When evaluating the diagnostic performance of clinical examination and lymph node ultrasonography, Krüger *et al.*[164] also used two reference standards. The reference standards were the removal of lymph nodes by excisional biopsy for histopathology of lymph node sections,[164] or follow-up among those identified as clinically negative to identify false-negative cases as follows:

- negative on sonography based on clinically suspicious lesions subsequently proven to be malignant by histopathology or
- negative based on lack of clinical suspicion and sonography (not histologically testable owing to lack of suspicious region/mass), with metastases identified during follow-up.

In the study by Krüger *et al.*,[164] reference standard techniques and instances when histopathology was indicative of the target condition were not reported with clarity.[164]

Risk of bias

Risk-of-bias assessments were performed using the QUADAS-2 tool across four domains.[153] The results of this are summarised in *Table 11*.

Methods of patient selection

The risk of patient selection bias was considered to be low for each study, although consecutive recruitment was reported for Doubrovsky *et al.*[163] only. Patients were identified over specified times either retrospectively[163] or prospectively.[164] There were no concerns regarding the included patients and settings not matching the review question in either study (both studies were deemed to have a low risk of bias regarding applicability).

The index test and how it was conducted and interpreted

The conduct of the index test was judged as being at low risk of bias in the study completed by Doubrovsky *et al.*[163] because sampling and interpretation of the index test was completed while blinded to the reference standard outcome. The reading of the index test (cytology slide interpretation) was completed by the cytopathologists who performed the FNB and prepared the cytology slides.[163] However, this is standard practice and unlikely to have promoted bias.

TABLE 11 Risk-of-bias assessments for the diagnostic accuracy review

Risk-of-bias domain (corresponding subchapter)	Study (first author and year of publication)			
	Doubrovsky *et al.*[163] 2008		Krüger *et al.*[164] 2011	
	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns
Patient selection[a] (5.4.3.1)	Low	Low	Low	Low
Index test[b] (5.4.3.2)	Low	Low	Low	Low
Reference standard[c] (5.4.3.3–5.4.3.4)	High	Low	High	Low
Flow and timing[d] (5.4.3.5–5.4.3.6)	High	Low	High	Low

a Doubrovsky *et al.*:[163] consecutive recruitment over 10 years; Krüger *et al.*:[164] prospective, consecutive sample of patients with a single melanoma.
b Doubrovsky *et al.*:[163] samples were taken and assessed by the cytopathologist, unblinded; Krüger *et al.*:[164] paired, non-blinded clinical and sonographic investigations.
c Doubrovsky *et al.*:[163] histopathology and clinical follow-up in those index test negative; Krüger *et al.*:[164] histopathology and clinical follow-up in those index test negative. Both studies utilised a divergent reference test based on index test outcome.
d Doubrovsky *et al.*:[163] index and reference tests conducted on the same day, but not for the large proportion of participants followed up clinically; Krüger *et al.*:[164] time interval between index and reference tests was not reported, but could have been long for those followed up clinically. Clinically negative investigations in a separate subgroup of the Krüger *et al.*[164] study were evaluated separately and not reported in the results.

The conduct of the index test in the study completed by Krüger *et al.*[164] was judged to be at low risk of bias as ultrasonography interpretation was completed while blinded to the reference standard. Furthermore, although sonographers completing the test made judgement on the outcome of the test (standard practice), prespecified thresholds for ultrasonography were outlined a priori and followed.[164]

Describe the reference standard and how it was conducted and interpreted

In the study by Doubrovsky *et al.*,[163] classification of the reference standard (histopathology following surgical resection) for the target condition was adequate and blinded appropriately. Krüger *et al.*[164] used histopathology as the reference test, but did not clearly report the technique used.

There were no concerns over whether or not the reference standards were identifying the target condition as defined by the review question (both studies were deemed to have a low risk of bias regarding applicability).[163,164] However, both studies used clinical follow-up as a second reference standard; although this is, realistically, unavoidable among those with no evidence of the target condition after index testing, it is a less reliable reference test, and its use based on the index test outcome may introduce differential verification bias.[163,164] For both studies, the definition used for a metastasis-free interval is likely to strongly affect the number of false-negative results, and the specificity of the investigations.[163,164] Given this, both studies were deemed to have a high risk of bias.

Flow and timing

The time interval between index and reference tests was not reported in the study by Krüger *et al.*,[164] but could have been long for those undergoing clinical follow-up. However, the study did adhere to the melanoma guidelines of the German Dermatological Society.[167] All 669 investigations conducted in patients with stage I melanoma were accounted for in the results.[164] SLNB or complete lymph node dissection was conducted in 272 out of 433 (63%) patients (stage of disease not reported) before clinical and sonographic follow-up at a mean interval of 166 days (± 101 days).[167] This surgery is unlikely to have been offered to patients with stage IA melanoma. Clinically negative investigations in a separate subgroup of the Krüger *et al.*[164] study were evaluated separately and not reported in the results; for this reason, this domain was judged to have a high risk of bias.

In the Doubrovsky *et al.*[163] study, clinically suspicious lesions were detected by palpation or by imaging techniques. These were then investigated by FNB. No other intermediate interventions were reported.[163] Of the whole study sample reported by Doubrovsky *et al.*,[163] 462 individuals had clinical follow-up as the reference standard.

The time interval between index and references tests in Doubrovsky *et al.*[163] was within the same day. However, not all patients received the reference standard, as some were followed up clinically; for these individuals, the time interval could have been long. For these reasons, this domain was judged to have a high risk of bias.[163]

Diagnostic accuracy

A meta-analysis was not conducted as there was only one study for each index test. The results of the studies by Doubrovsky *et al.*[163] and Krüger *et al.*[164] for diagnostic accuracy of FNB and ultrasonography, respectively, are shown in *Table 12*. Forest plots are also presented for the sensitivity and specificity of the index tests for the two studies (*Figures 6* and *7*).

Fine-needle biopsy

Doubrovsky *et al.*[163] excluded suspicious findings, but stated which were true suspicious and which were false suspicious.[163] We assumed that test results classified as true suspicious by Doubrovsky *et al.*[163] were true positives and, similarly, that false-suspicious test results were false positives.[163] Sensitivity was estimated as 0.94 (95% CI 0.90 to 0.97) and specificity as 0.95 (95% CI 0.90 to 0.97). When excluding suspicious findings from the analysis, the results were similar (see *Figure 6*).

TABLE 12 Diagnostic accuracy results of the included studies

Study (first author and year of publication)	AJCC stage I melanoma	Follow-up	True positive (*n*)	False negative (*n*)	True suspicious (*n*)	False suspicious (*n*)	False positive (*n*)	True negative (*n*)	Sensitivity (95% CI)	Specificity (95% CI)	Likelihood ratio (95% CI)	
											Positive	Negative
Doubrovsky *et al.*[163] 2008; FNB	Base case • Unit = FNB • *n* = 400 • Confirmed, *n* = 323	• Mean 50.2 months • Median 45.7 months • Range 6–144 months	150	9	0	0	9	155	0.94 (0.90 to 0.97)	0.95 (0.90 to 0.97)	17.2 (9.1 to 32.5)	0.06 (0.03 to 0.11)
	Excludes suspicious histopathology, *n* = 295	As above	128	9	22	6	3	155	0.93 (0.88 to 0.97)	0.98 (0.95 to 1.00)	49.2 (16.0 to 151)	0.07 (0.04 to 0.13)
Krüger *et al.*[164] 2011; clinical and sonography	Base case • Patient-level analysis • *n* = 224	Mean ± SD (days): 363.0 ± 318.4	1	0	–	–	3	220	1.00 (0.03 to 1.00)	0.99 (0.96 to 0.99)	74 (24 to 229)	0.00
	Unit = paired investigations *n* = 669	As above	1	0	–	–	9	659	1.00 (0.03 to 1.00)	0.99 (0.97 to 0.99)	74 (39 to 142)	0.00

Study (first author and year of publication)	TP (*n*)	FP (*n*)	FN (*n*)	TN (*n*)	Sensitivity (95% CI)	Specificity (95% CI)
Doubrovsky *et al.* 2008[163] (base-case analysis)	150	9	9	155	0.94 (0.90 to 0.97)	0.95 (0.90 to 0.97)
Doubrovsky *et al.* 2008[163] (excluding suspicious findings)	128	3	9	155	0.93 (0.88 to 0.97)	0.98 (0.95 to 1.00)

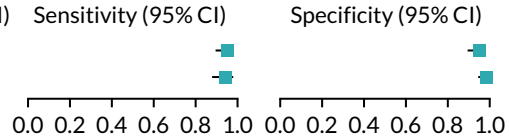


FIGURE 6 Forest plot of sensitivity and specificity of FNB for detection of stage I melanoma recurrence.[163] FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Study (first author and year of publication)	TP (*n*)	FP (*n*)	FN (*n*)	TN (*n*)	Sensitivity (95% CI)	Specificity (95% CI)
Krüger *et al.* 2011[164] (base-case analysis)	1	3	0	220	1.00 (0.03 to 1.00)	0.99 (0.96 to 1.00)
Krüger *et al.* 2011[164] (by investigations)	1	9	0	659	1.00 (0.03 to 1.00)	0.99 (0.97 to 0.99)

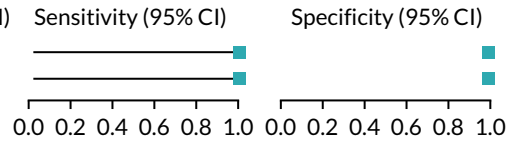


FIGURE 7 Forest plot of sensitivity and specificity of ultrasonography for detection of stage I melanoma recurrence.[164] FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Ultrasonography

In Krüger *et al.*,[164] the unit of analysis was the number of investigations in 433 patients (an average of three investigations per patient). To estimate results at a patient level, true-positive, false-positive, false-negative and true-negative results were divided by three. Krüger *et al.*[164] reported 217 stage I patients in the sample; this approximation was equivalent to 224 patients. For sensitivity, both the investigation and the patient-level analysis were similar, with very wide 95% CIs (see *Figure 7*). This result was expected, as only one true positive (and no false negatives) occurred in the sample. Specificity was high in both analyses.

Discussion

Summary of findings

This systematic review sought to estimate the diagnostic accuracy of ultrasonography and FNAC procedures, routinely used to detect recurrence and metastases during follow-up of patients initially diagnosed with stage I cutaneous melanoma. Comprehensive literature searches were performed in a range of bibliographic databases from 1998 to July 2019, and were complemented by grey literature searches for guidelines and other potentially relevant literature. Searches were restricted to 1998, which coincided with the introduction of SLNB assessment following melanoma diagnosis in some centres internationally. Despite this extensive searching, only two studies met the inclusion criteria. As SLNB involves the assessment of an initial lymph node basin, and often the complete removal of all associated nodes when a positive SLNB is identified, this would probably have an impact on the routine clinician examination findings of patients, as well as the potential choice and application of further investigation modalities in such patients. By limiting the searches to dates after this change in routine management, a more appropriate and contemporaneous assessment of the evidence could be made.

The two identified studies considered different tests (lymph node ultrasonography and FNB).[163,164] The findings reported for diagnostic accuracy of FNB in patients who were diagnosed initially with stage I melanoma were comparable to those reported for patients at stages II–IV.[163] Doubrovsky *et al.*[163] reported that the sensitivity was positively correlated with the following factors: the use of immunostaining (currently widely accepted[168]); if the cytopathologists had performed > 500 FNBs; the case mix, especially in patients presenting with ulcerated primary melanomas; and lesions located in the skin and subcutis. False-negative findings were reported to be associated with masses located in the axillary lymph nodes and when sampling required more than one needle pass.

The diagnostic accuracy of ultrasonography for stages II–IV was reported to be comparable to that for stage I. Krüger *et al.*[164] found that the sensitivity and specificity varied between melanoma subtypes (e.g. superficial or nodular).

In line with Cochrane guidance[82] on the use of narrative summary of findings tables, we did not include one for the systematic review of diagnostic accuracy of high-resolution ultrasonography, fine-needle aspiration or core biopsy to detect local recurrence, satellite lesions, new primary lesions, in-transit metastases, locoregional lymph node metastases or metastatic melanoma during surveillance. We did not calculate pooled results in this review; instead we presented the results of a single study. As a result, a narrative SoF table would not include any certainty of evidence judgements; thus, we felt that it was an unnecessary addition. Furthermore, heterogeneity is expected in diagnostic test accuracy reviews; therefore, with just one study included, by reporting on each test, we cannot make any inferences from the results presented.

Comparison of findings with other reviews

Four reviews retrieved by our searches had investigated the diagnostic accuracy of high-resolution ultrasonography or FNAC/core biopsy. None of the reviews conducted analyses by disease stage. Hall *et al.*[148] identified 10 studies published between 1980 and 2007 for inclusion. The review reported

summary estimates for both palpation and ultrasonography-guided FNAC together as a sensitivity of 0.97 (95% CI 0.95 to 0.98) and a specificity of 0.98 (95% CI 0.98 to 1.00). The positive likelihood ratio was 58 (95% CI 23 to 139). These data are similar to those reported by Doubrovsky *et al.*[163] on FNB in stage I patients.

A second review summarised the diagnostic accuracy of different imaging techniques for primary staging and surveillance of lymph nodes and distant metastases.[147] This review included 74 studies, of which 21 considered ultrasonography. Ultrasonography was reported at primary staging and surveillance. Twenty-one studies that considered ultrasonography and which were published between 1990 and 2007 met inclusion criteria for the review. The mean QUADAS-2 score across the 21 included studies was 5.8, with a SD of $\pm$ 2.5, from a potential score of 14. Ultrasonography during surveillance imaging had a median sensitivity of 0.96 [95% credible interval (CrI) 0.85 to 0.99] and median specificity of 0.99 (95% CrI 0.95 to 1.00). Although no analysis was conducted by disease stage in this review, the results were similar to those reported by Krüger *et al.*[164]

Another review compared ultrasonography and palpation to detect nodal invasion in patients with stage I–III melanoma.[169] Twelve studies published between 1997 and 2003 met the inclusion criteria. Positive likelihood ratios were 41.9 (95% CI 29 to 75) for ultrasonography and 4.55 (95% CI 2 to 18) for palpation. Negative likelihood ratios were 0.024 (95% CI 0.01 to 0.03) for ultrasonography and 0.22 (95% CI 0.06 to 0.31) for palpation.

A Cochrane review of ultrasonography and other imaging techniques for staging and re-staging of adults with cutaneous melanoma, at any stage of initial disease, included 11 studies.[170] It reported a summary sensitivity for ultrasonography alone of 0.35 (95% CI 0.17 to 0.59) and a specificity of 0.94 (95% CI 0.86 to 0 98) for detection of regional lymph node metastases before SLNB. A combination of pre-SLNB ultrasonography with FNAC reduced sensitivity to 0.18 (95% CI 0.036 to 0.57) and increased specificity to 1.00 (95% CI 0.99 to 1.00). These findings are at odds with those reported by Xing *et al.*[147] and those reported by Krüger *et al.*[164]

Overall, it is likely that differences in sensitivity between the reviews may be due to heterogeneity between studies of mixed populations, for example patients at different stages of disease and variations in case mix of melanoma subtypes.

Strengths and limitations

This review sought to systematically identify all relevant studies looking at the diagnostic performance of tests used to detect recurrance after treatment for stage I melanoma. It followed a prespecified protocol and conducted rigorous searches. The methods for the review correspond to best-practice recommendations.

The review is limited in a number of respects. First, it considered only two diagnostic tests. The rationale for this was based on clincial opinion about which tests might be viable for the NHS to consider, but it also took into account judgements about which tests would not be relevant; for example, sophisicated imaging was excluded for the review. Although these tests are not recommended by many guidelines currently, whether or not those recommendations should change has not been assessed (although recent surveillance reviews by NICE suggest not, at least for stage I disease).[54]

The dates of the searches were limited from 1998 to July 2019. This date was chosen because this is when SLNB became clinically utilised as a prognostic indicator.[100] However, it is possible that relevant studies could have been published before this date. The applicability of the findings of any such study to current practice is unclear, as staging criteria have changed several times since 1998.

The key limitation of this review is the lack of recent evidence on the diagnostic performance of ultrasonography and FNB in the target population. Planned meta-analyses were not possible as the

two studies included investigated different index tests. Furthermore, Krüger *et al.*[164] identified only one participant initially treated for stage I disease as a true positive. Consequently, the CI for sensitivity was very wide (95% CI 0.03 to 1.00), indicating that the study is non-informative.

Conclusion

Few data were found on the diagnostic accuracy of FNAC or high-frequency ultrasonography for detecting recurrence or metastasis in patients initially diagnosed with stage I melanoma. This may be, in part, because the natural history of AJCC stage I melanoma results in relatively few patients developing metastatic spread of the condition, thereby limiting the scope for analysis and the relative need for research, compared with higher-risk groups. The data applicable to stage I disease that are available do not provide strong evidence to support the use of these tests in this target patient group. However, the consistency of findings with other reviews (often looking at a wider population) provides some reassurance.

Chapter 6 Economic evaluation

Introduction and aim

This chapter presents the methods and results of a cost-effectiveness analysis based on an economic model. This economic evaluation aims to compare alternative surveillance strategies provided by the NHS for people who have been treated for AJCC stage IA and IB melanoma in the UK.

Methods

Development of a cost-effectiveness model

The development of the cost-effectiveness model started with a targeted review of the literature (see *Appendix 7*), which identified 15 economic evaluation studies of potential relevance.[24,171–184] None of these directly addressed the study question posed here. Therefore, a de novo model was developed to compare potential surveillance strategies provided by the NHS.

Using the information gathered from the systematic reviews reported in *Chapters 3–5* and extensive dialogue with the clinical members of the study team, we developed a Markov microsimulation model to evaluate the cost-effectiveness of plausible surveillance options based on differing clinical specialty, interval between visits and duration of follow-up. We used the most relevant UK epidemiological statistics and best available data from the literature to populate the model, reflecting the clinical reality of care for these patients. This included the probability of self-detection, 'false alarms' that result in unscheduled emergency appointments and a pathway of care that included investigational diagnosis and any subsequent treatment.

Model overview

The cost-effectiveness of surveillance strategies for stage I melanoma patients was assessed using a Markov microsimulation model. A microsimulation model is a form of economic modelling whereby individuals are simulated through the model one by one, rather than as a cohort. Individual results are then stored and the experience of the cohort is obtained by aggregating the individual results.[185] Markov health states are used to simply describe a patient's health status. In cancer models, disease progression will occur and the progression of the disease is described by a set of health states, typically of increasing severity. These states are commonly described as tunnel states. The minimum period of time (defined as the cycle length) that a person can be in a state before moving to another state was taken as 1 month. A 1-month cycle length was chosen as surveillance strategies were based on monthly surveillance interval differences.

The decision to use a microsimulation model as opposed to a simpler cohort-based Markov model[186] was based on the properties of a microsimulation. In a microsimulation, the memorylessness restrictions of a cohort-based Markov model can be overcome. Building the Markov microsimulation model in TreeAge Pro 2019 R1.0 (TreeAge Software Inc., Williamstown, MA, USA) enabled us to keep track of the treatment history, time since last treatment and recurrence of simulated patients, which all have an impact on the cost and effect estimates. In a cohort-based Markov model, movement from a given state is not affected by how the person arrived in that state (the model has no memory of prior events or care).

The microsimulation model describes the care pathway of individuals from the point when they received treatment (i.e. wide local excision) for stage I melanoma. The model allows for alternative surveillance strategies to be compared. The model seeks to estimate their longer-term (i.e. lifetime)

costs and consequences, including those that might arise from any subsequent melanoma diagnosis, be it for new or recurrent disease. Surveillance strategies were compared based on lifetime costs and health outcomes, expressed in terms of quality-adjusted life-years (QALYs). These data were then used to estimate cost-effectiveness. The costs were estimated in Great British pounds from an NHS and Personal Social Service perspective for the financial year 2017–18. As both costs and QALYs were estimated over a lifetime time horizon, they were discounted using a 3.5% annual discount rate, as per NICE's reference case.[187]

We used anonymised individual data from 160 early-stage patients obtained from a local hospital (University Hospital of North Durham, ethics approval: NREC 19/NE/004) to populate the basic demographic characteristics (mean age 56 years, range 17–98 years; ratio of males to females 36% : 64%). Stage I patients (ratio of stage IA to stage IB was 76% : 24%) had up to 8 years' follow-up on recurrence (15%) and mortality (melanoma-specific mortality over that time was 6% and all-cause mortality was 17%). Any additional melanoma skin cancer statistics were taken from the Cancer Research UK website.[188]

Model structure

Once discharged into follow-up care, a patient enters into the model as disease free. Surveillance is captured as an event undertaken at discrete time intervals. At any given point in time, an individual is in one of four Markov states: disease free, recurrence (diagnosed stage IA–IV), death from melanoma or death from other causes. There is a chance that, when a recurrence occurs and is not detected in the monthly cycle, disease progression will occur. This is depicted as a tunnel state in *Figure 8*. The detailed model structure is shown in *Appendix 8*.

In the microsimulation model, each individual has a chance (probability) of recurrence, metastasis and occurrence of new primary tumours in a given monthly cycle. In the model, any future melanoma diagnosis can be self-detected in a cycle by an individual/partner. This will result in an unscheduled emergency visit. There is also a chance that the emergency visit is a 'false alarm' and that the patient does not have melanoma. If the individual was scheduled for a surveillance visit that month, there is a chance that the attending clinician will detect the melanoma. There is also a chance that a clinician will fail to diagnose a melanoma if one is present.

For all suspicious lesions, the diagnostic process starts with a history and physical examination with dermoscopy by a clinician (note that, although it is recommended by NICE that all examinations should use dermoscopy, anecdotally there is varied uptake across different specialist settings). The use of

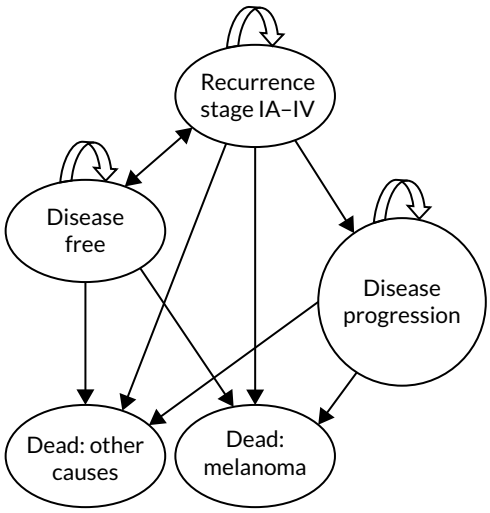


FIGURE 8 Description of a simplified version of the Markov microsimulation model.

subsequent investigations is determined by the findings of the prior investigations (i.e. if a local excision biopsy is positive, then SLNB will be performed; if SLNB is positive, then CT is performed).

If a cancer is present but is not detected during a 1-month cycle, then the cancer may progress. If this occurs, then, in the next cycle of the model, the individual moves into a tunnel health state. The tunnel state itself describes the AJCC staging of melanoma.

The whole process described here for the movement of an individual through the model is then repeated for each individual in the cohort, thereby generating individual life histories of the population modelled.

Plausible strategies

The development of the model strategies was based on an iterative consultation process with the local clinical expert team. Potentially thousands of strategies are possible based on the possible service provision over a patient lifetime.[25] We eliminated non-realistic strategies, such as those involving general practitioners (GPs), who do not have the skills, confidence or capacity to provide this service in the UK, let alone the support of specialist organisations (Dr Timothy Cunliffe, chair-elect, Primary Care Dermatology Society, November 2018, personal communication). Furthermore, it was established that:

- diagnostic imaging would not be considered as part of the surveillance regimen because the evidence from the USA[189] and the systematic review reported in *Chapter 5* did not support its use
- prognostic risk models had not been suitably validated to form part of a plausible surveillance strategy (see *Chapter 4*).

The three variables used to define variations between surveillance strategies were clinical specialty, surveillance intervals and duration of follow-up. In calculating the number of strategies for both stage IA and IB disease, different combinations of the above stratifying variables were used. Initially, 400 strategies were defined for stage IA and 600 for stage IB. The clinical team reviewed these strategies iteratively, and those considered as implausible were excluded from further consideration. Following this process, the number of plausible strategies was reduced to 75 strategies for stage IA and 75 for stage IB. To complete the strategy selection process, we added 12 additional low-resource strategies that might be options for people who had initially been treated for stage IB disease. Thus, for stage IB, a total of 87 strategies were included in the final model.

The clinical specialty options for conducting the surveillance are as follows: dermatologist; surgeon; and specialist dermatological nurse, also generically referred to as clinical nurse specialist (CNS) in the NHS. The realistic frequency of follow-up considered was follow-up every 3, 4, 6 and 12 months. The durations of follow-up considered for stage IA were 1, 2, 5 and 10 years. For stage IB, the durations of follow-up considered were 1, 3, 5, 10 and 20 years. We compared all strategies with current practice, based on that which has been recommended by the NICE 2015 guidelines[16] and the BAD 2010 revised UK guidelines[27] for the management of cutaneous melanoma. Both the NICE[16] and BAD[27] guidelines advise that patients who have stage I melanoma are followed up to detect signs of recurrence after history and examination. The NICE guidelines[16] recommend that patients with stage IA melanoma should be seen two to four times over 12 months following initial treatment, then discharged if no sign of recurrence or metastasis is found. Patients with stage IB melanoma should be seen every 3 months for 3 years, then every 6 months up to 5 years following initial treatment. At the end of that period, they should then be discharged if no sign of recurrence or metastasis is found.

Model inputs

To provide estimates of relative cost-effectiveness, the model required estimated values for a range of different types of parameters. We aimed to use the best available values derived in a systematic and reproducible manner to avoid bias caused by the distorted and selected use of data.[190] We focused on

identifying the most relevant data to the decision problem, which is the comparison of alternative surveillance regimens after treatment (i.e. local excision) for stage I melanoma.

We assembled the different types of data required for the economic model from analyses of existing data sets, a series of systematic reviews and focused searches for specific types of data (see *Appendix 9*). In brief, the broad types of data required to populate the economic model were as follows:

- patient behaviour with regards to surveillance and follow-up (e.g. self-detection/self-diagnosis and 'false alarms' resulting in emergency visits)
- clinical pathway once diagnosed with melanoma
- the performance of different regimens (e.g. clinical examinations) in terms of diagnostic accuracy (see *Diagnostic accuracy*)
- the prevalence, incidence and risk of progression of the disease, risk of recurrence [i.e. its epidemiology and natural history (see *Natural history of melanoma*)]
- resource use and unit costs required to estimate the costs of alternative surveillance regimens; the specific parameters and methods used to provide estimates that are relevant to the UK context (see *Resource use and unit costs*)
- health-state utilities (see *Health utilities*).

Patient behaviour

We based estimates of patient behaviour on data from the literature and on advice from local clinical experts. For example, the probability of self-diagnosis was based on the MELanoma Follow-Up (MELFO) study.[23] This study reported that 8 out of 17 (48%) recurrences among 180 patients in the 1-year time-frame were identified by self-diagnosis. Based on advice from the clinical experts in the study team, a value of 60% was used in the base-case scenario in the NHS setting based on their experiences. The probability of a 'false alarm' emergency visit occurring was based on an earlier Dutch study, which reported that almost 80% of patients (538/699) with a Breslow thickness of < 1 mm reported more frequent follow-up visits than the guideline recommends.[191] The clinical experts in the study team thought that an emergency visit would occur for between 80% and 90% of patients in a given year in their clinic; therefore, a yearly value of 85% was used. Annual probabilities of self-diagnosis and 'false alarm' emergency visits were converted into a monthly probability.

Clinical pathway

The 2015 NICE guidelines,[16] along with work from Wilson *et al.*,[182] was the starting basis of the clinical pathway modelled. As acknowledged in NICE's decision to update its guidelines,[16,192] new evidence has been identified that may change current recommendations. A plausible care pathway was modelled that used SLNB for staging; if positive, lymph node dissection was performed. The model did not include the ultrasonography-guided fine-needle aspiration option (because of lack of evidence; see *Chapter 5*), but did include the cost of treating advanced-stage disease with newer systematic targeted therapies and immunotherapies (Dr Janine Graham, South Tees Hospitals NHS Foundation Trust, June 2019, personal communication; see *Appendix 9*, *Table 24*, for list of clinical parameter values).

Epidemiology

National melanoma epidemiology statistics, such as summary stage of melanoma incidence and mean age at stage I diagnosis, were obtained from the National Cancer Registration and Analysis Service (NCRAS), which is now part of Public Health England (PHE) (NCRAS enquiries, February 2019, personal communication). The data on the natural history of melanoma were derived by expert elicitation among clinical experts from the UK.[193] The mean age of people diagnosed with AJCC stage I melanoma from 2000 to 2017 was 57.9 years (Charlotte Eversfield, PHE, NCRAS, personal communication).

How each set of data and the values used in the model were derived is described in more detail in the following sections.

Diagnostic accuracy

Diagnostic accuracy of practitioners

A Cochrane systematic review identified studies that described the accuracy of clinicians in identifying melanoma.[170] None of the studies identified in this review was from the UK. As part of the work conducted for the economic evaluation reported in this chapter, a meta-analysis of the diagnostic performance of different categories of staff was conducted using data from studies included in the Cochrane review. For dermatologists, data were pooled from 11 studies, all of which came from mainland Europe.[194–204] This gave a mean sensitivity of 0.875 (95% CI 0.784 to 0.931) and specificity of 0.893 (95% CI 0.792 to 0.949). For surgical oncologists, two studies (both from Italy) were pooled and gave a mean sensitivity of 0.886 (95% CI 0.795 to 0.940) and specificity of 0.734 (95% CI 0.688 to 0.775).[205,206] For specialist dermatological nurse or cancer nurse specialist, the only equivalent data were from a US study based on only eight physician assistants.[207] In that study, health-care professionals reviewed 65 dermoscopic images each; the physician assistants had a mean sensitivity of 0.800 (95% CI 0.590 to 0.930) and specificity of 0.470 (95% CI 0.320 to 0.640) (see *Appendix 9*, *Table 31–33*, for diagnostic accuracy values).

Diagnostic accuracy of staging disease

The sensitivity and specificity of local biopsy were obtained from a study that aimed to investigate how accurate and reproducible the results are of pathologists' diagnosis of melanocytic skin lesions.[208] A total of 240 skin biopsy cases were grouped into sets of 36 or 48 and were assessed by a randomised sample of US pathologists from 10 US states on two different occasions within a period of at least 8 months. The results of the paper indicate that 82.8% (95% CI 81.0% to 84.5%) of melanocytic skin biopsy diagnoses would have their diagnosis verified if reviewed by a consensus reference panel of experienced pathologists.

The accompanying review of the clinical evidence for the NICE guideline was the source of data for staging of melanoma.[16] For patients, the sensitivity of SLNB in identifying micrometastatic nodal/regional disease was estimated to be 86.6% (95% CI 84.6% to 88.4%), based on 47 studies with 19,607 data points. Specificity was 100%, as was reported in the review conducted for the NICE guideline. For advanced-stage disease, a meta-analysis assessed the clinical utility of various diagnostic tests for staging and surveillance of melanoma patients.[147] The 2015 NICE guideline recommended that CT staging be offered to people with stage III or suspected stage IV melanoma.[16] According to the meta-analyses, the median sensitivity of CT was 51% (95% CrI 24% to 76%) and specificity was 69% (95% CrI 30% to 92%) for staging distant metastasis. The corresponding CrIs were used in the model as beta distributions (see *Appendix 9*, *Tables 34–36*, for diagnostic accuracy values).[147]

Natural history of melanoma

Disease progression/transition probabilities

The natural history of undetected or untreated melanoma is unknown.[193] To get an estimate of the progression of melanoma within a certain time frame, expert elicitation was required. A novel approach to elicit expert opinion on the rate of progression from each stage to any other stage was developed and piloted on 14 clinical experts from the UK and Australia/New Zealand.[193] Participants were asked for their beliefs about the probability of progression from each of the starting stages stated (i.e. in situ to stage IV) to any other stage and death. Questions were asked in the format of: 'Imagine a cohort of 100 patients with stage X undiagnosed and hence untreated disease. After 6 months, the patients may be in any of the following stages'. Experts assigned probabilities using the quantile method, whereby median and upper and lower 95% CrIs were elicited. Wilson *et al.*[193] fitted a modified Connor–Mosimann (mCM) distribution to the elicited quantities from each expert. The mCM distribution is a generalisation of the Dirichlet distribution, which defines a multinomial distribution. The median value was used with a Dirichlet distribution fitted of the UK experts' opinions (*Table 13*). Because the probabilities derived reflected a 6-month cycle, they had to be converted to monthly probabilities for the purposes of the model. Furthermore, although the derived probabilities included transition to death, it was decided to

TABLE 13 Six-monthly transition probabilities based on UK clinical experts' opinions

	To stage									
From stage	IA	IB	IIA	IIB	IIC	IIIA	IIIB	IIIC	IV	Death
IA	0.714	0.209	0.022	0.000	0.000	0.000	0.055	0.000	0.000	–
IB		0.732	0.093	0.062	0.041	0.052	0.000	0.021	0.000	–
IIA			0.617	0.191	0.074	0.074	0.011	0.021	0.011	–
IIB				0.477	0.205	0.159	0.102	0.057	0.000	–
IIC					0.416	0.225	0.157	0.124	0.079	–
IIIA						0.654	0.198	0.099	0.049	–
IIB							0.365	0.435	0.200	–
IIIC								0.549	0.451	–
IV										0.450
Death										

use the ORs of death from a different study by the same author (see *Mortality rates*).[182] Given this adjustment, the sum of the probabilities for each stage was < 1, and thus had to be rescaled to ensure that they summed to 1.

Recurrence probabilities

Ideally, melanoma incidence data from the NCRAS/PHE would be used as a surrogate to find the recurrence probability in England/UK. However, the NCRAS/PHE data are based on the summary stage information (i.e. AJCC stage I–IV) of all patients diagnosed by Clinical Commissioning Groups (CCGs) in England. The Durham cohort did not record the stage of recurrence adequately. Therefore, more granular AJCC stage I–III incidence information (stages IA, IB, IIA, IIB, etc.) is needed for the model. A large registry-based German study ($n = 33{,}384$),[209] which is likely to be similar to English patient cohorts in terms of characteristics of individuals, was used for this purpose. From the German study, the proportion of recurrences was 43.9% for AJCC stage IA, 25.3% for stage IB, 12.5% for stage IIA, 7.7% for stage IIB, 2.6% for stage IIC, 1.0% for stage IIIA and 3.0% for stage IV.

Recurrence probabilities over time for AJCC stages IA–IIC were obtained from a different study, conducted in Australia by Turner *et al.*[210] The study authors analysed 2298 patient records for the development of recurrence and new primary melanoma up to 10 years. In this paper,[210] Kaplan–Meier curves showing time to recurrence for localised melanoma were able to be digitised using WebPlotDigitizer (https://automeris.io/WebPlotDigitizer; accessed 23 April 2020). Various points along the curve were chosen and the co-ordinates of those points were extracted. These were then used to calculate the lamda and gamma parameters of the Weibull distribution for AJCC stage IA. The two parameters were then used to calculate the transition probabilities (*tp*) for recurrence using the following equation:

$$tp(t\mu) = 1 - \exp[\lambda(t-\mu)^{\gamma} - \lambda t^{\gamma}], \quad (1)$$

where:

- t is time (measured in terms of the number of cycles; each cycle is equivalent to 1 month)
- λ is the scale parameter, which describes the probability that melanoma recurs for an individual, given that he/she is recurrence free during the current time period
- γ is the shape parameter, which describes the hazard function of Weibull function for the survival time
- μ is the length of the Markov cycle.

Moreover, for the calculation of the baseline transition probability, the following formula was used, in which μ is the length of the Markov cycle:

$$tp(t\mu) = 1 - \frac{s(t)}{s(t-\mu)}. \quad (2)$$

Recurrence rates for the remaining melanoma stages were computed as a function of the probability of recurrence of stage IA and the distribution of the hazard ratio of each stage up to stage IIC reported in the study. By using the CIs presented in the paper, the corresponding standard errors (SEs) were calculated, which, along with the hazard ratios, were used as Dirichlet distributions in the model (see *Appendix 9*, *Tables 26–30*, for recurrence values). Because it was not possible to source recurrence ratios for AJCC stages III and IV, it was assumed that the recurrence rates for these two stages are similar to those of stage IIC. Given the very poor prognosis for these stages, and the non-curative nature of treatments, this was not expected to cause any meaningful limitation to the model. Depending on the stage of the recurrence in the model, the monthly probabilities were selected (e.g. stage IA: 0.0022 or 0.22%; and stage IB: 0.0046 or 0.46%).

Mortality rates

Age-specific, all-cause mortality rates were derived from general population mortality statistics reported in national life tables from the Office for National Statistics.[211] The melanoma-specific risks of death for each stage, as reported by Wilson *et al.*,[182] were used and these were expressed as ORs. These mortality ORs are based on the original US-based AJCC Staging Database ($n = 17{,}600$) (see *Appendix 9*, *Tables 24* and *25*).[90]

Resource use and unit costs

The costs of surveillance were broken down into the following cost categories: scheduled surveillance (e.g. consultation time), further invasive tests (e.g. local biopsy, SLNB) and treatment (e.g. targeted therapy and immunotherapy).

We derived data on the costs incurred for the different surveillance regimens and their consequences from routine data sources, such as the NHS reference costs.[212] *Table 14* shows the cost estimates used in the economic model. The costs of drug treatment, targeted therapy [trametinib in combination with dabrafenib, NICE Technology Appraisal (TA) 396[213]] for AJCC stage III and immunotherapy for AJCC stage IV disease (nivolumab in combination with ipilimumab, NICE TA400[214]) were obtained from the *British National Formulary* as per current clinical treatment regimen.[215] Gamma distributions were assigned to cost values for stages III and IV treatments (see *Appendix 9*, *Table 38*).

Health utilities

As melanoma and its treatment affects not only survival, but also quality of life, a focused search of the literature and other relevant sources [e.g. the Sheffield School of Health and Related Research Health Utilities Database (ScHARRHUD) and the Health Economics Research Centre (HERC), Oxford, database of mapping studies] was initiated (see *Appendix 9*, *Table 37*). However, in 2018, a systematic review and meta-analysis of melanoma utility weights based on 33 studies was published.[216] This review included studies from Australia, Europe and North America, and sought to define post-treatment utilities by stage with time of data collection component up to 3 months, 3–12 months and > 12 months. The synthesis of health-state utilities is controversial. Although it aims to generate a more accurate estimate of the mean health-state utility and the associated uncertainty, a formal synthesis may not be meaningful considering the variability in measures [e.g. EuroQol-5 Dimensions (EQ-5D) vs. Short Form questionnaire-6 Dimensions], valuation method, types of anchors used, the country where the valuation was done and who provided the preference weights.[217]

TABLE 14 Resource use and costs

Description	Mean cost (£)	Source
Clinical specialty		
CNS	89	*Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*[212]
Surgeon	97	*Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*[212]
Dermatologist	108	*Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*[212]
Diagnostic investigations		
Local biopsy	387[a]	*Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*[212]
Wide local excision	218	*Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*[212]
CT	106	*Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*[212]
Surgical costs		
Radical lymph node dissection	1599	*Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*[212]
SLNB	1599	*Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*[212]
Treatment		
Stage III, per month (targeted therapy): • Dabrafenib, 150 mg twice daily • Trametinib, 2 mg once daily	10,400	Based on local clinical expert opinion. This is the average monthly cost of dabrafenib (£5600) + trametinib (£4800)
Stage IV, per month (immunotherapy): • Ipilimumab, 3 mg/kg every 3 weeks for four doses • Nivolumab, 240 mg every 2 weeks. Alternatively, 480 mg every 4 weeks for four doses. Then patients have nivolumab as a single agent until treatment failure/ significant toxicity	10,326	Based on local clinical expert opinion. This is the average monthly cost based on 12 months of combination treatment with ipilimumab (£15,000/vial of 200 mg/40 ml) + nivolumab (£2663/vial of 240 mg/24 ml)

a £387.64 used in the model.

As no relevant utility values derived from a UK population were available, the published values from the meta-analysis were used in the economic model (*Table 15*). It is further assumed that individuals in the 'disease-free' state have a health-state utility value of full health (utility score = 1), as per Wilson *et al.*[182] These are values without age adjustment. These values were chosen for the model as they are published in the literature and are consistent across all strategies. Beta distributions were assigned to utility values in the base-case analysis.

Main modelling assumptions

This section provides a brief summary of the key assumptions made when developing the economic model. It was assumed that the time interval between initial surgery and any subsequent event/ treatment is 1 month. No utility decrements associated with undertaking any of the tests (local biopsy, SLNB, etc.) or treatments (immunotherapies have severe adverse effects) were incorporated in the model. It was assumed that the delay in recurrence/new primary diagnosis would eventually be

TABLE 15 Utility scores based on time on treatment

Time on treatment (months)	Utility estimates (95% CI)
Disease free	1.000
AJCC stage I/II	
Day 1 to 3 months	0.772 (0.753 to 0.790)
3–12 months	0.852 (0.844 to 0.860)
> 12 months	0.857 (0.850 to 0.865)
AJCC stage III/IV	
Day 1 to 3 months	0.803 (0.783 to 0.823)
3–12 months	0.797 (0.786 to 0.809)
> 12 months	0.848 (0.787 to 0.910)
AJCC stage IV	
Day 1 to 3 months	0.653 (0.621 to 0.685)
3–12 months	0.831 (0.808 to 0.855)
> 12 months	0.833 (0.820 to 0.847)

captured by self-detection as advanced disease symptoms become evident. It was also assumed that individuals attend all of their scheduled follow-up appointments, that the initial surgical excision is curative and that individuals enter the model as disease free. Recurrence rates for stages IIIA, IIIB, IIIC and IV were assumed to be similar to the recurrence rates for stage IIC. Estimates of melanoma-specific risk of death for each stage were taken from Wilson *et al.*[182] It was assumed that patients received systemic treatment until treatment failure or death (Dr Janine Graham, personal communication). In the model, it was assumed that patients were *BRAF* positive and received dabrafenib and trametinib targeted therapy. The literature suggests that 50% of metastatic melanomas are *BRAF* positive.[218] If patients were *BRAF* negative, they would probably receive immunotherapy (e.g. ipilimumab).[219] The assumption that stages III and IV are now treated with new (expensive) therapies is captured in the model and the assumed costs are approximations. Finally, the 'no-surveillance' option was not considered in the model, as it was deemed to be unacceptable to patients by the advisory board.

Incremental cost-effectiveness analysis

Base-case analysis

The joint estimates of costs and effects were combined in an incremental analysis and presented in terms of the mean incremental cost-effectiveness ratio (ICER) for each comparator. The ICERs were calculated as the difference in costs divided by the difference in effects (QALYs) between treatment options. The ICERs were calculated for each successive alternative, from the least costly to the most costly. To help identify the optimal approach, the net monetary benefit (NMB) framework was used, whereby the NMB for a given strategy is equal to the accrued QALYs multiplied by the ceiling ratio (λ) of willingness to pay (WTP) per QALY, minus the strategy costs:

$$\text{NMB} = (\text{QALYs} \times \lambda) - \text{costs}. \quad (3)$$

A value of £20,000, which is typically used by NICE as a threshold to inform judgements on cost-effectiveness, was placed on the ceiling ratio.[220] The threshold means that NICE is prepared to pay £20,000 for each extra QALY gained.

Measures of variance for the joint incremental costs and effects were obtained using Monte Carlo simulation in the probabilistic sensitivity analysis (PSA) (see *Probabilistic sensitivity analysis*) and presented graphically using cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs).[221]

Sensitivity analyses

Both deterministic and probabilistic sensitivity analyses were used to explore parameter uncertainty surrounding estimates of cost-effectiveness. Threshold analyses were used to explore further considerations around the main parameters of interest.

Deterministic sensitivity analyses

Deterministic sensitivity analysis is a method that can be used to investigate the sensitivity of the results to variations in a specific input parameter or set of parameters. The parameters of most interest were manually changed (across a plausible range) and the results were analysed to determine to what extent the change affects the output values.

Given that microsimulation models are computationally burdensome, and owing to the large number of strategies, deterministic analysis also helped exclude strategies that were deemed never to be cost-effective. Therefore, many of the model parameters were subject to one-way sensitivity analyses, using hypothetical increases or decreases, to determine the key drivers of the model results. Deterministic sensitivity analyses were also carried out to test for the effect of assumptions and variability.

Probabilistic sensitivity analysis

When available, data were also entered into the model as distributions to more fully incorporate the uncertainty around parameter values so that a PSA could be undertaken. In decision modelling, many of the parameter values are often estimated with a degree of uncertainty. The probabilistic distribution for each parameter was defined by considering the mean, SE and anticipated shape of the distribution. The PSA was run with 1000 simulations for each individual and CEACs were produced to identify the probability of the different strategies being cost-effective across a range of WTP thresholds. Estimation of costs and QALYs were calculated as the expectation over the joint distribution of the parameters.

One-way sensitivity and scenario analyses

One-way sensitivity analyses can indicate the effect of parameter values on the expected net benefit, but not the overall decision on what strategy the NHS should approve. A key feature in the model is the role of self-diagnosis (self-detection). In their paper, Turner *et al.*[210] reported that self-detection rates for recurrence ranged from as low as 15%[222] up to 90%.[223] For new primaries, the literature suggests plausible self-detection rates as low as 5%[222] and as high as 50%.[224] The aforementioned studies do not explicitly mention a time period over which this self-detection occurs. Therefore, additional analyses are warranted to determine if optimal strategies, in terms of NMB, change in response to changes in self-detection rates.

The nature of recurrence of early-stage disease in the base-case analysis was taken from an Australian study.[210] It is likely that the natural history of melanoma is similar around the world, irrespective of geography. However, there is uncertainty in the estimate and additional analyses are warranted to determine if there is a threshold when, in terms of NMB, optimal strategies change in response to changes in recurrence rates.

The use of a meta-analysis for utility values estimate is controversial.[217] As the model is set up in a way that recurrences are detected ranging from stage IA to IV, the utility values are likely to have a plausible range that is greater than the range in the base case. This sensitivity analysis assessed whether or not the relative efficiency of strategies changed when alternative utility values were used.

From an NHS health-care management perspective, the role of specialist dermatological nurses in melanoma surveillance of stage I patients is worth exploring further. This is to ascertain what level of diagnostic accuracy they need to have to be considered a viable strategy, compared with the other strategies considered.

Even though a validated prognostic risk model was not identified in the systematic review, reported in *Chapter 4*, the hypothetical scenario of a prognostic approach (be that as a risk model, a biomarker test, or a combination of the two) is explored here. As evidence emerges in this field about such technologies, it is plausible that this may lead to further plausible surveillance strategies that could be potentially cost-effective.

Value-of-information analysis

A critical role of the PSA was that it facilitated the estimation of the value of further research. Therefore, in addition to exploring the cost-effectiveness of different follow-up strategies for individuals with stages IA and IB melanoma, the main uncertainties were quantified to inform any future direction of research. This was explored through the expected value of partial perfect information (EVPPI) analysis. EVPPI analysis can be used to estimate the expected value of removing the uncertainty of specific parameters or groups of parameters in a given model. It can be used to help guide decisions on where future research should be directed to obtain more precise and reliable estimates of specific parameters.[225] EVPPI places an upper value on conducting further research on a specific area of information. It indicates the maximum gain that could be obtained by removing all uncertainty in that area. In reality, any piece of future research would remove only part of that uncertainty, but the approach can still help indicate where the most gain could be made from further research.

To calculate the population EVPPI, the individual EVPPI is first calculated by the model. The individual EVPPI is then multiplied by the size of the population that will be affected by the information over the anticipated lifetime of the technology. For the estimation of the size of the population, the annual prevalence (the number of patients affected by the decision each year) of stage I disease is needed. The assumption is that the annual incidence of stage I melanoma is, in essence, the number of patients affected by the decision each year. In 2017, 8555 patients were diagnosed with stage I disease.[226] The proportion (63% for IA vs. 37% for IB) was provided by Leiter *et al.*[209] in their work based in Germany.

Two-level simulations were conducted to estimate the EVPPI. The first level happened in the microsimulation by randomly selecting individuals of different age and sex from the individual-level data. Then, each selected individual was simulated 10,000 times (PSA) and values for the parameters were selected from the prespecified distributions. Results from the PSA were then used in the Sheffield Accelerated Value of Information (SAVI) tool to estimate the EVPPI.[227]

Four groups of parameters were considered in the EVPPI analysis:

1. health utility values
2. diagnostic accuracy of health-care professional to detect a melanoma
3. probability of transitioning between stages
4. recurrence of melanoma.

Model validation

Model validation was achieved by clinical expert concurrence, by a review by modelling colleagues at Newcastle University and by being transparent in reporting. As laid out in a review on validating cost-effectiveness models, 'internal validity' related to input parameters, whereas 'descriptive validity' of a clinical pathway requires analysts to make trade-offs between accuracy, complexity and fulfilling the purpose of the model.[228] For this project, the objective was to compare surveillance strategies for stage I disease, rather than capturing the complex and evolving management of advanced-stage disease.

Results

Base-case analysis results

Given the computational burden of running the model with 10,000 PSA simulations over 87 strategies, the model was first run deterministically to identify 20 strategies that were considered to have the most promise of being cost-effective for the stage IA model. This was repeated for the stage IB model. These strategies were then used in the base-case analysis.

Base-case analysis: stage IA

Monte Carlo simulation was performed to obtain probabilistic estimates of the cost-effectiveness of the different surveillance strategies, compared with NICE's recommended follow-up schedule. The 20 strategies included in the analysis are presented in *Table 16*.

Table 17 and *Figure 9* (which excludes strategies that are, on average, dominated) report the results of the base-case analyses for the average patient treated for stage IA melanoma. Results are reported in

TABLE 16 Strategies for surveillance of stage IA melanoma included in the base-case analysis

Strategy	Duration of follow-up	Intervals of follow-up each year	Health-care professional undertaking screening
1–NICE	1 year	Every 3 months for 1 year	Dermatologist
4	10 years	Every 3 months for 1 year and every 6 months thereafter	Dermatologist
7	10 years	Every 3 months the first year and every 12 months thereafter	Dermatologist
14	10 years	Every 4 months the first year and every 12 months thereafter	Dermatologist
15	1 year	Every 6 months for 1 year	Dermatologist
16	3 years	Every 6 months for 3 years	Dermatologist
19	3 years	Every 6 months for 1 year and every 12 months thereafter	Dermatologist
21	10 years	Every 6 months for 1 year and every 12 months thereafter	Dermatologist
22	1 year	Once for 1 year	Dermatologist
23	3 years	Every 12 months for 3 years	Dermatologist
29	10 years	Every 3 months for 1 year and every 6 months thereafter	Surgeon
32	10 years	Every 3 months for 1 year and every 12 months thereafter	Surgeon
37	3 years	Every 4 months for 1 year and every 12 months thereafter	Surgeon
39	10 years	Every 4 months for 1 year and every 12 months thereafter	Surgeon
40	1 year	Every 6 months for 1 year	Surgeon
41	3 years	Every 6 months for 3 years	Surgeon
44	3 years	Every 6 months for 1 year and every 12 months thereafter	Surgeon
46	10 years	Every 6 months for 1 year and every 12 months thereafter	Surgeon
47	1 year	Once for 1 year	Surgeon
48	3 years	Every 12 months for 3 years	Surgeon

TABLE 17 Results of the base-case analysis for patients treated for stage IA melanoma

Strategy[a]	Cost (£)	Incremental cost (£)	QALY	Incremental QALY	ICER (£) (Δcost/ΔQALY)	NMB (£)	Probability of being cost-effective for different threshold values for society's WTP for a QALY (%)		
							£20,000	£30,000	£50,000
22	**8476**		**14.74**			**286,321**	**13.2**	**9.7**	**6.2**
15	8824	348	14.75	0.01	33,367 (extendedly dominated[b])	286,181	7.1	5.6	4.1
23	8874	50	14.73	−0.02	Absolutely dominated[c]	285,820	10.8	9.5	7.6
19	**9215**	**391**	**14.77**	**0.02**	**25,894[d]**	**286,153**	**9.8**	**8.4**	**6.7**
1-NICE	9336	121	14.76	−0.00	Absolutely dominated[c]	285,952	12.4	10.3	8.2
16	9613	399	14.77	0.00	199,677(extendedly dominated[b])	285,794	7.2	6.8	6.0
21	10,254	641	14.76	−0.01	Absolutely dominated[c]	284,952	8.3	8.2	7.3
14	10,494	881	14.76	−0.01	Absolutely dominated[c]	284,747	11.3	11.4	10.4
7	**10,768**	**1154**	**14.78**	**0.01**	**103,890[e]**	**284,899**	**7.8**	**7.9**	**7.1**
4	12,222	1455	14.78	0.00	1,462,441 (extendedly dominated[b)]	283,464	6.3	7.6	8.1
47	13,909	1687	14.74	−0.04	Absolutely dominated[c]	280,893	1.7	2.7	3.6
40	14,427	2205	14.75	−0.04	Absolutely dominated[c]	280,522	0.1	0.8	2.2
48	14,579	2356	14.74	−0.05	Absolutely dominated[c]	280,173	0.7	1.4	2.2
44	15,072	2850	14.76	−0.03	Absolutely dominated[c]	280,089	0.6	1.5	3.0
37	15,518	3296	14.76	−0.02	Absolutely dominated[c]	279,749	0.7	1.5	2.6
41	15,754	3532	14.76	−0.02	Absolutely dominated[c]	279,501	0.3	0.8	1.7
46	16,923	4701	14.77	−0.02	Absolutely dominated[c]	278,449	0.5	1.2	3.2
39	17,367	5145	14.78	−0.00	Absolutely dominated[c]	278,227	0.6	2.1	4.0
32	17,765	5543	14.78	0.00	10,081,791 (extendedly dominated[b])	277,932	0.3	1.9	3.6
29	**20,319**	**2554**	**14.79**	**0.01**	**1,335,810[f]**	**275,490**	0.3	0.7	2.2

a See *Table 16* for strategy description.
b This strategy has a higher ICER (relative to a third option) and fewer benefits than the alternative. Strategy 22 (common baseline as it is least costly option) is the comparison or reference strategy.
c This strategy is always more costly and less beneficial than comparator strategies.
d Compared with strategy 22.
e Compared with strategy 19.
f Compared with strategy 7.

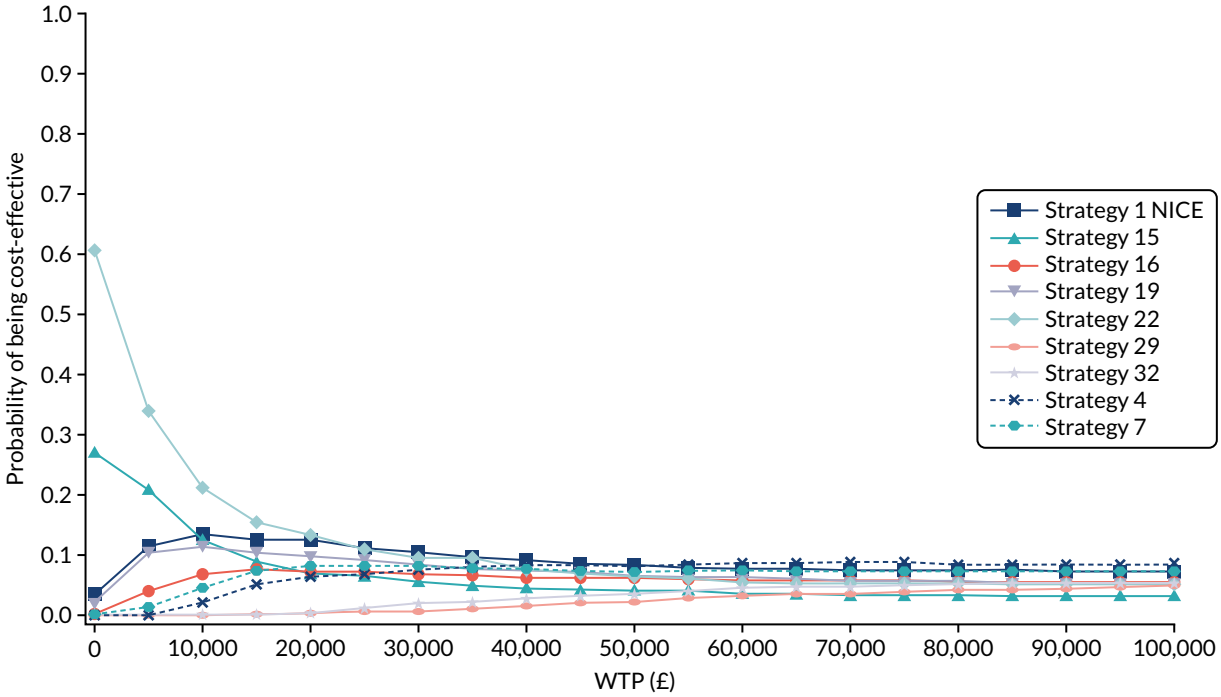


FIGURE 9 The CEAC for stage IA melanoma (dominated strategies excluded). See *Table 16* for strategy description.

terms of total and incremental costs and effectiveness, incremental cost per QALY (ICER) and NMB (were society's WTP for a QALY (λ) = £20,000 per QALY), alongside the probability that each strategy would have the highest NMB for different threshold values used by NICE.

For stage IA, over a lifetime time horizon, none of the strategies involving a specialist dermatological nurse is likely to be considered cost-effective. It is worth noting that, overall, all strategies for stage IA produced very similar QALYs. This might be expected, given the relatively low chance of recurrence or metastasis and the high rate of self-diagnosis (it has been assumed that 60% of all melanomas are self-diagnosed within 12 months). Therefore, the lifetime costs are the main driver for evaluating surveillance strategies.

Strategy 22, which is follow-up once at 12 months by a dermatologist, is, on average, the least costly surveillance strategy (£8476 per person) and generates 14.74 QALYs. This strategy also has the highest NMB (£286,321) at a WTP threshold of £20,000 per QALY. Should the NHS wish to make a decision on cost alone, then this strategy has approximately a 60% chance of being considered the most cost-effective. As society's willingness for a QALY increases, other, more effective, strategies become increasingly cost-effective. At a £20,000 threshold for society's WTP for a QALY, strategy 22 has almost a 13% probability of being considered cost-effective. At the same threshold, three other strategies have a probability of > 10% of being cost-effective. These are strategy 23 (follow-up every 12 months for 3 years), strategy 1 (the NICE-recommended strategy: every 3 months for 1 year) and strategy 14 (surveillance every 4 months for the first year and then every 12 months for the next 9 years). Two strategies have a probability of 8–10% of being cost-effective: strategy 19 (every 6 months for 1 year and every 12 months for the next 2 years) and strategy 21 (every 6 months for 1 year and every 12 months for the next 9 years).

Given that 20 strategies are being compared, should each strategy be equally likely to be considered cost-effective, it would be expected that each strategy would have only a 5% chance of being considered cost-effective. Several of the dermatology-based strategies (including the NICE-recommended strategy) have probabilities greater than this, and four strategies have probabilities of > 10%. These four include strategies that are both less intensive and more intensive than the strategy recommended by NICE.

Base-case analyses: stage IB

The 20 strategies included in the analysis for stage IB disease are presented in *Table 18*. As with strategies for stage IA, strategies involving the specialist dermatological nurse were judged very unlikely to be considered cost-effective and were not included in the base-case analysis.

Table 19 and *Figure 10* report the results of the base-case analyses for stage IB melanoma patients. None of the strategies involving a surgeon conducting the surveillance had a probability of > 1% of being cost-effective across any of the values for society's WTP for a QALY. The strategy recommended by NICE (strategy 1, surveillance by a dermatologist every 3 months for the first 3 years and then every 6 months thereafter for the next 2 years) also never had a probability of > 6% of being cost-effective over any of the values for society's WTP for a QALY. As with the analysis for stage IA disease, given that 20 strategies are being compared, should each strategy be equally likely to be considered cost-effective, it would be expected that each strategy would have only a 5% chance of being considered cost-effective.

TABLE 18 Strategies for surveillance of stage IB melanoma included in the base-case analysis

Strategy	Duration of follow-up	Intervals of follow-up each year	Health-care professional undertaking screening
1–NICE	5 years	Every 3 months for the first 3 years and every 6 months thereafter	Dermatologist
2	3 years	Every 3 months for 3 years	Dermatologist
4	20 years	Every 3 months for the first 3 years and every 6 months thereafter	Dermatologist
5	5 years	Every 3 months for the first 3 years and every 12 months thereafter	Dermatologist
8	3 years	Every 4 months for 3 years	Dermatologist
9	5 years	Every 4 months for the first 3 years and every 6 months thereafter	Dermatologist
11	20 years	Every 4 months for the first 3 years and every 6 months thereafter	Dermatologist
15	3 years	Every 6 months for 3 years	Dermatologist
18	20 years	Every 6 months for 20 years	Dermatologist
23	5 years	Every 12 months for 5 years	Dermatologist
25	20 years	Every 12 months for 20 years	Dermatologist
29	20 years	Every 3 months for the first 3 years and every 6 months thereafter	Surgeon
77	1 year	Once for 1 year	Dermatologist
78	1 year	Once for 1 year	Surgeon
80	2 years	Every 12 months for 2 years	Dermatologist
81	2 years	Every 12 months for 2 years	Surgeon
82	2 years	Every 6 months for 2 years	Surgeon
83	2 years	Every 6 months for 2 years	Dermatologist
86	1 year	Every 6 months for 1 year	Dermatologist
87	1 year	Every 6 months for 1 year	Surgeon

TABLE 19 Results of the base-case analysis for melanoma treated for stage IB melanoma

Strategy[a]	Cost (£)	Incremental cost (£)	QALY	Incremental QALY	ICER (£) (Δcost/ΔQALY)	NMB (£)	Probability of being cost-effective for different threshold values for society's WTP for a QALY (%)		
							£20,000	£30,000	£50,000
77	**9536**		**14.57**	**0.00**		**281,945**	**12.9**	**9.7**	**6.8**
80	**9833**	**297**	**14.61**	**0.04**	**7818**	**282,408**	**9.8**	**7.9**	**6.2**
86	9971	141	14.62	0.00	47,383 (extendedly dominated[b])	282,326	10.4	8.4	6.0
23	10,320	346	14.60	−0.01	Absolutely dominated[c]	281,727	8.6	7.7	6.1
83	**10,475**	**501**	**14.63**	**0.02**	**31,172**	**282,146**	**6.2**	**5.2**	**4.0**
15	10,696	221	14.62	−0.01	Absolutely dominated[c]	281,642	4.5	3.6	3.2
8	11,420	945	14.61	−0.02	Absolutely dominated[c]	280,811	6.5	5.7	5.0
25	11,910	1435	14.63	−0.01	Absolutely dominated[c]	280,600	7.9	7.7	7.1
9	12,158	1683	14.62	−0.01	Absolutely dominated[c]	280,332	6.5	6.8	7.1
2	12,187	1712	14.64	0.01	175,956 (extendedly dominated[b])	280,629	3.6	3.6	4.0
5	12,485	297	14.64	−0.01	Absolutely dominated[c]	280,317	3.6	4.1	4.4
1–NICE	12,748	561	14.64	0.00	3,079,045 (extendedly dominated[b])	280,072	4.1	4.9	5.5
18	14,607	1858	14.65	0.01	139,038 (extendedly dominated[b])	278,480	5.5	7.1	8.1
78	14,970	363	14.59	−0.06	Absolutely dominated[c]	276,838	0.2	1.6	2.8
11	15,352	745	14.65	−0.00	Absolutely dominated[c]	277,657	4.7	6.1	7.5
81	15,391	784	14.62	−0.04	Absolutely dominated[c]	276,926	0.7	1.2	2.0
87	15,552	946	14.62	−0.04	Absolutely dominated[c]	276,748	0.8	1.5	2.3
4	**16,045**	**1438**	**14.69**	**0.03**	**45,598**	**277,673**	2.8	4.6	6.2
82	16,353	309	14.64	−0.05	Absolutely dominated[c]	276,400	0.7	1.8	2.6
29	25,871	9826	14.69	−0.00	Absolutely dominated[c]	267,840	0.0	0.8	3.1

a See *Table 18* for strategy description.
b This strategy has a higher ICER (relative to a third option) and fewer benefits than the alternative. Strategy 77 (common baseline as it is least costly option) is the comparison or reference strategy.
c This strategy is always more costly and less beneficial than comparator strategies.

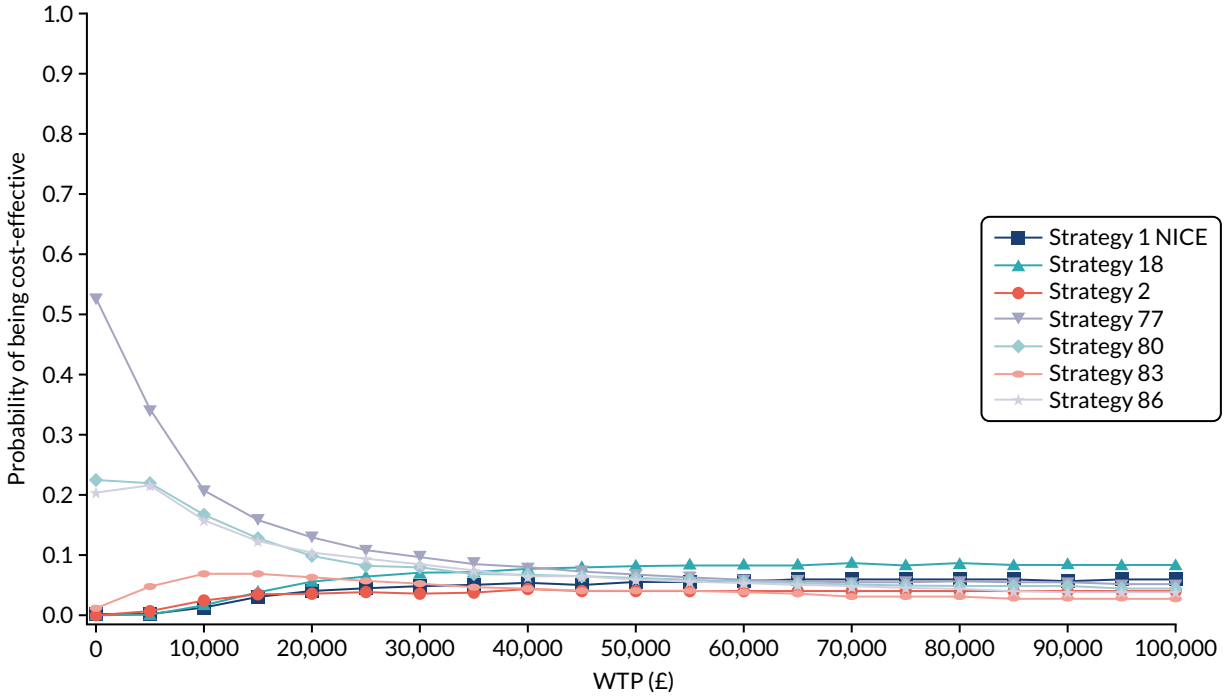


FIGURE 10 The CEAC for stage IB strategies (dominated strategies excluded). See *Table 18* for strategy description.

For stage IB, over a lifetime time horizon, strategy 77 (follow-up once for 1 year by a dermatologist) is, on average, the least costly strategy (£9536) and generates 14.57 QALYs. Should society not be willing to pay for additional QALYs, then this strategy has a 55% chance of being cost-effective. As society's WTP for a QALY increases, the probability that strategy 77 would be considered cost-effective falls, but strategy 77 has the highest probability of being cost-effective (12.9%) when society is willing to pay £20,000 per QALY. At the same threshold value, only one other strategy considered had a probability of being cost-effective of > 10%: strategy 86 (surveillance by a dermatologist every 6 months for 1 year). It is worth noting that, at a WTP threshold of £20,000 per QALY, strategy 80 (surveillance by a dermatologist every 12 months for 2 years) had the highest NMB and an ICER of £7818 per QALY gained, compared with strategy 77. Three further strategies had a probability of being cost-effective of between 7% and 10%: strategy 80, strategy 23 (surveillance by a dermatologist every 12 months for 5 years) and strategy 25 (surveillance by a dermatologist every 12 months for 20 years). With the exception of strategy 25, all of the aforementioned strategies are less intensive than the NICE-recommended strategy. Strategy 25 has a longer follow-up duration (20 years), with patients followed up annually.

Sensitivity analyses results

The results of the PSA base case best capture the uncertainty in the decision problem. The one-way sensitivity can help indicate the effect of parameter values on the expected NMB of each comparator strategy (see *Appendix 10*).

Keeping all other parameters at their base-case values, the impact of the probability of self-diagnosis on the NMB was examined. The model uses a yearly probability of 60%, which was transformed to a monthly probability (0.074 or 7%). A number of sensitivity analyses were run for lower and higher values of self-diagnosis. Strategies with a probability of > 1% of being cost-effective at a WTP of £20,000 were chosen for stages IA and IB. Results are presented in *Figures 11* and *12* in terms of NMB, and in *Appendix 10*, *Tables 39* and *40*, in terms of cost, effectiveness, ICER and NMB of each strategy. *Figure 11* and *Appendix 10*, *Table 39*, report the results for stage IA; *Figure 12* and *Appendix 10*, *Table 40*, report the results for stage IB. The results indicate that, without self-diagnosis, more

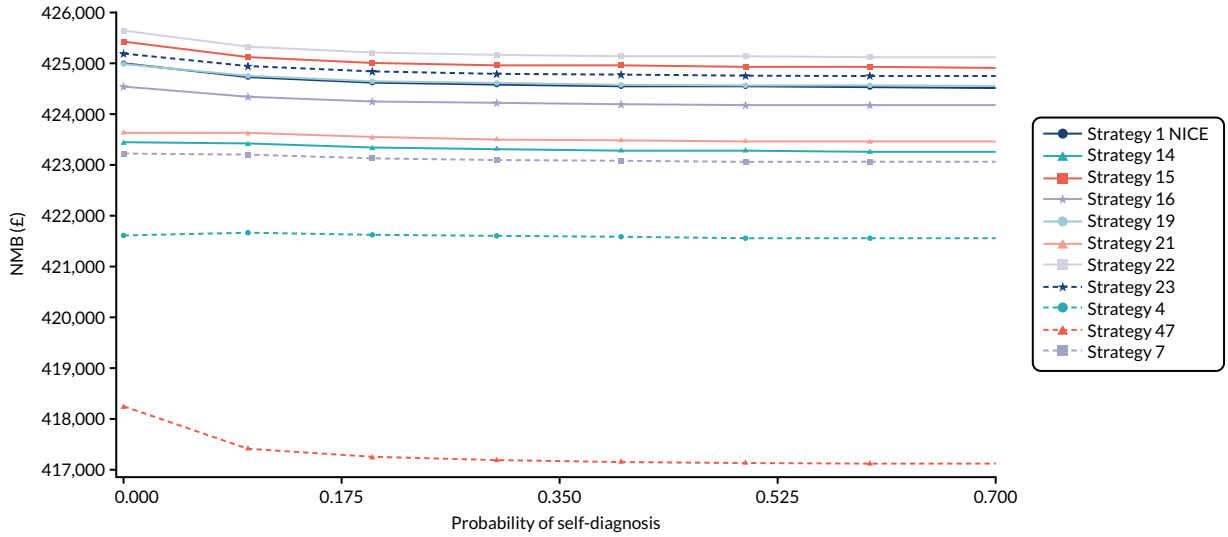


FIGURE 11 One-way sensitivity analysis of monthly probability of self-diagnosis for stage IA patients.

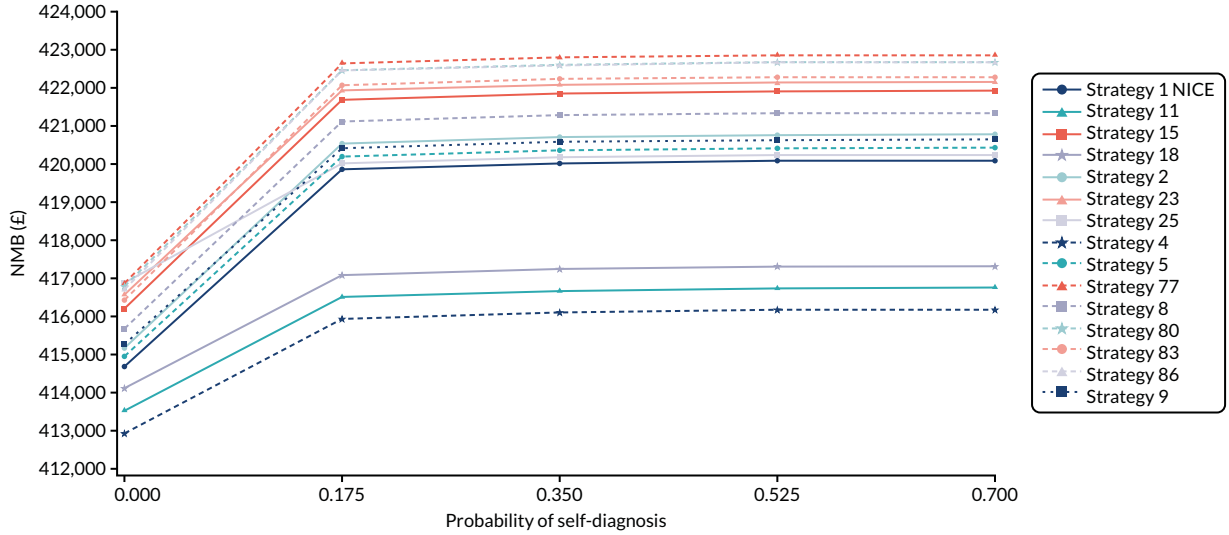


FIGURE 12 One-way sensitivity analysis of monthly probability of self-diagnosis for stage IB patients.

intensive and longer surveillance strategies for stage IA (strategies 21, 14, 7 and 4) and stage IB (strategies 25, 18, 11 and 4) produce a fraction more QALYs (< 0.2 over the estimated patient's average life expectancy) and are more expensive (stage IA: £2000–5000; stage IB: £4000–10,000). At higher monthly probabilities of self-diagnosis (≥ 0.35, or 35%), the benefits of surveillance are questionable as there is such little gain in QALYs. Thus, more costs are incurred from the continuous surveillance without any substantial gain in QALYs.

Similar patterns are observed when the probabilities of recurrence are varied and all other parameters are kept at their base-case values. Results are presented in *Figures 13* and *14* and in *Tables 19* and *20* for stages IA and IB, respectively (further results are reported in *Appendix 10*, *Tables 41* and *42*). At no or very low monthly probability of recurrence, more intensive and longer strategies for stage IA (strategies 21, 14, 7 and 4) and stage IB (strategies 25, 1-NICE, 18, 11 and 4) are more costly without producing any additional QALYs. As monthly recurrence probabilities increase, QALYs tend to drop (as would be expected because there are reductions in quality of life and survival). When monthly recurrence probability reaches ≥ 0.29, effectiveness between high and low resource use surveillance

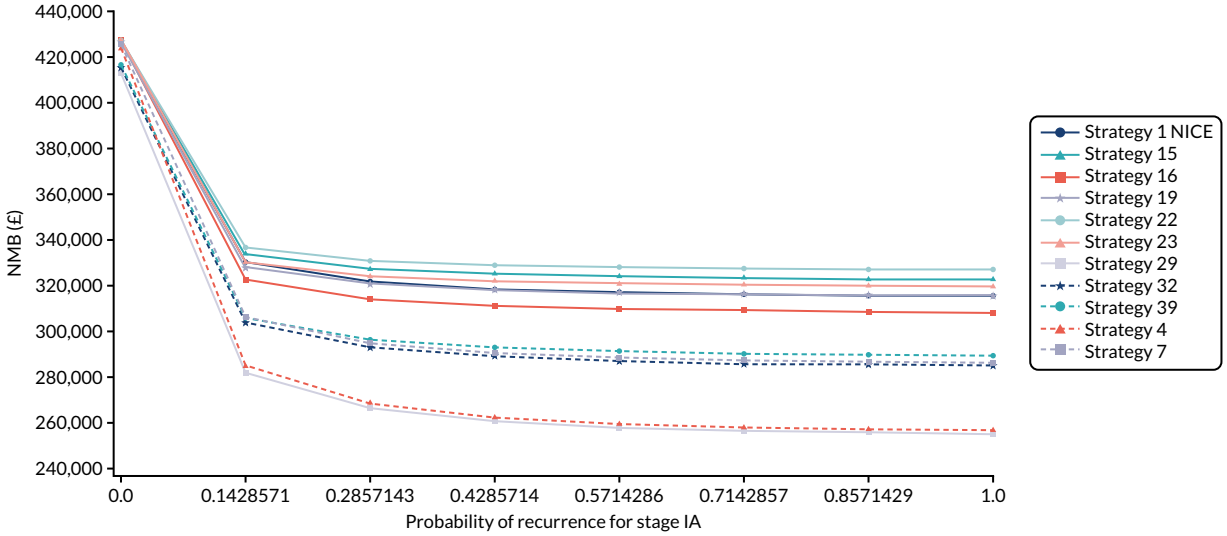


FIGURE 13 One-way sensitivity analysis of monthly probability of recurrence for stage IA patients.

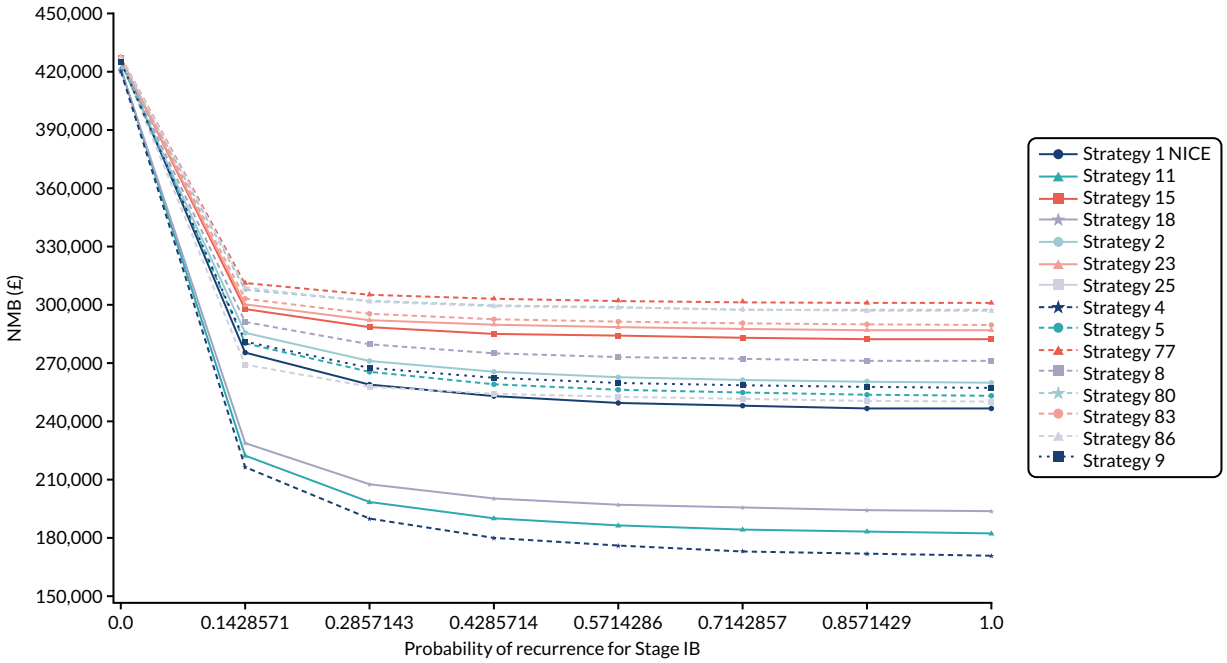


FIGURE 14 One-way sensitivity analysis of monthly probability of recurrence for stage IB patients.

strategies, in terms of QALYs gained, is low (e.g. for stage IB, for strategy 11, there are 19.49 QALYs, and for strategy 77, there are 19.45 QALYs). Although the benefits of surveillance are faster detection of the recurrence, the actual gain in QALYs will be minimal.

In the case of utility, *Figures 15* and *16* show that, as long as the utility value of the recurrence is > 0.1 for stage IA and IB patients, then the NMBs of all the surveillance strategies are positive. The most important utility values are related to stage I, as approximately 70% of all disease is diagnosed when the cancer is stage I.[226] Without data to the contrary, it is likely that most second primary melanoma diagnosis or first primary recurrences are also diagnosed at stage I.

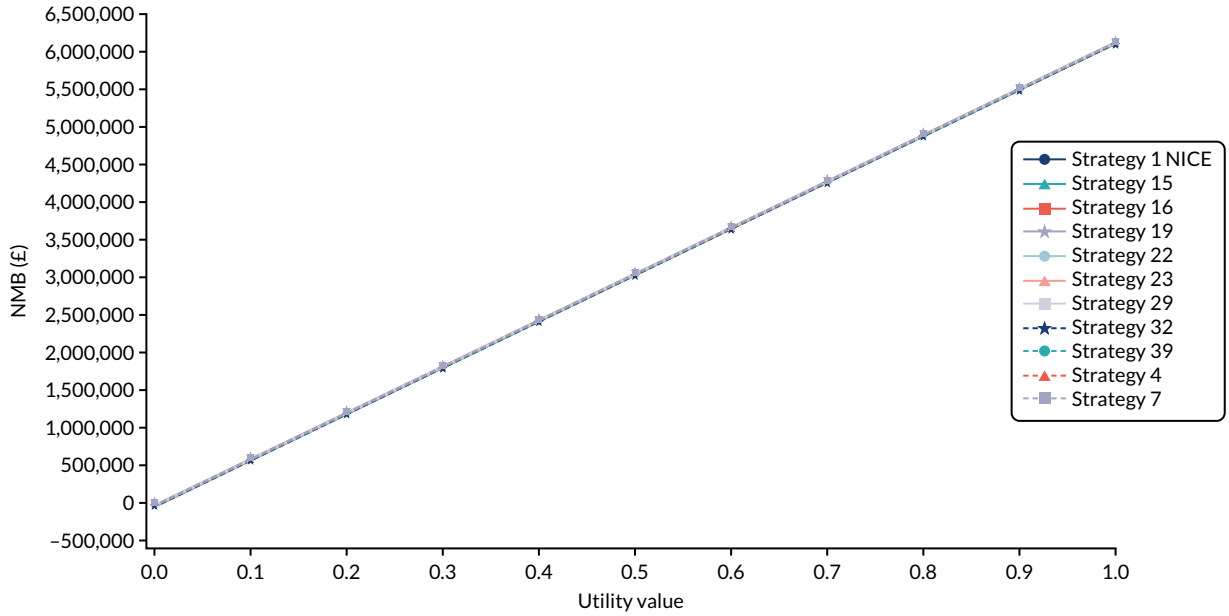


FIGURE 15 One-way sensitivity analysis for utility values for stage IA patients.

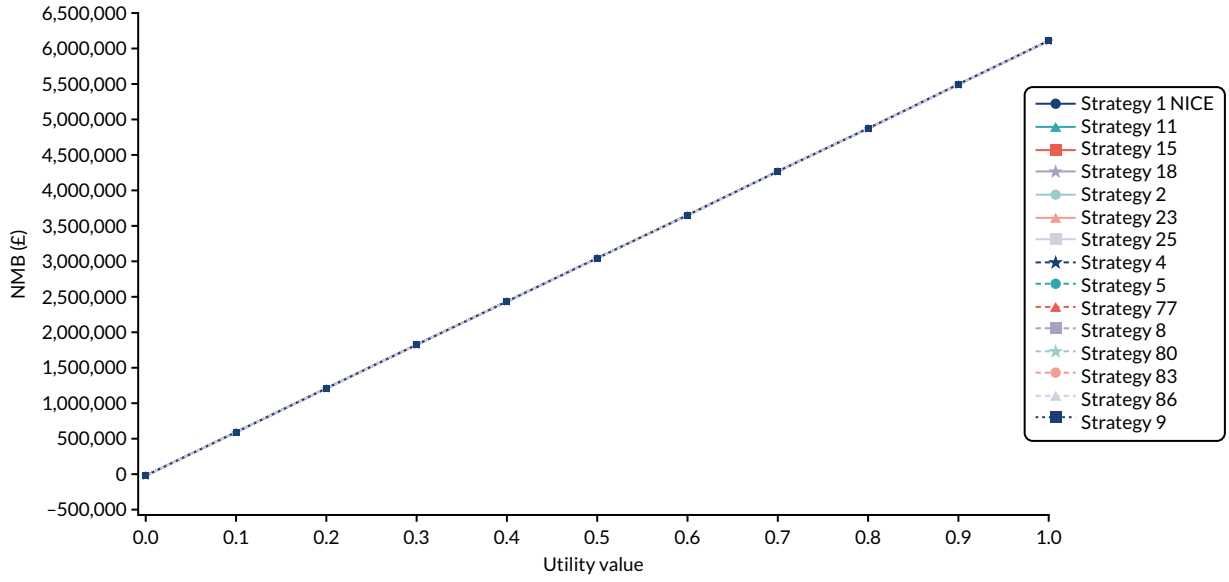


FIGURE 16 One-way sensitivity analysis for utility values for stage IB patients.

TABLE 20 The EVPPIs for pertinent parameters (stage IA)

Parameter	Per person EVPPI (£)	EVPPI for England per year (£)	EVPPI for England over 20 years (£)
Probabilities of transitioning between stages	3422	19,024,352	380,487,047
Diagnostic accuracy	2483	13,804,052	276,081,046
Health utility values	864	4,803,343	96,066,864
Recurrence of melanoma	224	1,245,311	24,906,224

Role of specialist dermatological nurse in surveillance

One of the research questions at the start of this analysis was whether or not specialist dermatological nurses/CNSs have a role to play in surveillance. From the initial analyses presented in the base-case analyses, none of the nurse strategies was cost-effective; therefore, these strategies were not included in the PSA. To explore the reason why they were not cost-effective, additional analyses, based on the diagnostic accuracy of the strategies, were performed (i.e. the ability of dermatological nurses to correctly diagnose whether or not a recurrence has occurred). Given that the specificity and sensitivity of nurses were obtained from a single study, which assessed the performance of eight 'physician assistants' in rating 173 dermoscopic images of skin lesions with known histological diagnosis,[207] sensitivity analysis explored the impact on the cost-effectiveness of changes in the sensitivity and specificity of diagnosing recurrences in the nurse-based strategies. One-way sensitivity analysis showed that, at low values of specificity, nurse strategies are not expected to be cost-effective. This is because individuals are referred for further tests instead of being discharged. However, as specificity increases, nurses become more cost-effective, compared with alternative strategies. When the value of specificity reaches 0.876 for stage IA and 0.878 for stage IB, nurses become as cost-effective as dermatologists (*Figures 17* and *18*).

Potential role of prognostic test

A hypothetical scenario in which a validated prognostic test, be it a risk factor-based model or a biomarker or a combination of both, is available to the NHS (cost: £250) was added to the model to compare surveillance strategies. It was assumed that the test has a sensitivity of 0.8 and a specificity of 0.8 in identifying high- and low-risk patients. It was also assumed that the 'true' prevalence of recurrence over a lifetime is 20%, based on an Australian study.[210] Those identified as being at high risk received the recommended NICE strategy and those identified as being at low risk received the strategy that was identified in the base-case analysis as being cost-effective (stage IA: strategy 22; stage IB: strategy 77). This hypothetical prognostic test strategy was compared with a few strategies, including the current NICE strategy for stage IA and IB patients.

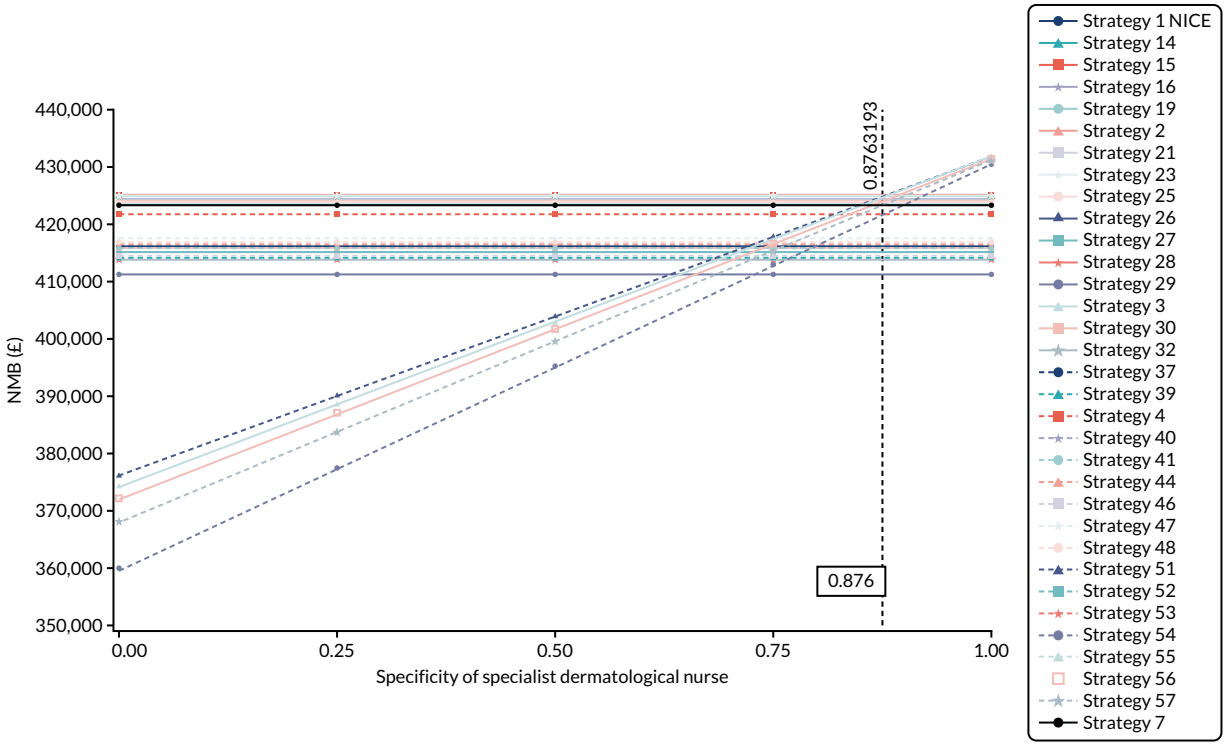


FIGURE 17 One-way sensitivity analysis of specificity of specialist dermatological nurses for stage IA patients.

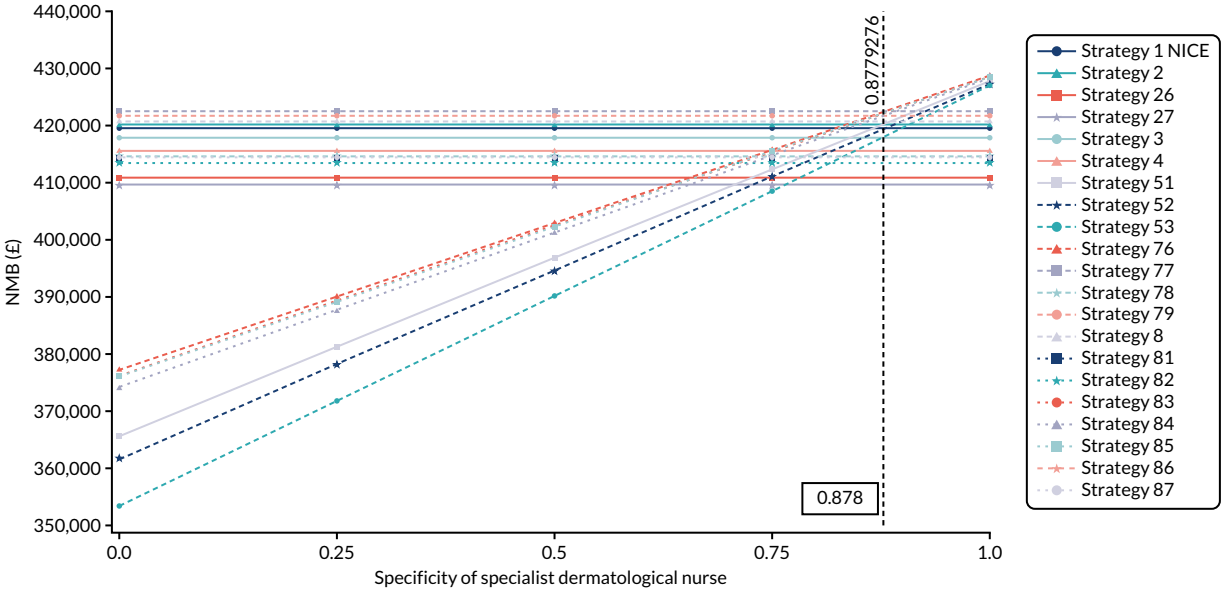


FIGURE 18 One-way sensitivity analysis of specificity of specialist dermatological nurses for stage IB patients.

The results of the costs, effectiveness, ICER and NMB are presented in *Appendix 10*, *Table 43*. In this simplified analysis, the least costly option had the highest NMB for stage IB patients. However, the prognostic test strategy had the highest NMB in stage IA patients and performed slightly better than the NICE strategy in both subgroups. When the associated CEACs are taken into account, the NICE strategy is likely to be the most cost-effective strategy at WTP thresholds of > £12,000 for stage IA and £26,000 for stage IB (*Figures 19* and *20*).

In these illustrated examples (see *Figures 19* and *20*), only a few of the possible strategies were compared in this analysis (compared with the 20 strategies in the base-case analysis). Furthermore, the way in which the prognostic test is incorporated into surveillance strategy in this sensitivity analysis may not be optimal. Both these points will have an effect on what is deemed cost-effective. What these results illustrate is that, other things being equal, increases to the sensitivity and specificity of the prognostic test (or reductions in

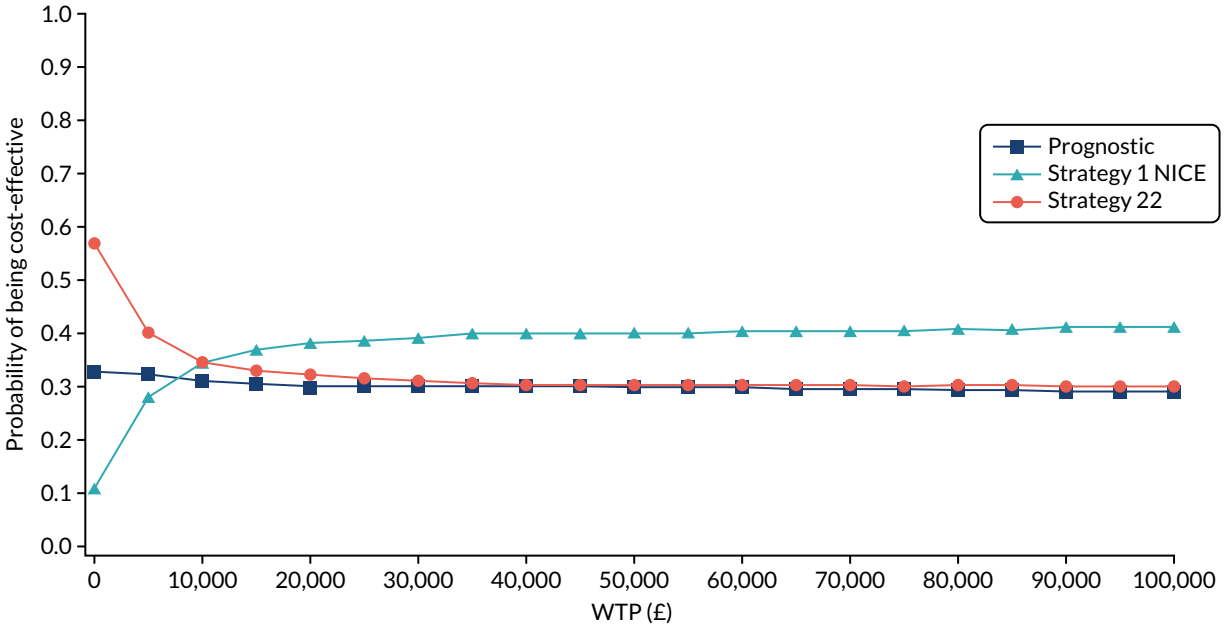


FIGURE 19 The CEAC for prognostic test strategy for stage IA patients (only three strategies included).

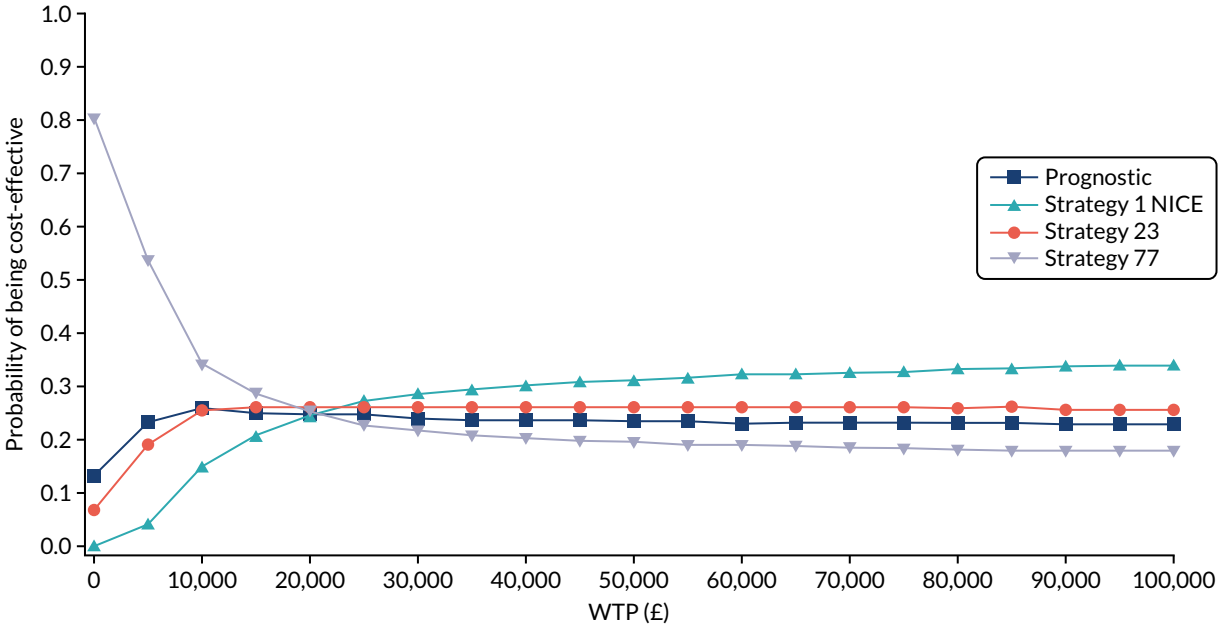


FIGURE 20 The CEAC for prognostic test strategy for stage IB patients (only four strategies included).

the cost of the prognostic test) could markedly change which strategy is considered cost-effective (and may lead to additional strategies being considered cost-effective). If a prognostic test that can acceptably differentiate stage IA and IB into high and low risk of recurrence were to become available in the future at a reasonable price, then the recommended surveillance strategy may change.

Value-of-information analysis

The usefulness of a value-of-information/expected value of perfect information (EVPI) analysis is in calculating a maximum reasonable price for information. Additional evidence is valuable because it can improve patient outcomes by resolving existing uncertainty about the cost-effectiveness of the interventions available, thereby informing treatment choices for subsequent patients.[229] It also allows for the comparison of the cost of uncertainty with the cost of obtaining additional evidence.

The overall EVPI for IA is £5545 per person affected by the decision, and the population EVPI is £30,824,655 per year in England. The overall EVPI for stage IB is £6339 per person affected by the decision, and the population EVPI is £20,256,750 per year in England. We assume that the population EVPI exceeds the expected cost of research and consider the EVPPI for groups of parameters for which it is likely that a new study (or studies) would be informative for the whole group, rather than for individual parameters.

The parameters that were considered to contribute most significantly to EVPPI estimates for both stage IA and IB by the SAVI tool were probabilities of transitioning between stages (natural history), diagnostic accuracy, health utility values and recurrence of melanoma on the overall decision uncertainty in the economic evaluation model. These parameters are potentially relevant for the design of future research, as well as for broader policy questions. *Table 20* reports the EVPPI for the four groups of parameters, with the overall decision EVPI at a threshold of £20,000 per QALY. The EVPPI associated with probabilities of transitioning between stages is relatively high for both stage IA (see *Table 20*) and stage IB (*Table 21*). This research is more likely to take the form of an epidemiological study. Other parameters, such as diagnostic accuracy and utility values, require other forms of study design.

For an NHS decision-maker, the EVPPI for England per year is based on the number of patients affected by the decision each year. This was assumed to be equal to the annual incidence of stage I melanoma. In 2017, 8555 patients were diagnosed with stage I melanoma.[226] The split in IA and IB patients (63% : 37%)

TABLE 21 The EVPPIs for pertinent parameters (stage IB)

Parameters	Per person EVPPI (£)	EVPPI for England per year (£)	EVPPI for England over 20 years (£)
Probabilities of transitioning between stages	4109	22,843,677	456,873,547
Diagnostic accuracy	3013	9,628,267	192,565,350
Health utility values	1371	4,381,133	87,622,667
Recurrence of melanoma	617	1,971,670	39,433,396

was taken from incident data from a German Registry data set.[209] EVPI is expressed for the total population of patients who stand to benefit from additional information over the expected lifetime of the technology. The EVPPI over an arbitrary 20-year time horizon is presented.

Discussion

This chapter presented the methods and results of a cost-effectiveness analysis that compared alternative surveillance strategies for people who have been treated for AJCC stage IA and IB melanoma by the NHS in the UK. The paucity and limitations of existing evidence from English/UK data sources resulted in many data values from various international sources being used in the economic model. Therefore, the evidence on cost-effectiveness should be treated cautiously, in part because of the inherent problems in combining data from multiple sources.

Summary of the cost-effectiveness analysis

The base-case PSAs economic model presented in this report compared 20 different follow-up surveillance strategies for stage IA melanoma and 20 different strategies for stage IB melanoma. Cumulative costs and QALYs were compared over a lifetime time horizon, with the results suggesting that a less intensive follow-up strategy for both stages IA and IB has a higher NMB than the surveillance strategy recommended in the NICE guidelines. The results indicate that, for the strategies examined, the main difference is in the costs of each strategy, and there are only minimal differences observed in QALYs between strategies. Strategy 22 for stage IA and strategy 77 for stage IB are the least costly options. However, in the base-case analyses, only restricted strategies were analysed and, by omitting relevant comparators, an underestimation of ICERs may have occurred.[230] However, the NMB statistic highlights the similarity of all evaluated strategies.

The reported one-way sensitivity analyses carried out on the base-case model results suggest that altering the probability of self-diagnosis did not make a major impact on the NMB. Recurrence rates are estimated to be 0.22% for stage IA and 0.46% for stage IB per month and, even if all patients had a recurrence, the analysed surveillance strategies would still have a positive NMB. This suggests that surveillance is worthwhile.

There is uncertainty in the appropriate utility values to use in the base-case analysis. In the one-way sensitivity analysis, varying utility values associated with recurrence between zero and one was explored. As utility values associated with recurrence increase, so does the NMB. It is likely that most recurrences are going to be stage I and utility values are likely to be > 0.7. However, the analysis is not able to distinguish between strategies and, therefore, is not as informative to decision-makers.

Strategies with dermatological nurses/CNSs as the main health-care professional performing the screening were the least cost-effective, despite their lower resource use cost. This was because their diagnostic accuracy and, most importantly, their specificity was lower than those of dermatologists/surgeons.

The consequence of this is that patients who are initially seen by the dermatological nurse/CNS accrue further costs because of the additional diagnostic tests performed.

A hypothetical prognostic test is likely to result in some savings, relative to the current NICE strategy, and may have a role in surveillance. Further work is warranted to identify a relevant and viable prognostic test and how surveillance strategies may be altered to accommodate such a test in a cost-effective manner. The analysis reported in this chapter has been partial with respect to how surveillance strategies may be altered to accommodate a prognostic test; further work would be warranted if and when a viable prognostic test becomes available.

Value-of-information analyses, and specifically EVPPI, were also carried out on the base-case analysis. This analysis sought to estimate the value of removing uncertainty around particular parameters or groups of parameters in the model. The results indicate that further research is valuable and warranted in removing uncertainty around the diagnostic accuracy of health-care professionals, the probability of transitioning between stages, health-state utility values and the recurrence of melanoma to make a confident decision to change current NICE guidelines.

Strengths and limitations

The main strength of the economic evaluation is that a de novo model has been created to answer the research question from the perspective of the NHS. The research team has attempted to use rigorous and systematic methods to obtain parameter inputs into the economic evaluation. These were then assembled in the economic model, whose structure was informed by detailed discussions with the clinical members of the research team. One of the most important challenges faced when conducting this economic evaluation was incorporating patient behaviour. A unique feature of this model is that it considers self-diagnosis by patients and the heightened anxiety melanoma survivors experience that result in 'false-alarm' appointments.

A number of limitations in the economic evaluation need to be acknowledged. With respect to patient characteristics, the subtype of melanoma was not accounted for in the model. The location of the primary site of the melanoma was available in the Durham cohort, but was not utilised in the model. Accordingly, it was assumed that recurrence is independent of primary location of melanoma. Recent Australian data suggested that head and neck location of the primary tumour had the highest rate of 2-year recurrence, compared with primary tumours occurring on the trunk (HR 1.67, 95% CI 1.01 to 2.67), with upper and lower limbs having lower HRs than the trunk area.[231]

The model was based on current NICE guidelines; given the complex nature of the model, the assessment of structural uncertainty (i.e. whether or not relevant processes are reflected in the model and whether or not the model reflects reality) was considered with respect to the number of follow-up regimens considered, including the frequency and duration of follow-up, and who completes that follow-up. Other areas of structural uncertainty were not addressed. Few data were available for many of the model parameters. What data were available were not ideally suited to the research question being addressed. For example, health-state 'utilities' values are based on patients with first primary melanoma diagnosis, rather than recurrence. Utility values are required to estimate QALYs. In this analysis, data were taken from a meta-analysis of utility values, which could be questioned.[216]
As highlighted in an editorial,[232] instruments or questionnaires used should be sensitive to the domains of quality of life that are likely to change as a result of the disease, the routine treatment or a targeted intervention. It is unclear whether or not the EQ-5D is sufficiently sensitive to changes in some of the domains of quality of life that might occur. For example, a European study[233] reported that over half the patients with melanoma in their study reported anxiety. However, the EuroQol-5 Dimensions, five-level version, instrument was somewhat insensitive to detecting this.[233]

Given the limitations of the data, the base-case and sensitivity analyses were not able to provide a strong basis on which the NHS could draw robust conclusions about what surveillance strategy would be best for the UK to adopt. Instead, 'best bets' that would be worthy of further consideration were identified.

Impact and implications

The impact of the work presented in this chapter is important for patients, practitioners, the NHS and for researchers. One key feature of melanoma surveillance is the important role of patients (and their partners) in detecting changes in their moles, and thus detecting recurrences or new primary melanomas. This suggests that patient education is important. The evidence suggests that dermatologists have the highest accuracy in detecting melanomas. For nurses to be considered the main health-care provider in surveillance, further work would be needed to determine and, if necessary, support the development of their diagnostic performance. Should their diagnostic performance be less than dermatologists, then training may be worthwhile. This is because, other things being equal, nurses providing surveillance would be less costly than dermatologists and it may be relatively easier to increase the cadre of nurses able to provide surveillance, rather than increasing the numbers of dermatologists.

Although no viable stratification approach above AJCC staging has thus far been identified, should one be developed, then the initial results suggest that there may be merit in further research to identify how it could be used to refine surveillance in the optimal way.

Chapter 7 Discussion

Summary of findings

Cutaneous melanoma is a cancer developing from melanocytes, which are the pigment-producing cells in the skin. It is one of the most deadly of all skin cancers, with metastatic disease being highly aggressive.[1,2] Until the introduction of targeted therapies and immunotherapies, median overall survival was between 6 and 10 months once metastasis had occurred.[3] In the UK, it is the leading cause of cancer-related death among people aged 20–35 years.

The incidence of melanoma is increasing worldwide; currently, approximately 2% of the population develop melanoma each year.[5] In the UK, there are approximately 17,000 cases of melanoma per year; the incidence rate has increased by 134% since the 1990s, and melanoma now makes up 5% of new cancer cases. Reflecting the impact on mortality, melanoma affects a disproportionate number of people aged < 50 years compared with other cancers.[7]

Primary melanomas are staged according to the AJCC staging criteria and the focus of this research has been on AJCC stage I melanomas. Although these tumours have the lowest mortality risk (at approximately 14% over 10 years) compared with other stages of the disease, approximately 10,500 of the 17,000 melanomas occurring each year are stage I. Thus, for the majority of people who develop stage I melanoma, surgical treatment is effective. However, for a sizable minority of people, the disease recurs. Hence, follow-up and surveillance of those initially treated for stage I melanoma is required.

How surveillance is organised for stage I disease varies considerably between countries. In the Netherlands, following excision of stage IA disease, individuals receive a one-off appointment 1 month after diagnosis.[58] In contrast, the German guidelines recommend that individuals treated for stage IA disease receive a follow-up appointment every 6 months for the first 3 years and are then followed up annually for a further 7 years.[59] The current NICE guideline recommendation for stage IA lies between these two, with follow-up visits recommended every 3–6 months for 1 year.[16] The 2019 surveillance report for NICE suggested no change to recommendations for surveillance of stage I disease.[54] The differences in recommendations for surveillance between guidelines of stage I melanomas (summarised in *Chapter 2*) partly reflect the fact that the evidence base is from non-randomised and anecdotal evidence, along with expert opinion. Recommendations are based on the assumption that earlier detection of metastatic disease results in improved overall outcome, but often do not consider the potential physical, psychological and economic costs of these regimens.

With the increasing melanoma rates, the pressures on the NHS will increase. One approach to managing these pressures is to reconsider how surveillance is organised for those treated for stage I disease. To begin this, consideration must first be given to what is known to form judgements as to whether or not the current evidence base is sufficient to justify changes in practices and, if it is, what is the effectiveness and cost-effectiveness of alternative ways of organising surveillance.

This evidence synthesis work has attempted to do this. It started by systematically reviewing the existing evidence base on the relative effectiveness of surveillance and follow-up strategies following surgical excision of a stage I tumour. Only one RCT (reported in two papers) was eligible.[85] The one study[71] for which data were available suggested that an educational intervention for patients and their partners improved self-identification of new primary melanomas (rather than locoregional metastatic disease). Even if the data were available from both studies, quantitative synthesis would not be possible, as they employed different follow-up strategies. Furthermore, because the study was not conducted in the UK, the applicability of the findings to the UK may be limited and the evidence from

this systematic review was judged to be of low certainty[234] because of the few data available. The findings from this systematic review are similar to those found by Cromwell *et al.*[25]

To address the limitations of the empirical evidence, an evidence synthesis approach was used to construct a comparison of alternative surveillance strategies following treatment for stage I melanoma. As part of this evidence synthesis, a systematic review of the prognostic accuracy of risk prediction models to predict recurrence, new primary tumours and metastases was conducted. The purpose of this review was to identify if there was any evidence that might enable surveillance to be stratified by risk. Such stratification could enable low-risk patients to safely receive less intensive surveillance than higher-risk patients. This review identified 11 different risk prediction models, with the number of predictors per model ranging from 3 to 11. The models differed in the predictors used, the outcomes of interest and the statistical measures used to assess model performance. Consequently, no quantitative synthesis of their results was performed. None of the identified risk prediction models has undergone rigorous validation. The data elements most commonly used in these models are patient demographic information or histological features of the primary tumour.[136] Overall, the identified evidence did not allow accurate prognostication of melanoma.

Any surveillance strategy would also require that further disease can be accurately diagnosed if it occurs. For this reason, a systematic review was conducted that explored the diagnostic test accuracy of FNB and ultrasonography to detect recurrence and locoregional metastases during follow-up of stage I melanoma. Despite extensive searching, only two studies assessing different index tests met the inclusion criteria. One considered FNB[163] and was judged to be at high risk of bias and one assessed ultrasonography[164] and was also considered to be at high risk of bias. The findings reported for diagnostic accuracy of FNB in patients who were diagnosed initially as having stage I melanoma were comparable to those reported at stages II–IV.[163] They were also similar to those reported in another systematic review on the subject.[148] Similarly, the diagnostic performance of ultrasonography estimated by Kruger *et al.*[164] was similar to that reported by two different reviews, even though no analysis was conducted by disease stage.[147,170]

No existing economic evaluation that directly addressed the study question was identified. Therefore, using the information gathered from the systematic reviews and discussions among the study team, a Markov microsimulation model was developed. The economic evaluation compared alternative surveillance strategies for people who have been treated for AJCC stage IA and IB melanoma provided by the NHS in the UK.

Initially, 400 surveillance strategies were defined for stage IA and 600 for stage IB. The clinical team reviewed these and implausible strategies were excluded from further consideration. Initial modelling focused on 75 strategies for stage IA and 87 strategies for stage IB. Surveillance strategies varied in terms of who performed the surveillance (dermatologist, surgeon or cancer nurse specialist), frequency of follow-up (i.e. every 3, 4, 6 or 12 months) and duration of follow-up (for stage IA, follow-up was 1, 3, 5 or 10 years, and for stage IB, follow-up was 1, 3, 5, 10 or 20 years).

This initial modelling showed that strategies involving a cancer nurse specialist were highly unlikely to be cost-effective. The reason for this is that, although the cost of using cancer nurse specialists to provide surveillance was low, the data available suggested that cancer nurse specialists were not good at identifying people without disease. Hence, a considerable number of individuals were referred on for further (costly) investigation. Nevertheless, the evidence on the diagnostic performance of cancer nurse specialists was poor. However, this suggests that training cancer nurse specialists to accurately diagnose recurrent melanoma could potentially increase capacity within the NHS to undertake surveillance. A sensitivity analysis showed that, should it be possible to do this so that the diagnostic performance of cancer nurse specialists approached that estimated for dermatologists, strategies involving cancer nurse specialists might be cost-effective. Currently, training support for cancer nurse

specialists in this role is limited, but the results of the analysis suggest that work to develop training for these staff may be worthwhile.

Following the initial modelling, 20 strategies for stage IA and 20 for stage IB disease were selected based on relevance to the NHS, the potential to be cost-effective and the ability to extrapolate findings to similar strategies that were not modelled. For each stage of the disease, the current NICE recommendations for surveillance were modelled. The results of these analyses showed that there were only very small differences in QALYs between strategies, with the main differences being in costs. This was primarily because rates of recurrence were expected to be comparatively low (approximately 14% of people who have been treated for stage I experience a recurrence over 10 years).

For stage IA patients, the strategy of follow-up once for 1 year by a dermatologist was, on average, the least costly and most effective. For stage IB, the strategy of follow-up once for 1 year by a dermatologist was, on average, the least costly and least effective. If all strategies were equally likely to be cost-effective, then each strategy would have a 5% chance of being considered cost-effective at any given threshold. For both stages IA and IB, the strategy of follow-up once for 1 year by a dermatologist was the most likely to be considered cost-effective at a £20,000 threshold per QALY (13% and 12.9% for stages IA and IB, respectively). For stage IA disease, the NICE strategy has a $\approx$ 9% chance of being considered cost-effective at the same threshold. For stage IB, the NICE strategy (surveillance by a dermatologist every 3 months for the first 3 years and then every 6 months for the next 2 years) never has a probability of $>$ 6% of being cost-effective over any of the values for society's WTP for a QALY.

Overall, no strategy clearly stands out as superior, unless the NHS was to make a decision between the strategies based solely on cost (follow-up once for 1 year by a dermatologist had a $>$ 50% chance of being less costly for both stages IA and IB disease). This suggests that, based on the reviews and economic modelling, there are no firm grounds to suggest changing current NICE recommendations. However, the analyses also indicate that there are plausible strategies that could represent a more efficient use of NHS resources for both stages IA and IB. These strategies include ones both less intensive and more intensive than the strategy recommended by NICE. This suggests that further research to explore these strategies may be worthwhile.

In the model, a high rate of self-diagnosis of recurrence was assumed (60% of all recurrences were assumed to be self-diagnosed within 12 months). This high rate of self-diagnosis restricts the gains that any surveillance strategy could provide. The evidence supporting the self-diagnosis rate is weak, and rates of self-diagnosis could be much lower. When modelling the impact of much lower rates of self-diagnosis, those strategies that are more intensive than the NICE-recommended strategies become more likely to be considered cost-effective. However, the gains in QALYs are small, primarily because the underlying rates of recurrence are low for stage I melanomas. For some groups, rates of self-diagnosis may be low[235] (e.g. males aged $>$ 50 years, those with a darker skin colour, patients living alone); a valid question is whether or not it would be worth investing in melanoma awareness campaigns designed to increase the rates of self-diagnosis.

A hypothetical scenario in which a validated prognostic test, be it a risk factor-based model, a biomarker or a combination of both, is available to the NHS was also modelled. As *Chapter 4* shows, no good risk prediction model currently exists. However, there is considerable interest in developing tools to help with prognostication. Indeed, initial scoping searches for prognostic biomarkers for melanoma conducted as part of the review of risk prediction models identified several tens of thousands of potentially relevant titles and abstracts. Such tools offer the potential to focus surveillance resources on those most likely to experience a recurrence or metastasis, while allowing low-risk individuals to be either discharged or followed up in a far less intensive manner. The analyses conducted illustrated the potential scope for such tests, especially for the higher-risk stage IB patients, provided the costs of the test were modest (we modelled a cost per test of £250). The analysis also illustrated that much more

thought could go into how a surveillance strategy could be adapted to incorporate such a test. This is because the analyses conducted may not have optimised the use of such a test. However, work in this area is of less importance until a viable tool becomes close to being validated for use.

Uncertainty about the assessment

Key uncertainties for the assessment relate to the paucity of data available for evidence synthesis, few of which related to the latest version of the AJCC staging. The various clinical guidelines when making recommendations on surveillance reflect this, as did the review of different surveillance strategies. Considerable efforts were made in the three reviews reported (see *Chapters 3–5*) to identify relevant data and all used relevant tools to assess the quality of the evidence identified that was relevant to the questions posed by that review.

The review of different surveillance strategies identified just one comparative study.[85] Given the very large number of potential strategies that could be considered for surveillance, the evidence base is sparse. However, although there is a need for well-designed studies in this area comparing alternative strategies, it is currently unclear which strategies should be compared. The economic evaluation identified a very large number of potential strategies, of which a small number might be worthy of further consideration. Further work would be needed to develop strategies that might be relevant to trial further. As a surveillance strategy is a complex intervention containing several components, such development and evaluation work should be placed in the context of the Medical Research Council's framework for complex interventions.[236]

The current evidence synthesis provides no evidence that any possible alternative surveillance strategies should include a tool to help identify those at low and those at high risk of recurrence and progression. A number of risk prediction models were identified, but none met the recommendations set out by TRIPOD.[135] Only a small number of models were externally validated, limiting judgements on their applicability to the NHS. However, considerable research efforts are ongoing internationally to develop tools that could be used to identify those at low risk and those at high risk of recurrence and progression. The economic modelling showed that, should a tool be developed (even with only modest performance of 0.80 sensitivity and specificity), further work to determine whether or not and how to integrate such a tool into a surveillance strategy is necessary.

There was also considerable uncertainty about how and for whom a diagnosis of recurrent disease might be made. With respect to diagnostic tests, based on clinical judgement, the focus was on FNB and ultrasonography to detect melanomas. The uncertainty surrounding the performance of different types of practitioners (as discussed earlier in this chapter for cancer nurse specialists) was one aspect of this uncertainty, but few data were available for the performance of the specified tests for stage I melanoma. The data identified appeared to be consistent with those drawn from other reviews, which did not differentiate by stage.[147,169,170]

The uncertainties that remain after the reviews are compounded in the economic model, as the economic model is a further level of evidence synthesis. In addition, the economic model itself may not have fully captured the complexity of melanoma follow-up. Although considerable efforts were used to develop a model that reflected the disease process and UK practice, inevitably, simplifications were made in the model structure. Furthermore, although a very large number of surveillance strategies were considered in the base-case analysis, only a restricted number of strategies were analysed. Although interpretation of this restricted number was not straightforward, by omitting relevant comparators, ICERs may have been underestimated[230] and probabilities of a given strategy being cost-effective may have been misspecified. Furthermore, the economic evaluation considered NHS costs only. Costs to patients and their families of accessing surveillance and treatment were omitted. Such costs may be disproportionately felt by those least able to bear these costs. Other aspects of

patient outcomes that might not be adequately captured by the EQ-5D and the utilities included in the model were also omitted. These include the psychological effects of surveillance and diagnosis and treatment.

Given the uncertainties, the economic evaluation sought to identify the areas where further research would be worthwhile. For this, it estimated the EVPI and the EVPPI. These analyses relate to the value of information for the model (so if the model is misspecified, then the EVPI and EVPPI will be misspecified). Furthermore, the EVPI and EVPPI represent the total value of removing all uncertainty. In reality, any study could remove only some of this uncertainty. Nevertheless, the EVPI and EVPPI show that there could be considerable value in removing uncertainty overall, and specifically in research, to improve estimates for the probabilities of transitioning between stages for both stages IA and IB. There may be less (but still considerable) value in research around improving estimates of diagnostic accuracy and for utility values.

Chapter 8 Conclusions

Based on the results of the evidence synthesis and economic evaluation, the following implications for practice and research can be drawn out.

Implications for practice

- Few data were available specific to surveillance of people after treatment for melanoma. Furthermore, few data were available for key components of a surveillance strategy that could be used to model alternative strategies. Therefore, the results are imprecise and considerable uncertainty exists.
- There is insufficient evidence to recommend any changes to the current guidelines produced by NICE with respect to surveillance. For stage IA disease, there are plausible surveillance strategies that may perform better than the current recommendation of clinical follow-up by a dermatologist every 3–6 months for 1 year. However, the NICE strategy still performs comparatively well compared with these. For stage IB disease, the NICE-recommended strategy of follow-up every 3 months for 3 years then every 6 months for a further 2 years performs poorly compared with other strategies considered, but there is currently insufficient evidence to support any changes.
- Surveillance strategies whereby the clinical follow-up is conducted by a cancer nurse specialist may ease pressure on dermatologists. However, methods to enhance their diagnostic performance may be needed, as the current limited evidence base suggests that their ability to correctly identify who has or does not have a recurrence is lower than that of dermatologists.
- Encouraging and supporting patients in making accurate self-diagnosis of recurrence in stage I disease may reduce the need for any active surveillance strategy for those initially treated for stage I disease.

Implications for research

It is tempting to recommend that a RCT should be conducted to compare alternative surveillance strategies. However, a surveillance strategy is a complex intervention and research should first establish what sensible comparators should be used against current practice. What an appropriate comparator would be may vary between stages IA and IB disease, and establishing this requires improved evidence on:

- How disease in patients with stage I disease develops over time especially as defined by the latest AJCC staging criteria. The economic modelling shows that both the incidence of recurrent and metastatic disease over time and how the disease progresses are important. Such data would inform whether or not a more or less intensive surveillance strategy than the ones recommended by NICE for stage IA or IB should be considered. The value-of-information analysis suggested that this is where there would be the greatest value in removing all uncertainty. Potential research techniques may include constructing cohorts relying on routine data collection.
- How well recurrent and metastatic disease is diagnosed. This may also include research designed to improve the diagnostic performance of particular practitioner groups in the clinical workforce, especially when their increased use may alleviate current or impending capacity constraints in the availability of dermatologists. Study designs may include behaviour change studies designed to improve the diagnostic performance of practitioners.

- Low-cost tools that can better stratify patients into low or high risk of future recurrence and metastasis. This would help develop better stratified follow-up, allowing some of the ≈ 85% of patients treated for stage I melanoma based on the AJCC's seventh edition classification who do not develop further disease to be safely discharged more quickly. It would also allow more focused use of scarce resources for those at higher risks.
- Identifying patients' preferences for alternative methods of surveillance, the impacts on health-related quality of life of surveillance (or no surveillance) and the longer-term consequences of melanoma.

Acknowledgements

We would like to thank the following people:

- Kirsty Laing for contributing to the literature searches for *Chapter 2*
- those who assisted with the review reported in *Chapters 3* and *4* – Ryan Kenny and Hosein Shabaninejad, and Akvile Stoniute for taking part in the updating of the screening for reviews
- Professor J Robinson for providing data for the subset of stage I melanoma patients for use in the review of surveillance strategies
- those who assisted with the review reported in *Chapter 5* – Theophile Birgirumurame for translating a French article, Cristina Fernandez-Garcia for taking part in the translation for one Spanish paper, Svetlana Glinianaia for translating a Russian article for *Melanoma Review* and Jac Dinnes for providing timely assistance in providing specialist diagnostic accuracy data
- Ed Wilson for providing timely assistance in model development and data gathering for Review 3
- Eugenie Johnson for providing administrative support to the project
- Janice Legge for taking part in the administrative support
- Dr Janine Graham for providing expert opinions related to oncological elements of the study.

Contributions of authors

Luke Vale (https://orcid.org/0000-0001-8574-8429) (Professor of Health Economics) was responsible for management of the study and contributed to the conception of the study and the conduct of all elements.

Patience Kunonga (https://orcid.org/0000-0002-6193-1365) (Systematic Reviewer) assisted in design of the surveillance review with Margaret Astin and was responsible for the design of the review for risk prediction models.

Diarmuid Coughlan (https://orcid.org/0000-0002-5348-3750) (Health Economist) and **Vasileios Kontogiannis (https://orcid.org/0000-0002-9823-992X)** (Health Economist) assisted in the design of the economic evaluation and were equally responsible for the conduct and reporting of the economic evaluation.

Margaret Astin (https://orcid.org/0000-0002-3482-5567) (Systematic Reviewer) was responsible for the protocol and conduct of the diagnostic accuracy review and assisted with the surveillance and prediction models reviews, from data extraction to reporting of study results.

Fiona Beyer (https://orcid.org/0000-0002-6396-3467) (Senior Research Associate – Evidence Synthesis) provided methodological expertise and advice for the systematic review element of the study, and was responsible for developing searches for the surveillance review.

Catherine Richmond (https://orcid.org/0000-0002-2940-5197) (Information Specialist) was responsible for developing searches for the review of diagnostic accuracy, managing update searches and reference management.

Dor Wilson (https://orcid.org/0000-0002-3034-001X) (Information Specialist) was responsible for developing the searches for the review of risk prediction models.

Dalvir Bajwa (https://orcid.org/0000-0002-0965-7012) (Dermatologist) designed and collected the data from the patient cohort.

Mehdi Javanbakht (https://orcid.org/0000-0002-8661-8439) (Health Economist) contributed to the conception, study and design of the economic evaluation.

Andrew Bryant (https://orcid.org/0000-0003-4351-8865) (Statistician) contributed to the systematic reviews of risk prediction models and diagnostic tests.

Wanwuri Akor (https://orcid.org/0000-0001-9291-3816) (Health Services Researcher) conducted the structured review of clinical guideline recommendations on surveillance following treatment for stage I melanoma.

Dawn Craig (https://orcid.org/0000-0002-5808-0096) (Professor of Evidence Synthesis) advised on the conduct of the study.

Penny Lovat (https://orcid.org/0000-0002-6657-6825) (Professor of Cellular Dermatology and Oncology), **Marie Labus (https://orcid.org/0000-0002-8339-7423)** (Business Development Manager), **Batoul Nasr (https://orcid.org/0000-0002-4670-0171)** (Consultant Dermatologist) and **Timothy Cunliffe (https://orcid.org/0000-0001-6081-5891)** (Consultant Dermatologist) contributed to the conception of the study and advised on the conduct of the study.

Helena Hinde (https://orcid.org/0000-0003-0057-6758) (Dermatology Cancer Nurse Specialist) advised on the conduct of the study.

Mohamed Shawgi (https://orcid.org/0000-0002-1021-8439) (Consultant Radiologist) advised on the conduct of the diagnostic accuracy review.

Daniel Saleh (https://orcid.org/0000-0003-4849-6933) (Consultant Plastic Surgeon) advised on the conduct of the study.

Pam Royle (https://orcid.org/0000-0001-8033-6556) (Patient Representative), **Paul Steward (https://orcid.org/0000-0002-5460-231X)** (Patient Representative) and **Rachel Lucas (https://orcid.org/0000-0002-3612-6495)** (Patient Representative) contributed to the conception of the study and advised on the conduct of the study.

Robert Ellis (https://orcid.org/0000-0001-6305-2121) (Principal Investigator, Consultant Dermatologist) contributed to the conception of the study and the conduct of all elements.

All authors contributed to the interpretation of the results and the writing/editing of the report.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review.

Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: https://understandingpatientdata.org.uk/data-citation.

References

1. Matthews NH, Li WQ, Qureshi AA, Weinstock MA, Cho E. Epidemiology of melanoma. In Ward WH, Farma JM, editors. *Cutaneous Melanoma: Etiology and Therapy*. Brisbane, QLD: Codon Publications; 2017. https://doi.org/10.15586/codon.cutaneousmelanoma.2017.ch1
2. American Cancer Society. *Cancer Facts and Figures 2018*. URL: www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html (accessed 8 October 2019).
3. 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. *J Clin Oncol* 2008;**26**:527–34. https://doi.org/10.1200/JCO.2007.12.7837
4. Luke JJ, Flaherty KT, Ribas A, Long GV. Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol* 2017;**14**:463–82. https://doi.org/10.1038/nrclinonc.2017.43
5. de Vries E, Sierra M, Piñeros M, Loria D, Forman D. The burden of cutaneous melanoma and status of preventive measures in Central and South America. *Cancer Epidemiol* 2016;**44**(Suppl. 1):100–9. https://doi.org/10.1016/j.canep.2016.02.005
6. World Cancer Research Fund International. *Skin Cancer Statistics*. 2018. URL: www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics (accessed 8 August 2019).
7. Cancer Research UK. *Skin Cancer: Risks and Causes*. 2017. URL: www.cancerresearchuk.org/about-cancer/skin-cancer/risks-causes (accessed 8 August 2019).
8. Ballantine KR, Watson H, Macfarlane S, Winstanley M, Corbett RP, Spearing R, *et al.* Small numbers, big challenges: adolescent and young adult cancer incidence and survival in New Zealand. *J Adolesc Young Adult Oncol* 2017;**6**:277–85. https://doi.org/10.1089/jayao.2016.0074
9. Watson M, Geller AC, Tucker MA, Guy GP Jr, Weinstock MA. Melanoma burden and recent trends among non-Hispanic whites aged 15–49 years, United States. *Prev Med* 2016;**91**:294–8. https://doi.org/10.1016/j.ypmed.2016.08.032
10. Cancer Research UK. *Melanoma Skin Cancer Incidence Statistics*. 2019. URL: www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer/incidence#ref-0 (accessed 8 August 2019).
11. International Agency for Research on Cancer. *Cancer Today*. 2018. URL: http://gco.iarc.fr/today/home (accessed 8 August 2019).
12. Miiskin. *Facts About Melanoma and Other Skin Cancers*. 2019. URL: https://miiskin.com/melanoma-skin-cancer-facts/ (accessed 8 August 2019).
13. Melanoma Patient Network Europe. *Melanoma – The Facts*. 2013. URL: www.melanomapatientnetworkeu.org/melanoma.html (accessed 8 August 2019).
14. Colebatch AJ, Scolyer RA. Trajectories of premalignancy during the journey from melanocyte to melanoma. *Pathology* 2018;**50**:16–23. https://doi.org/10.1016/j.pathol.2017.09.002
15. Miller AJ, Mihm MC. Melanoma. *N Engl J Med* 2006;**355**:51–65. https://doi.org/10.1056/NEJMra052166
16. National Institute for Health and Care Excellence. *Melanoma: Assessment and Management*. 2015. URL: www.nice.org.uk/guidance/ng14 (accessed 17 July 2019).

17. Voss RK, Woods TN, Cromwell KD, Nelson KC, Cormier JN. Improving outcomes in patients with melanoma: strategies to ensure an early diagnosis. *Patient Relat Outcome Meas* 2015;**6**:229–42. https://doi.org/10.2147/PROM.S69351

18. Watts CG, Dieng M, Morton RL, Mann GJ, Menzies SW, Cust AE. Clinical practice guidelines for identification, screening and follow-up of individuals at high risk of primary cutaneous melanoma: a systematic review. *Br J Dermatol* 2015;**172**:33–47. https://doi.org/10.1111/bjd.13403

19. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 8th edn. Chicago, IL: American Joint Committee on Cancer/Springer; 2016.

20. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 7th edn. Chicago, IL: American Joint Committee on Cancer; 2010.

21. Ribero S, Argenziano G, Lallas A, Moscarella E, Benati E, Raucci M, *et al.* Dermoscopic features predicting the presence of mitoses in thin melanoma. *J Dermatol Sci* 2017;**86**:158–61. https://doi.org/10.1016/j.jdermsci.2017.01.013

22. Svedman FC, Pillas D, Taylor A, Kaur M, Linder R, Hansson J. Stage-specific survival and recurrence in patients with cutaneous malignant melanoma in Europe – a systematic review of the literature. *Clin Epidemiol* 2016;**8**:109–22. https://doi.org/10.2147/CLEP.S99021

23. Damude S, Hoekstra-Weebers JE, Francken AB, Ter Meulen S, Bastiaannet E, Hoekstra HJ. The MELFO-Study: prospective, randomized, clinical trial for the evaluation of a stage-adjusted reduced follow-up schedule in cutaneous melanoma patients-results after 1 year. *Ann Surg Oncol* 2016;**23**:2762–71. https://doi.org/10.1245/s10434-016-5263-7

24. Hofmann U, Szedlak M, Rittgen W, Jung EG, Schadendorf D. Primary staging and follow-up in melanoma patients – monocenter evaluation of methods, costs and patient survival. *Br J Cancer* 2002;**87**:151–7. https://doi.org/10.1038/sj.bjc.6600428

25. Cromwell KD, Ross MI, Xing Y, Gershenwald JE, Royal RE, Lucci A, *et al.* Variability in melanoma post-treatment surveillance practices by country and physician specialty: a systematic review. *Melanoma Res* 2012;**22**:376–85. https://doi.org/10.1097/CMR.0b013e328357d796

26. Marsden JR, Newton-Bishop JA, Burrows L, Cook M, Corrie PG, Cox NH, *et al.* Revised UK guidelines for the management of cutaneous melanoma 2010. *J Plast Reconstr Aesthet Surg* 2010;**63**:1401–19. https://doi.org/10.1016/j.bjps.2010.07.006

27. Marsden JR, Newton-Bishop JA, Burrows L, Cook M, Corrie PG, Cox NH, *et al.* Revised U.K. guidelines for the management of cutaneous melanoma 2010. *Br J Dermatol* 2010;**163**:238–56. https://doi.org/10.1111/j.1365-2133.2010.09883.x

28. Weinstein D, Leininger J, Hamby C, Safai B. Diagnostic and prognostic biomarkers in melanoma. *J Clin Aesthet Dermatol* 2014;**7**:13–24.

29. Guo HB, Stoffel-Wagner B, Bierwirth T, Mezger J, Klingmüller D. Clinical significance of serum S100 in metastatic malignant melanoma. *Eur J Cancer* 1995;**31**:924–8. https://doi.org/10.1016/0959-8049(95)00087-9

30. Melanoma Focus. *The Current Role of Sentinel Lymph Node Biopsy in the Management of Cutaneous Melanoma – a UK Consensus Statement*. 2019. URL: https://melanomafocus.com/wp-content/uploads/2019/01/SNB-Consensus-Final-1.pdf (accessed 8 August 2019).

31. Cornett WR, McCall LM, Petersen RP, Ross MI, Briele HA, Noyes RD, *et al.* Randomized multicenter trial of hyperthermic isolated limb perfusion with melphalan alone compared with melphalan plus tumor necrosis factor: American College of Surgeons Oncology Group Trial Z0020. *J Clin Oncol* 2006;**24**:4196–201. https://doi.org/10.1200/JCO.2005.05.5152

32. Madu MF, Deken MM, van der Hage JA, Jóźwiak K, Wouters MWJM, van Akkooi ACJ. Isolated limb perfusion for melanoma is safe and effective in elderly patients. *Ann Surg Oncol* 2017;**24**:1997–2005. https://doi.org/10.1245/s10434-017-5803-9

33. Grünhagen DJ, Verhoef C. Isolated limb perfusion for stage III melanoma: does it still have a role in the present era of effective systemic therapy? *Oncology* 2016;**30**:1045–52.

34. Keenan LG, O'Sullivan S, Glynn A, Higgins M, Flavin A, Brennan S. Clinical review of treatment outcomes and patterns of failure with adjuvant radiotherapy in node-positive malignant melanoma. *J Med Imaging Radiat Oncol* 2017;**61**:258–62. https://doi.org/10.1111/1754-9485.12536

35. Liu JB, Bilimoria KY. Weighing the value of completion nodal dissection for melanoma. *J Surg Oncol* 2016;**114**:281–7. https://doi.org/10.1002/jso.24273

36. Raigani S, Cohen S, Boland GM. The role of surgery for melanoma in an era of effective systemic therapy. *Curr Oncol Rep* 2017;**19**:17. https://doi.org/10.1007/s11912-017-0575-8

37. Weber JS, D'Angelo SP, Minor D, Hodi FS, Gutzmer R, Neyns B, *et al.* Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* 2015;**16**:375–84. https://doi.org/10.1016/S1470-2045(15)70076-8

38. Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, *et al.* Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* 2015;**372**:320–30.

39. Weber J, Mandala M, Del Vecchio M, Gogas HJ, Arance AM, Cowey CL, *et al.* Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. *N Engl J Med* 2017;**377**:1824–35. https://doi.org/10.1056/NEJMoa1709030

40. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, *et al.* Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 2015;**373**:23–34. https://doi.org/10.1056/NEJMoa1504030

41. Ribas A, Puzanov I, Dummer R, Schadendorf D, Hamid O, Robert C, *et al.* Pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory melanoma (KEYNOTE-002): a randomised, controlled, phase 2 trial. *Lancet Oncol* 2015;**16**:908–18. https://doi.org/10.1016/S1470-2045(15)00083-2

42. Dummer R, Ascierto PA, Gogas HJ, Arance A, Mandala M, Liszkay G, *et al.* Encorafenib plus binimetinib versus vemurafenib or encorafenib in patients with BRAF-mutant melanoma (COLUMBUS): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* 2018;**19**:603–615. https://doi.org/10.1016/S1470-2045(18)30142-6

43. Robert C, Karaszewska B, Schachter J, Rutkowski P, Mackiewicz A, Stroiakovski D, *et al.* Improved overall survival in melanoma with combined dabrafenib and trametinib. *N Engl J Med* 2015;**372**:30–9. https://doi.org/10.1056/NEJMoa1412690

44. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* 2011;**364**:2507–16. https://doi.org/10.1056/NEJMoa1103782

45. Long GV, Stroyakovskiy D, Gogas H, Levchenko E, de Braud F, Larkin J, *et al.* Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N Engl J Med* 2014;**371**:1877–88. https://doi.org/10.1056/NEJMoa1406037

46. Long GV, Hauschild A, Santinami M, Atkinson V, Mandalà M, Chiarion-Sileni V, *et al.* Adjuvant dabrafenib plus trametinib in stage III BRAF-mutated melanoma. *N Engl J Med* 2017;**377**:1813–23. https://doi.org/10.1056/NEJMoa1708539

47. National Institute for Health and Care Excellence. *NICE Pathways – Melanoma Overview*. 2019. URL: https://pathways.nice.org.uk/pathways/melanoma (accessed 8 August 2019).

48. D'Aniello C, Perri F, Scarpati GDV, Pepa CD, Pisconti S, Montesarchio V, *et al.* Melanoma adjuvant treatment: current insight and clinical features. *Curr Cancer Drug Targets* 2018;**18**:442–56. https://doi.org/10.2174/1568009617666170208163714

49. Vallejo-Torres L, Morris S, Kinge JM, Poirier V, Verne J. Measuring current and future cost of skin cancer in England. *J Public Health* 2014;**36**:140–8. https://doi.org/10.1093/pubmed/fdt032

50. Rychetnik L, McCaffery K, Morton R, Irwig L. Psychosocial aspects of post-treatment follow-up for stage I/II melanoma: a systematic review of the literature (provisional abstract). *Psycho-Oncology* 2013;**22**:721–36. https://doi.org/10.1002/pon.3060

51. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, *et al.* AGREE II: advancing guideline development, reporting and evaluation in health care. *J Clin Epidemiol* 2010;**63**:1308–11. https://doi.org/10.1016/j.jclinepi.2010.07.001

52. Appraisal of Guidelines for Research and Evaluation (AGREE) Research Trust. *AGREE Enterprise Website*. 2014. URL: www.agreetrust.org/ (accessed 11 September 2019).

53. Barbour A, Millward M, Morton R, Saw R, Cancer Council Australia Melanoma Guidelines Working Party. *Clinical Practice Guidelines for the Diagnosis and Management of Melanoma*. 2018. URL: https://wiki.cancer.org.au/australia/Guidelines:Melanoma (accessed 8 August 2019).

54. National Institute for Health and Care Excellence. *Surveillance Proposal Consultation Document: 2019 Surveillance of Melanoma (NICE Guidelines NG14 and CSG8)*. 2019. URL: www.nice.org.uk/guidance/ng14/documents/surveillance-review-proposal (accessed 3 October 2019).

55. National Comprehensive Cancer Network. *Cutaneous Melanoma*. 2019. URL: www.nccn.org/professionals/physician_gls/default.aspx#melanoma (accessed 2 October 2019).

56. Dummer R, Hauschild A, Lindenblatt N, Pentheroudakis G, Keilholz U, ESMO Guidelines Committee. Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015;**26**(Suppl. 5):v126–32. https://doi.org/10.1093/annonc/mdv297

57. Swetter SM, Tsao H, Bichakjian CK, Curiel-Lewandrowski C, Elder DE, Gershenwald JE, *et al.* Guidelines of care for the management of primary cutaneous melanoma. *J Am Acad Dermatol* 2019;**80**:208–50. https://doi.org/10.1016/j.jaad.2018.08.055

58. Dutch Working Group on Melanoma. *Melanoma Guideline 2012*. 2013. URL: www.oncoline.nl/uploaded/docs/melanoom/201208_vertaling%20Richtlijn%20melanoom%20def.pdf (accessed 1 October 2019).

59. Pflugfelder A, Kochs C, Blum A, Capellaro M, Czeschik C, Dettenborn T, *et al.* S3-guideline 'diagnosis, therapy and follow-up of melanoma' – short version. *J Dtsch Dermatol Ges* 2013;**11**:563–602. https://doi.org/10.1111/ddg.12044

60. Dummer R, Siano M, Hunger RE, Lindenblatt N, Braun R, Michielin O, *et al.* The updated Swiss guidelines 2016 for the treatment and follow-up of cutaneous melanoma. *Swiss Med Wkly* 2016;**146**:w14279. https://doi.org/10.4414/smw.2016.14279

61. Castro LG, Bakos RM, Duprat Neto JP, Bittencourt FV, Di Giacomo TH, Serpa SS, *et al.* Brazilian guidelines for diagnosis, treatment and follow-up of primary cutaneous melanoma – Part II. *An Bras Dermatol* 2016;**91**:49–58. https://doi.org/10.1590/abd1806-4841.20164715

62. Cox NH, Williams HC. The British Association of Dermatologists therapeutic guidelines: can we AGREE? *Br J Dermatol* 2003;**148**:621–5. https://doi.org/10.1046/j.1365-2133.2003.05241.x

63. Einwachter-Thompson J, MacKie RM. An evidence base for reconsidering current follow-up guidelines for patients with cutaneous melanoma less than 0.5 mm thick at diagnosis. *Br J Dermatol* 2008;**159**:337–41. https://doi.org/10.1111/j.1365-2133.2008.08641.x

64. Griffiths CE. The British Association of Dermatologists guidelines for the management of skin disease *Br J Dermatol* 1999;**141**:396–7. https://doi.org/10.1046/j.1365-2133.1999.3029a.x

65. National Comprehensive Cancer Network. *About the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®)*. URL: www.nccn.org/professionals/ (accessed 26 August 2020).

66. National Comprehensive Cancer Network. *NCCN Categories of Evidence and Consensus*. URL: www.nccn.org/guidelines/guidelines-process/development-and-update-of-guidelines (accessed 26 August 2020).

67. ESMO. *ESMO Clinical Practice Guidelines: Melanoma*. URL: www.esmo.org/guidelines/melanoma (accessed 12 December 2020).

68. Dykewicz CA, Centers for Disease Control and Prevention (U.S.). Summary of the guidelines for preventing opportunistic infections among hematopoietic stem cell transplant recipients. *Clin Infect Dis* 2001;**33**:139–44. https://doi.org/10.1086/321805

69. ESMO Guidelines Committee. *Standard Operating Procedures (SOPs) Instructions for Authors and Templates for Standard ESMO Clinical Practice Guidelines (CPGs) and ESMO-MCBS Scores*. 2019. URL: www.esmo.org/content/download/77789/1426712/file/ESMO-Clinical-Practice-Guidelines-Standard-Operating-Procedures.pdf (accessed 1 October 2019).

70. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B, Bowman M. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *J Am Board Fam Pract* 2004;**17**:59–67. https://doi.org/10.3122/jabfm.17.1.59

71. Association of Community Cancer Centers. *Cancer Survivorship Care*. 2011. URL: www.oncoline.nl/cancer-survivorship-care (accessed 1 October 2019).

72. Association of Community Cancer Centers. *Cancer Screening for Psychosocial Distress*. 2011. URL: www.oncoline.nl/screening-for-psychosocial-distress (accessed 1 October 2019).

73. German Guideline Program in Oncology. *S3-Leitlinie zur Diagnostik, Therapie and Nachsorge des Melanoms*. 2018. URL: www.leitlinienprogramm-onkologie.de/leitlinien/melanom/ (accessed 8 August 2019).

74. Scottish Intercollegiate Guidelines Network. *SIGN 50: A Guideline Developer's Handbook*. 2015. URL: www.sign.ac.uk/sign-50.html (accessed 2 October 2019).

75. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *JAMA* 1993;**270**:2093–5. https://doi.org/10.1001/jama.1993.03510170083036

76. Francken AB, Bastiaannet E, Hoekstra HJ. Follow-up in patients with localised primary cutaneous melanoma. *Lancet Oncol* 2005;**6**:608–21. https://doi.org/10.1016/S1470-2045(05)70283-7

77. National Institute for Health and Care Excellence. *Melanoma: Assessment and Management. 1.9 Follow-up After Treatment for Melanoma*. 2015. URL: www.nice.org.uk/guidance/ng14/chapter/1-recommendations#follow-up-after-treatment-for-melanoma-2 (accessed 8 August 2019).

78. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic reviews and Meta-Analyses: the PRISMA statement. *J Clin Epidemiol* 2009;**62**:1006–12. https://doi.org/10.1016/j.jclinepi.2009.06.005

79. Javanbakht M, Ellis R, Vale L, Lovat P, Bryant A, Kunonga P, *et al. Optimal Surveillance Strategies for AJCC Stage I Cutaneous Melanoma Post Primary Tumour Excision: An Evidence Synthesis and Economic Evaluation*. 2019. URL: www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018086784 (accessed 23 May 2019).

80. Whitlock EP, Lopez SA, Chang S, Helfand M, Eder M, Floyd N. AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. *J Clin Epidemiol* 2010;**63**:491–501. https://doi.org/10.1016/j.jclinepi.2009.03.008

81. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;**5**:210. https://doi.org/10.1186/s13643-016-0384-4

82. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, *et al. Cochrane Handbook for Systematic Reviews of Interventions Version 6.1 [updated September 2020]*. London: Cochrane; 2020.

83. Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, *et al.* A revised tool for assessing risk of bias in randomized trials. In Chandler J, McKenzie J, Boutron I, Welch V, editors. *Cochrane Methods. Cochrane Database of Systematic Reviews 2016, Issue 10 (Suppl 1)*. Hoboken, NJ: John Wiley & Sons, Inc.; 2016. pp. 29–30.

84. McMaster University, Evidence Prime Inc. *GRADEpro GDT: GRADEpro Guideline Development Tool*. URL: https://gradepro.org/ (accessed 8 August 2019).

85. Robinson JK, Wayne JD, Martini MC, Hultgren BA, Mallett KA, Turrisi R. Early detection of new melanomas by patients with melanoma and their partners using a structured skin self-examination skills training intervention: a randomized clinical trial. *JAMA Dermatol* 2016;**152**:979–85. https://doi.org/10.1001/jamadermatol.2016.1985

86. Robinson JK, Gaber R, Hultgren B, Eilers S, Blatt H, Stapleton J, *et al.* Skin self-examination education for early detection of melanoma: a randomized controlled trial of Internet, workbook, and in-person interventions. *J Med Internet Res* 2014;**16**:e7. https://doi.org/10.2196/jmir.2883

87. Turrisi R, Hultgren B, Mallett KA, Martini M, Robinson JK. Comparison of efficacy of differing partner-assisted skin examination interventions for melanoma patients: a randomized clinical trial. *JAMA Dermatol* 2015;**151**:945–51. https://doi.org/10.1001/jamadermatol.2015.0690

88. Deckers E, Hoekstra-Weebers J, Damude S, Francken A, Ter Meulen S, Bastiaannet E, *et al.* The MELFO-Study: a multi-center prospective randomized clinical trial on the effects of a reduced stage-adjusted follow-up schedule on cutaneous melanoma IB–IIC patients: results after 3-years. *Ann Surg Oncol* 2018;**25**:S40. https://doi.org/10.1245/s10434-019-07825-7

89. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;**98**:691–8. https://doi.org/10.1136/heartjnl-2011-301247

90. Balch CM, Gershenwald JE, Soong SJ, Thompson JF, Atkins MB, Byrd DR, *et al.* Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol* 2009;**27**:6199–206. https://doi.org/10.1200/JCO.2009.23.4799

91. Bigby M, Popescu C. Sentinel lymph node biopsy in melanoma. *BMJ* 2015;**351**:h5940. https://doi.org/10.1136/bmj.h5940

92. Ferrara G, Partenzi A, Filosa A. Sentinel node biopsy in melanoma: a short update. *Dermatopathology* 2018;**5**:21–5. https://doi.org/10.1159/000484892

93. Botar-Jid CM, Cosgarea R, Bolboacă SD, Şenilă SC, Lenghel LM, Rogojan L, Dudea SM. Assessment of cutaneous melanoma by use of very high-frequency ultrasound and real-time elastography. *AJR Am J Roentgenol* 2016;**206**:699–704. https://doi.org/10.2214/AJR.15.15182

94. Albertini MR. The age of enlightenment in melanoma immunotherapy. *J Immunother Cancer* 2018;**6**:80. https://doi.org/10.1186/s40425-018-0397-8

95. Lugowska I, Teterycz P, Rutkowski P. Immunotherapy of melanoma. *Contemp Oncol* 2018;**22**:61–7. https://doi.org/10.5114/wo.2018.73889

96. Koppolu V, Rekha Vasigala VK. Checkpoint immunotherapy by nivolumab for treatment of metastatic melanoma. *J Cancer Res Ther* 2018;**14**:1167–75. https://doi.org/10.4103/jcrt.JCRT_1290_16

97. Usher-Smith JA, Emery J, Kassianos AP, Walter FM. Risk prediction models for melanoma: a systematic review. *Cancer Epidemiol Biomarkers Prev* 2014;**23**:1450–63. https://doi.org/10.1158/1055-9965.EPI-14-0295

98. Vuong K, McGeechan K, Armstrong BK, Cust AE. Risk prediction models for incident primary cutaneous melanoma: a systematic review. *JAMA Dermatol* 2014;**150**:434–44. https://doi.org/10.1001/jamadermatol.2013.8890

99. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLOS ONE* 2012;**7**:e32844. https://doi.org/10.1371/journal.pone.0032844

100. Thomson DR, Rughani MG, Kuo R, Cassell OCS. Sentinel node biopsy status is strongly predictive of survival in cutaneous melanoma: extended follow-up of Oxford patients from 1998 to 2014. *J Plast Reconstr Aesthet Surg* 2017;**70**:1397–403. https://doi.org/10.1016/j.bjps.2017.05.025

101. Steyerberg EW. *Clinical Prediction Models*. New York, NY: Springer Publishing Company; 2009.

102. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLOS Med* 2014;**11**:e1001744. https://doi.org/10.1371/journal.pmed.1001744

103. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies (Prediction model Risk Of Bias ASsessment Tool). *Ann Intern Med* 2019;**170**:51–8. https://doi.org/10.7326/M18-1376

104. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, *et al.* Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;**350**:h870. https://doi.org/10.1136/bmj.h870

105. Baade PD, Royston P, Youl PH, Weinstock MA, Geller A, Aitken JF. Prognostic survival model for people diagnosed with invasive cutaneous melanoma. *BMC Cancer* 2015;**15**:27. https://doi.org/10.1186/s12885-015-1024-4

106. Balch CM, Soong SJ, Gershenwald JE, Thompson JF, Reintgen DS, Cascinelli N, *et al.* Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system. *J Clin Oncol* 2001;**19**:3622–34. https://doi.org/10.1200/JCO.2001.19.16.3622

107. Cochran AJ, Elashoff D, Morton DL, Elashoff R. Individualized prognosis for melanoma patients. *Hum Pathol* 2000;**31**:327–31. https://doi.org/10.1016/S0046-8177(00)80246-4

108. Gimotty PA, Elder DE, Fraker DL, Botbyl J, Sellers K, Elenitsas R, *et al.* Identification of high-risk patients among those diagnosed with thin cutaneous melanomas. *J Clin Oncol* 2007;**25**:1129–34. https://doi.org/10.1200/JCO.2006.08.1463

109. Gimotty PA, Guerry D, Ming ME, Elenitsas R, Xu X, Czerniecki B, *et al.* Thin primary cutaneous malignant melanoma: a prognostic tree for 10-year metastasis is more accurate than American Joint Committee on Cancer staging. *J Clin Oncol* 2004;**22**:3668–76. https://doi.org/10.1200/JCO.2004.12.015

110. Maurichi A, Miceli R, Camerini T, Mariani L, Patuzzo R, Ruggeri R, *et al.* Prediction of survival in patients with thin melanoma: results from a multi-institution study. *J Clin Oncol* 2014;**32**:2479–85. https://doi.org/10.1200/JCO.2013.54.2340

111. Rosenbaum BE, Schafer CN, Han SW, Osman I, Zhong H, Brinster N. Computer-assisted measurement of primary tumor area is prognostic of recurrence-free survival in stage IB melanoma patients. *Mod Pathol* 2017;**30**:1402–10. https://doi.org/10.1038/modpathol.2017.64

112. Saldanha G, Yarrow J, Pancholi J, Flatman K, Teo KW, Elsheik S, *et al.* Breslow density is a novel prognostic feature that adds value to melanoma staging. *Am J Surg Pathol* 2018;**42**:715–25. https://doi.org/10.1097/PAS.0000000000001034

113. Soong SJ, Ding S, Coit D, Balch CM, Gershenwald JE, Thompson JF, Gimotty P, AJCC Melanoma Task Force. Predicting survival outcome of localized melanoma: an electronic prediction tool based on the AJCC Melanoma Database. *Ann Surg Oncol* 2010;**17**:2006–14. https://doi.org/10.1245/s10434-010-1050-z

114. Tsai CA, Chen DT, Chen JJ, Balch CM, Thompson JF, Soong SJ. An integrated tree-based classification approach to prognostic grouping with application to localized melanoma patients. *J Biopharm Stat* 2007;**17**:445–60. https://doi.org/10.1080/10543400701199585

115. Vollmer RT, Seigler HF. Using a continuous transformation of the Breslow thickness for prognosis in cutaneous melanoma. *Am J Clin Pathol* 2001;**115**:205–12. https://doi.org/10.1309/WAVR-560R-NU5E-4Q96

116. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* 2005;**92**:965–70. https://doi.org/10.1093/biomet/92.4.965

117. Stadelmann WK, Rapaport DP, Soong SJ, Reintgen DS, Buzaid AC, Balch CM. Prognostic factors that influence melanoma outcome. In Balch CM, Sober AJ, Houghton AN, Soong SJ, editors. *Cutaneous Melanoma*. 3rd edn. St Louis, MO: Quality Medical Publishing; 1998. pp. 11–35.

118. Burnham K, Anderson D. *Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach*. 2nd edn. New York, NY: Springer-Verlag; 2002.

119. Deeny SR, Steventon A. Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data. *BMJ Qual Saf* 2015;**24**:505–15. https://doi.org/10.1136/bmjqs-2015-004278

120. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, *et al.* PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;**170**:W1–W33. https://doi.org/10.7326/M18-1377

121. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. *J Thorac Dis* 2019;**11**:S574. https://doi.org/10.21037/jtd.2019.01.25

122. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, Collins GS. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016;**353**:i3140. https://doi.org/10.1136/bmj.i3140

123. Debray TP, Koffijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KG. Meta-analysis and aggregation of multiple published prediction models. *Stat Med* 2014;**33**:2341–62. https://doi.org/10.1002/sim.6080

124. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;**21**:128–38. https://doi.org/10.1097/EDE.0b013e3181c30fb2

125. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004;**23**:723–48. https://doi.org/10.1002/sim.1621

126. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013;**13**:33. https://doi.org/10.1186/1471-2288-13-33

127. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using LOESS smoothers. *Stat Med* 2014;**33**:517–35. https://doi.org/10.1002/sim.5941

128. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;**54**:774–81. https://doi.org/10.1016/S0895-4356(01)00341-9

129. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;**68**:279–89. https://doi.org/10.1016/j.jclinepi.2014.06.018

130. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *J Thromb Haemost* 2013;**11**(Suppl. 1):129–41. https://doi.org/10.1111/jth.12262

131. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;**338**:b604. https://doi.org/10.1136/bmj.b604

132. Assel M, Sjoberg DD, Vickers AJ. The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. *Diagn Progn Res* 2017;**1**:19. https://doi.org/10.1186/s41512-017-0020-3

133. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;**338**:b605. https://doi.org/10.1136/bmj.b605

134. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;**338**:b606. https://doi.org/10.1136/bmj.b606

135. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;**13**:1. https://doi.org/10.1186/s12916-014-0241-z

136. Ankeny JS, Labadie B, Luke J, Hsueh E, Messina J, Zager JS. Review of diagnostic, prognostic, and predictive biomarkers in melanoma. *Clin Exp Metastasis* 2018;**35**:487–93. https://doi.org/10.1007/s10585-018-9892-z

137. Hyams DM, Cook RW, Buzaid AC. Identification of risk in cutaneous melanoma patients: Prognostic and predictive markers. *J Surg Oncol* 2019;**119**:175–86. https://doi.org/10.1002/jso.25319

138. Sartore L, Papanikolaou GE, Biancari F, Mazzoleni F. Prognostic factors of cutaneous melanoma in relation to metastasis at the sentinel lymph node: a case-controlled study. *Int J Surg* 2008;**6**:205–9. https://doi.org/10.1016/j.ijsu.2008.03.003

139. Kettlewell S, Moyes C, Bray C, Soutar D, MacKay A, Byrne D, *et al.* Value of sentinel node status as a prognostic factor in melanoma: prospective observational study. *BMJ* 2006;**332**:1423. https://doi.org/10.1136/bmj.38849.680509.AE

140. Catalano O, Siani A. Cutaneous melanoma: role of ultrasound in the assessment of locoregional spread. *Curr Probl Diagn Radiol* 2010;**39**:30–6. https://doi.org/10.1067/j.cpradiol.2009.04.001

141. Voit CA, van Akkooi AC, Schäfer-Hesterberg G, Schoengen A, Schmitz PI, Sterry W, *et al.* Rotterdam criteria for sentinel node (SN) tumor burden and the accuracy of ultrasound (US)-guided fine-needle aspiration cytology (FNAC): can US-guided FNAC replace SN staging in patients with melanoma? *J Clin Oncol* 2009;**27**:4994–5000. https://doi.org/10.1200/JCO.2008.19.0033

142. Voit C, Van Akkooi AC, Schäfer-Hesterberg G, Schoengen A, Kowalczyk K, Roewert JC, *et al.* Ultrasound morphology criteria predict metastatic disease of the sentinel nodes in patients with melanoma. *J Clin Oncol* 2010;**28**:847–52. https://doi.org/10.1200/JCO.2009.25.7428

143. Bossi MC, Sanvito S, Lovati E, De Fiori E, Testori A, Bellomi M. [Role of high resolution color-Doppler US of the sentinel node in patients with stage I melanoma.] *Radiol Med* 2001;**102**:357–62.

144. Hocevar M, Bracko M, Pogacnik A, Vidergar-Kralj B, Besic N, Zgajnar J, Music MM. The role of preoperative ultrasonography in reducing the number of sentinel lymph node procedures in melanoma. *Melanoma Res* 2004;**14**:533–6. https://doi.org/10.1097/00008390-200412000-00015

145. van Rijk MC, Teertstra HJ, Peterse JL, Nieweg OE, Olmos RA, Hoefnagel CA, Kroon BB. Ultrasonography and fine-needle aspiration cytology in the preoperative evaluation of melanoma patients eligible for sentinel node biopsy. *Ann Surg Oncol* 2006;**13**:1511–16. https://doi.org/10.1245/s10434-006-9106-9

146. Ulrich J, van Akkooi AJ, Eggermont AM, Voit C. New developments in melanoma: utility of ultrasound imaging (initial staging, follow-up and pre-SLNB). *Expert Rev Anticancer Ther* 2011;**11**:1693–701. https://doi.org/10.1586/era.11.115

147. Xing Y, Bronstein Y, Ross MI, Askew RL, Lee JE, Gershenwald JE, *et al.* Contemporary diagnostic imaging modalities for the staging and surveillance of melanoma patients: a meta-analysis. *J Natl Cancer Inst* 2011;**103**:129–42. https://doi.org/10.1093/jnci/djq455

148. Hall BJ, Schmidt RL, Sharma RR, Layfield LJ. Fine-needle aspiration cytology for the diagnosis of metastatic melanoma: systematic review and meta-analysis. *Am J Clin Pathol* 2013;**140**:635–42. https://doi.org/10.1309/AJCPWSDDHLLW40WI

149. Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, Glanville JM. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database Syst Rev* 2013;**9**:MR000022. https://doi.org/10.1002/14651858.MR000022.pub3

150. Vincent S, Greenley S, Beaven O. Clinical Evidence diagnosis: Developing a sensitive search strategy to retrieve diagnostic studies on deep vein thrombosis: a pragmatic approach. *Health Info Libr J* 2003;**20**:150–9. https://doi.org/10.1046/j.1365-2532.2003.00427.x

151. Simel DL, Feussner JR, DeLong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. *Med Decis Making* 1987;**7**:107–14. https://doi.org/10.1177/0272989X8700700208

152. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;**6**:e012799. https://doi.org/10.1136/bmjopen-2016-012799

153. 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. *Ann Intern Med* 2011;**155**:529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

154. 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. *J Clin Epidemiol* 2005;**58**:982–90. https://doi.org/10.1016/j.jclinepi.2005.02.022

155. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–58. https://doi.org/10.1002/sim.1186

156. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;**48**:193–204. https://doi.org/10.1016/j.jbi.2014.02.013

157. Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. *J Clin Diagn Res* 2016;**10**:YE01–YE06. https://doi.org/10.7860/JCDR/2016/18129.8744

158. Pánczél G, Liszkay G, Borbola K, Balatoni T, Hunyadi J. [The importance of fine needle aspiration cytology in the management of recurrent and metastatic melanoma.] *Orv Hetil* 2012;**153**:1419–23. https://doi.org/10.1556/OH.2012.29434

159. Blum A, Schlagenhauff B, Stroebel W, Breuninger H, Rassner G, Garbe C. Ultrasound examination of regional lymph nodes significantly improves early detection of locoregional metastases during the follow-up of patients with cutaneous melanoma: results of a prospective study of 1288 patients. *Cancer* 2000;**88**:2534–9. https://doi.org/10.1002/1097-0142(20000601)88:11<2534::AID-CNCR15>3.0.CO;2-2

160. Dalle S, Paulin C, Lapras V, Balme B, Ronger-Savle S, Thomas L. Fine-needle aspiration biopsy with ultrasound guidance in patients with malignant melanoma and palpable lymph nodes. *Br J Dermatol* 2006;**155**:552–6. https://doi.org/10.1111/j.1365-2133.2006.07361.x

161. Hayes AJ, Moskovic E, O'Meara K, Smith HG, Pope RJE, Larkin J, Thomas JM. Prospective cohort study of ultrasound surveillance of regional lymph nodes in patients with intermediate-risk cutaneous melanoma. *Br J Surg* 2019;**106**:729–34. https://doi.org/10.1002/bjs.11112

162. Rubaltelli L, Beltrame V, Tregnaghi A, Scagliori E, Frigo AC, Stramare R. Contrast-enhanced ultrasound for characterizing lymph nodes with focal cortical thickening in patients with cutaneous melanoma. *AJR Am J Roentgenol* 2011;**196**:W8–12. https://doi.org/10.2214/AJR.10.4711

163. Doubrovsky A, Scolyer RA, Murali R, McKenzie PR, Watson GF, Lee CS, *et al.* Diagnostic accuracy of fine needle biopsy for metastatic melanoma and its implications for patient management. *Ann Surg Oncol* 2008;**15**:323–32. https://doi.org/10.1245/s10434-006-9341-0

164. Krüger U, Kretschmer L, Thoms KM, Padeken M, Peter Bertsch H, Schön MP, Zutt M. Lymph node ultrasound during melanoma follow-up significantly improves metastasis detection compared with clinical examination alone: a study on 433 patients. *Melanoma Res* 2011;**21**:457–63. https://doi.org/10.1097/CMR.0b013e328348dad3

165. Solbiati L, Rizzatto G, Bellotti E. *High resolution sonography of cervical lymph nodes in head and neck cancer: criteria for differentiation of reactive versus malignant lymph nodes.* Proceedings of the 74th Meeting of the 462 Melanoma Research 2011, Vol. 21 No. 5 Radiologic Society of North America. Chicago, Radiologic Society of North America, 1988. abstr 113ff.

166. Vassallo P, Wernecke K, Roos N, Peters PE. Differentiation of benign from malignant superficial lymphadenopathy: the role of high resolution US. *Radiology* 1992;**183**:215–20. https://doi.org/10.1148/radiology.183.1.1549675

167. Garbe C, Schadendorf D, Stolz W, Volkenandt M, Reinhold U, Kortmann RD, *et al.* Short German guidelines: malignant melanoma. *J Dtsch Dermatol Ges* 2008;**6**(Suppl. 1):9–14. https://doi.org/10.1111/j.1610-0387.2008.06711.x

168. Knackstedt T, Knackstedt RW, Couto R, Gastman B. Malignant melanoma: diagnostic and management update. *Plast Reconstr Surg* 2018;**142**:202e–216e. https://doi.org/10.1097/PRS.0000000000004571

169. Bafounta ML, Beauchet A, Chagnon S, Saiag P. Ultrasonography or palpation for detection of melanoma nodal invasion: a meta-analysis. *Lancet Oncol* 2004;**5**:673–80. https://doi.org/10.1016/S1470-2045(04)01609-2

170. Dinnes J, Ferrante di Ruffano L, Takwoingi Y, Cheung ST, Nathan P, Matin RN, *et al.* Ultrasound, CT, MRI, or PET-CT for staging and re-staging of adults with cutaneous melanoma. *Cochrane Database Syst Rev* 2019;**7**:CD012806. https://doi.org/10.1002/14651858.CD012806.pub2

171. Mooney MM, Mettlin C, Michalek AM, Petrelli NJ, Kraybill WG. Life-long screening of patients with intermediate-thickness cutaneous melanoma for asymptomatic pulmonary recurrences: a cost-effectiveness analysis. *Cancer* 1997;**80**:1052–64.

172. Bassères N, Grob JJ, Richard MA, Thirion X, Zarour H, Noe C, *et al.* Cost-effectiveness of surveillance of stage I melanoma. A retrospective appraisal based on a 10-year experience in a dermatology department in France. *Dermatology* 1995;**191**:199–203. https://doi.org/10.1159/000246546

173. Hengge UR, Wallerand A, Stutzki A, Kockel N. Cost-effectiveness of reduced follow-up in malignant melanoma. *J Dtsch Dermatol Ges* 2007;**5**:898–907. https://doi.org/10.1111/j.1610-0387.2007.06454.x

174. Morton RL, Howard K, Thompson JF. The cost-effectiveness of sentinel node biopsy in patients with intermediate thickness primary cutaneous melanoma. *Ann Surg Oncol* 2009;**16**:929–40. https://doi.org/10.1245/s10434-008-0164-z

175. Serra-Arbeloa P, Rabines-Juárez ÁO, Álvarez-Ruiz MS, Guillén-Grima F. Sentinel node biopsy in patients with primary cutaneous melanoma of any thickness: a cost-effectiveness analysis. *Surg Oncol* 2016;**25**:205–11. https://doi.org/10.1016/j.suronc.2016.05.020

176. Freedberg KA, Geller AC, Miller DR, Lew RA, Koh HK. Screening for malignant melanoma: a cost-effectiveness analysis. *J Am Acad Dermatol* 1999;**41**:738–45. https://doi.org/10.1016/S0190-9622(99)70010-1

177. Losina E, Walensky RP, Geller A, Beddingfield FC, Wolf LL, Gilchrest BA, Freedberg KA. Visual screening for malignant melanoma: a cost-effectiveness analysis. *Arch Dermatol* 2007;**143**:21–8. https://doi.org/10.1001/archderm.143.1.21

178. Krug B, Crott R, Roch I, Lonneux M, Beguin C, Baurain JF, *et al.* Cost-effectiveness analysis of FDG PET-CT in the management of pulmonary metastases from malignant melanoma. *Acta Oncol* 2010;**49**:192–200. https://doi.org/10.3109/02841860903440254

179. Kansal AR, Shaul AJ, Stern S, Busam K, Doucet CA, Chalfin DB. Cost-effectiveness of a FISH assay for the diagnosis of melanoma in the USA. *Expert Rev Pharmacoecon Outcomes Res* 2013;**13**:371–80. https://doi.org/10.1586/erp.13.22

180. Gordon LG, Brynes J, Baade PD, Neale RE, Whiteman DC, Youl PH, *et al.* Cost-effectiveness analysis of a skin awareness intervention for early detection of skin cancer targeting men older than 50 years. *Value Health* 2017;**20**:593–601. https://doi.org/10.1016/j.jval.2016.12.017

181. Kurtz J, Beasley GM, Agnese D, Kendra K, Olencki TE, Terando A, Howard JH. Surveillance strategies in the follow-up of melanoma patients: too much or not enough? *J Surg Res* 2017;**214**:32–7. https://doi.org/10.1016/j.jss.2017.02.070

182. Wilson ECF, Usher-Smith JA, Emery J, Corrie P, Walter FM. A modeling study of the cost-effectiveness of a risk-stratified surveillance program for melanoma in the United Kingdom. *Value Health* 2018;**21**:658–68. https://doi.org/10.1016/j.jval.2017.11.009

183. Wilson EC, Emery JD, Kinmonth AL, Prevost AT, Morris HC, Humphrys E, *et al.* The cost-effectiveness of a novel SIAscopic diagnostic aid for the management of pigmented skin lesions in primary care: a decision-analytic model. *Value Health* 2013;**16**:356–66. https://doi.org/10.1016/j.jval.2012.12.008

184. Edwards SJ, Mavranezouli I, Osei-Assibey G, Marceniuk G, Wakefield V, Karner C. VivaScope® 1500 and 3000 systems for detecting and monitoring skin lesions: a systematic review and economic evaluation. *Health Technol Assess* 2016;**20**(58). https://doi.org/10.3310/hta20580

185. York Health Economics Consortium. *Micro-simulation*. 2016. URL: www.yhec.co.uk/glossary/micro-simulation/ (accessed 2 October 2019).

186. York Health Economics Consortium. *Markov Model*. 2016. URL: www.yhec.co.uk/glossary/markov-model/ (accessed 2 October 2019).

187. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. URL: www.nice.org.uk/process/pmg9/chapter/the-reference-case (accessed 2 October 2019).

188. Cancer Research UK. *Melanoma Skin Cancer Statistics*. 2019. URL: www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer#heading-One (accessed 23 May 2019).

189. Rueth NM, Xing Y, Chiang YJ, Cromwell KD, Ross MI, Lee JE, *et al.* Is surveillance imaging effective for detecting surgically treatable recurrences in patients with melanoma? A comparative analysis of stage-specific surveillance strategies. *Ann Surg* 2014;**259**:1215–22. https://doi.org/10.1097/SLA.0000000000000233

190. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al.* Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**(36). https://doi.org/10.3310/hta8360

191. Holterhues C, van de Poll-Franse LV, de Vries E, Neumann HA, Nijsten TE. Melanoma patients receive more follow-up care than current guideline recommendations: a study of 546 patients from the general Dutch population. *J Eur Acad Dermatol Venereol* 2012;**26**:1389–95. https://doi.org/10.1111/j.1468-3083.2011.04297.x

192. National Institute for Health and Care Excellence. *Improving Outcomes for People with Skin Tumours Including Melanoma*. 2006. URL: www.nice.org.uk/guidance/csg8 (accessed 2 October 2019).

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

194. Ahnlide I, Bjellerup M, Nilsson F, Nielsen K. Validity of ABCD rule of dermoscopy in clinical practice. *Acta Derm Venereol* 2016;**96**:367–72. https://doi.org/10.2340/00015555-2239

195. Bauer P, Cristofolini P, Boi S, Burroni M, Dell'Eva G, Micciolo R, Cristofolini M. Digital epiluminescence microscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. A statistical comparison between visual and computer inspection. *Melanoma Res* 2000;**10**:345–9. https://doi.org/10.1097/00008390-200008000-00005

196. 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 Res Technol* 2002;**8**:255–9. https://doi.org/10.1034/j.1600-0846.2002.00335.x

197. Carli P, De Giorgi V, Donati E, Pestelli E, Giannotti B. Epiluminescence microscopy reduces the risk of removing clinically atypical, but histologically common, melanocytic lesions. *G Ital Dermatol Venereol* 1994;**129**:599–605.

198. 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 Res* 2009;**19**:180–4. https://doi.org/10.1097/CMR.0b013e32832a1e41

199. Gokdemir A, Guler OM, Bek Y, Aydin F, Senturk N, Canturk T. Dermoscopic and histopathological correlation in melanocytic and non-melanocytic lesions. *Turkiye Klinikleri Dermatoloji* 2011;**21**:7–16.

200. 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. *J Am Acad Dermatol* 2010;**62**:785–93. https://doi.org/10.1016/j.jaad.2009.08.049

201. Morales-Callaghan AM, Castrodeza-Sanz J, Martínez-García G, Peral-Martínez I, Miranda-Romero A. [Correlation between clinical, dermatoscopic, and histopathologic variables in atypical melanocytic nevi.] *Actas Dermosifiliogr* 2008;**99**:380–9. https://doi.org/10.1016/S0001-7310(08)74697-0

202. 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. *J Am Acad Dermatol* 1994;**30**:551–9. https://doi.org/10.1016/S0190-9622(94)70061-3

203. Soyer HP, Smolle J, Leitinger G, Rieger E, Kerl H. Diagnostic reliability of dermoscopic criteria for detecting malignant melanoma. *Dermatology* 1995;**190**:25–30. https://doi.org/10.1159/000246629

204. Guitera P, Pellacani G, Longo C, Seidenari S, Avramidis M, Menzies SW. In vivo reflectance confocal microscopy enhances secondary evaluation of melanocytic lesions. *J Invest Dermatol* 2009;**129**:131–8. https://doi.org/10.1038/jid.2008.193

205. Bono A, Bartoli C, Baldi M, Tomatis S, Bifulco C, Santinami M. Clinical and dermatoscopic diagnosis of small pigmented skin lesions. *Eur J Dermatol* 2002;**12**:573–6.

206. 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**:362–6. https://doi.org/10.1159/000066436

207. Ferris LK, Harkes JA, Gilbert B, Winger DG, Golubets K, Akilov O, Satyanarayanan M. Computer-aided classification of melanocytic lesions using dermoscopic images. *J Am Acad Dermatol* 2015;**73**:769–76. https://doi.org/10.1016/j.jaad.2015.07.028

208. Elmore JG, Barnhill RL, Elder DE, Longton GM, Pepe MS, Reisch LM, *et al.* Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. *BMJ* 2017;**357**:j2813. https://doi.org/10.1136/bmj.j2813

209. Leiter U, Buettner PG, Eigentler TK, Bröcker EB, Voit C, Gollnick H, *et al.* Hazard rates for recurrent and secondary cutaneous melanoma: an analysis of 33,384 patients in the German Central Malignant Melanoma Registry. *J Am Acad Dermatol* 2012;**66**:37–45. https://doi.org/10.1016/j.jaad.2010.09.772

210. Turner RM, Bell KJ, Morton RL, Hayen A, Francken AB, Howard K, *et al.* Optimizing the frequency of follow-up visits for patients treated for localized primary cutaneous melanoma. *J Clin Oncol* 2011;**29**:4641–6. https://doi.org/10.1200/JCO.2010.34.2956

211. Office for National Statistics. *National Life Tables: UK*. 2019. URL: www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables (accessed 27 August 2019).

212. Department of Health and Social Care. *Reference Costs 2017/18: Highlights, Analysis and Introduction to the Data*. 2018. URL: https://improvement.nhs.uk/documents/1972/1_-_Reference_costs_201718.pdf (accessed 27 August 2019).

213. National Institute for Health and Care Excellence. *Trametinib in Combination with Dabrafenib for Treating Unresectable or Metastatic Melanoma*. NICE technology appraisal guidance TA396. 2016. URL: www.nice.org.uk/guidance/ta396/ (accessed 8 August 2019).

214. National Institute for Health and Care Excellence. *Nivolumab in Combination with Ipilimumab for Treating Advanced Melanoma*. NICE technology appraisal guidance TA400. 2016. URL: www.nice.org.uk/guidance/ta400/ (accessed 3 October 2019).

215. Joint Formulary Committee. *British National Formulary* (online). London: BMJ Group and Pharmaceutical Press. URL: www.medicinescomplete.com (accessed 3 October 2019).

216. Tran AD, Fogarty G, Nowak AK, Espinoza D, Rowbotham N, Stockler MR, Morton RL. A systematic review and meta-analysis of utility estimates in melanoma. *Br J Dermatol* 2018;**178**:384–93. https://doi.org/10.1111/bjd.16098

217. Brazier J, Ara R, Azzabi I, Busschbach J, Chevrou-Séverac H, Crawford B, *et al*. Identification, review, and use of health state utilities in cost-effectiveness models: an ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health* 2019;**22**:267–75. https://doi.org/10.1016/j.jval.2019.01.004

218. Ascierto PA, Kirkwood JM, Grob JJ, Simeone E, Grimaldi AM, Maio M, *et al*. The role of BRAF V600 mutation in melanoma. *J Transl Med* 2012;**10**:85. https://doi.org/10.1186/1479-5876-10-85

219. National institute for Health and Care Excellence. *Nivolumab for Treating Advanced (Unresectable or Metastatic) Melanoma*. 2016. URL: www.nice.org.uk/guidance/ta384/ (accessed 3 October 2019).

220. Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, Parkin D. The influence of cost-effectiveness and other factors on NICE decisions. *Health Econ* 2015;**24**:1256–71. https://doi.org/10.1002/hec.3086

221. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves – facts, fallacies and frequently asked questions. *Health Econ* 2004;**13**:405–15. https://doi.org/10.1002/hec.903

222. Garbe C. A rational approach to the follow-up of melanoma patients. *Recent Results Cancer Res* 2002;**160**:205–15. https://doi.org/10.1007/978-3-642-59410-6_24

223. Shumate CR, Urist MM, Maddox WA. Melanoma recurrence surveillance. Patient or physician based? *Ann Surg* 1995;**221**:566–9. https://doi.org/10.1097/00000658-199505000-00014

224. Francken AB, Shaw HM, Accortt NA, Soong SJ, Hoekstra HJ, Thompson JF. Detection of first relapse in cutaneous melanoma patients: implications for the formulation of evidence-based follow-up guidelines. *Ann Surg Oncol* 2007;**14**:1924–33. https://doi.org/10.1245/s10434-007-9347-2

225. Coyle D, Oakley J. Estimating the expected value of partial perfect information: a review of methods. *Eur J Health Econ* 2008;**9**:251–9. https://doi.org/10.1007/s10198-007-0069-y

226. National Cancer Registration and Analysis Service. *Survival by Stage*. 2019. URL: www.ncin.org.uk/publications/survival_by_stage (accessed 25 September 2019).

227. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Med Decis Making* 2014;**34**:311–26. https://doi.org/10.1177/0272989X13505910

228. McCabe C, Dixon S. Testing the validity of cost-effectiveness models. *PharmacoEconomics* 2000;**17**:501–13. https://doi.org/10.2165/00019053-200017050-00007

229. Drummond M. *Methods for the Economic Evaluation of Health Care Programmes*. 4th edn. Oxford: Oxford University Press; 2015.

230. O'Mahony JF, Naber SK, Normand C, Sharp L, O'Leary JJ, de Kok IM. Beware of kinked frontiers: a systematic review of the choice of comparator strategies in cost-effectiveness analyses of human papillomavirus testing in cervical screening. *Value Health* 2015;**18**:1138–51. https://doi.org/10.1016/j.jval.2015.09.2939

231. von Schuckmann LA, Hughes MCB, Ghiasvand R, Malt M, van der Pols JC, Beesley VL, *et al.* Risk of melanoma recurrence after diagnosis of a high-risk primary tumor. *JAMA Dermatol* 2019;**155**:688–93. https://doi.org/10.1001/jamadermatol.2019.0440

232. Morton RL. Essential inputs for studies of cost-effectiveness analysis in melanoma. *Br J Dermatol* 2014;**171**:1294–5. https://doi.org/10.1111/bjd.13455

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

234. Schünemann H, Brożek J, Guyatt G, Oxman A, The GRADE Working Group. *GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations*. 2013. URL: https://gdt.gradepro.org/app/handbook/handbook.html (accessed 25 February 2019).

235. Agbai ON, Buster K, Sanchez M, Hernandez C, Kundu RV, Chiu M, *et al.* Skin cancer and photoprotection in people of color: a review and recommendations for physicians and the public. *J Am Acad Dermatol* 2014;**70**:748–62. https://doi.org/10.1016/j.jaad.2013.11.038

236. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655. https://doi.org/10.1136/bmj.a1655

237. 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. *Arch Dermatol* 2010;**146**:257–64. https://doi.org/10.1001/archdermatol.2009.370

238. Tufts Medical Center. *CEA Registry*. 2018. URL: www.cevr.tuftsmedicalcenter.org/databases/cea-registry (accessed 15 August 2019).

239. Health Economics Research Centre. *HERC Database of Mapping Studies*. 2019. URL: www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies (accessed 15 August 2019).

240. National Institute for Health Research. *PROSPERO: International Prospective Register of Systematic Reviews*. 2019. URL: www.crd.york.ac.uk/PROSPERO (accessed 15 August 2019).

241. Canadian Agency for Drugs and Technologies in Health. *Strings Attached: CADTH Database Search Filters*. 2016. URL: www.cadth.ca/resources/finding-evidence/strings-attached-cadths-database-search-filters#health (accessed 29th August 2019).

Appendix 1 The MEDLINE strategy for the systematic review of surveillance strategies

MEDLINE (via Ovid)

Date range searched: 1946 to April week 4 2018.

Date searched: 4 April 2018.

#	Searches	Results (*n*)
1	Melanoma/or Hutchinson's Melanotic Freckle/or Melanoma, Amelanotic/	77,620
2	melanoma*.ti,ab,kf.	93,966
3	((skin or freckle* or lentigo* or lentigin*) adj3 (melanotic or tumor* or tumour* or cancer* or maligna*)).ti,ab,kf.	29,969
4	(cutaneous adj3 (tumor* or tumour* or cancer*)).ti,ab,kf.	3679
5	dubreuilh.ti,ab,kf.	41
6	or/1-5	130,861
7	Public Health Surveillance/	1907
8	Population Surveillance/	54,425
9	Mass Screening/	93,615
10	Aftercare/	7802
11	exp Genetic Testing/	40,228
12	or/7-11	194,425
13	6 and 12	1201
14	Melanoma/pc [Prevention & Control]	1697
15	Hutchinson's Melanotic Freckle/pc [Prevention & Control]	3
16	Melanoma, Amelanotic/pc [Prevention & Control]	3
17	((melanoma* or melanotic or skin cancer*) adj6 (monitor* or surveill* or follow up or followup or screen* or posttreatment or post treatment or after care or aftercare or check up or checkup or examin*)).ti,ab.	5206
18	or/14-17	6646
19	13 or 18	7254
20	limit 19 to (humans and yr = '2011 -Current')	2182
21	(comment or editorial or letter or news).pt.	1,652,061
22	20 not 21	2050

Search update carried out on 2 July 2019, added to the end of the search strategy described above.

MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations (via Ovid)

Date range searched: 1946 to 1 July 2019.

Date searched: 2 July 2019.

#	Searches	Results (*n*)
22	20 not 21	2409
23	(201803* or 201804* or 201805* or 201806* or 201807* or 201808* or 201809* or 20181* or 2019*).ed.	1,355,420
24	2019*.dp.	601,134
25	23 or 24	1,818,467
26	22 and 25	434

Appendix 2 Systematic review of surveillance strategies: excluded studies

Study	Reason(s) for exclusion
Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, *et al*. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;**391**:1023–75	Insufficient stage I participants
Auckland R, Wassell P, Hall S, Nicolson MC, Murchie P. Exploring patterns of recurrent melanoma in Northeast Scotland to inform the introduction a digital self-examination intervention. *BMC Dermatol* 2014;**14**:4	Insufficient stage I participants
Balamurugan A, Rees JR, Kosary C, Rim SH, Li J, Stewart SL. Subsequent primary cancers among men and women with in situ and invasive melanoma of the skin. *J Am Acad Dermatol* 2011;**65**(Suppl. 5):69–77	Stage not reported
Chen J, Xu Y, Zhou Y, Wang Y, Zhu H, Shi Y. Prognostic role of sentinel lymph node biopsy for patients with cutaneous melanoma: a retrospective study of surveillance, epidemiology, and end-result population-based data. *Oncotarget* 2016;**7**:45671–45677	Prognostic study
Chen T, Fallah M, Försti A, Kharazmi E, Sundquist K, Hemminki K. Risk of next melanoma in patients with familial and sporadic melanoma by number of previous melanomas. *JAMA Dermatol* 2015;**151**:607–15	Stage not reported
Czajkowska Z, Hall NC, Sewitch M, Wang B, Körner A. The role of patient education and physician support in self-efficacy for skin self-examination among patients with melanoma. *Patient Educ Couns* 2017;**100**:1505–10	Less than 80% of participants at stage I
Danielsen M, Højgaard L, Kjær A, Fischer BM. Positron emission tomography in the follow-up of cutaneous malignant melanoma patients: a systematic review. *Am J Nucl Med Mol Imaging* 2013;**4**:17–28	Systematic review of melanoma stages II–III
Dika E, Chessa MA, Veronesi G, Ravaioli GM, Fanti PA, Ribero S, *et al*. A single institute experience on melanoma prognosis: a long term follow-up. *G Ital Dermatol Venereol* 2018;**12**:12	No surveillance and insufficient stage I participants
de Vries M, Speijers MJ, Bastiaannet E, Plukker JT, Brouwers AH, van Ginkel RJ, *et al*. Long-term follow-up reveals that ulceration and sentinel lymph node status are the strongest predictors for survival in patients with primary cutaneous melanoma. *Eur J Surg Oncol* 2011;**37**:681–7	Prognostic study
Glenn BA, Chen KL, Chang LC, Lin T, Bastani R. Skin examination practices among melanoma survivors and their children. *J Cancer Educ* 2017;**32**:335–43	Stage not reported
Jones MS, Torisu-Itakura H, Flaherty DC, Schoellhammer HF, Lee J, Sim MS, Faries MB. Second primary melanoma: risk factors, histopathologic features, survival, and implications for follow-up. *Am Surg* 2016;**82**:1009–13	Less than 80% of participants at stage I
Lee HJ, Jin H, You HS, Shim WH, Kim JM, Kim GW, *et al*. Various dermatoses what the patients with cutaneous melanoma had anxiety for the recurrence during postoperative surveillance. *Ann Dermatol* 2017;**29**:433–7	Insufficient stage I participants
Memari N, Hayen A, Bell KJ, Rychetnik L, Morton RL, McCaffery K, *et al*. How often do patients with localized melanoma attend follow-up at a specialist center? *Ann Surg Oncol* 2015;**22**:S1164–71	Less than 80% of participants at stage I
Nahar VK, Allison Ford M, Brodell RT, Boyas JF, Jacks SK, Biviji-Sharma R, *et al*. Skin cancer prevention practices among malignant melanoma survivors: a systematic review. *J Cancer Res Clin Oncol* 2016;**142**:1273–83	Systematic review of prevention strategies of survivors
Pomerantz H, Huang D, Weinstock MA. Risk of subsequent melanoma after melanoma in situ and invasive melanoma: a population-based study from 1973 to 2011. *J Am Acad Dermatol* 2015;**72**:794–800	All stages of invasive melanomas

Study	Reason(s) for exclusion
Ribero S, Podlipnik S, Osella-Abate S, Sportoletti-Baduel E, Manubens E, Barreiro A, *et al*. Ultrasound-based follow-up does not increase survival in early-stage melanoma patients: a comparative cohort study. *Eur J Cancer* 2017;**85**:59–66	Less than 80% of participants at stage I
Rodríguez VM, Berwick M, Hay JL. Communication about melanoma and risk reduction after melanoma diagnosis. *Psycho-Oncology* 2017;**26**:2142–8	Prevention strategies
Rohren EM. PET/computed tomography and patient outcomes in melanoma. *PET Clin* 2015;**10**:243–54	Diagnostic study
Saiag P, Aegerter P, Vitoux D, Lebbé C, Wolkenstein P, Dupin N, *et al.* Prognostic Value of 25-hydroxyvitamin D3 levels at diagnosis and during follow-up in melanoma patients. *J Natl Cancer Inst* 2015;**107**:djv264	Prognostic study
Salerni G, Lovatto L, Carrera C, Puig S, Malvehy J. Melanomas detected in a follow-up program compared with melanomas referred to a melanoma unit. *Arch Dermatol* 2011;**147**:549–55	Less than 80% of participants at stage I
Sanlorenzo M, Ribero S, Osella-Abate S, Zugna D, Marenco F, Macripo G, *et al.* Prognostic differences across sexes in melanoma patients: what has changed from the past? *Melanoma Res* 2014;**24**:568–76	Prognostic study
Sjoestroem C, Khosravi S, Cheng Y, Safaee Ardekani G, Martinka M, Li G. DLC1 expression is reduced in human cutaneous melanoma and correlates with patient survival. *Mod Pathol* 2014;**27**:1203–11	Prognostic study
Solivetti FM, Elia F, Graceffa D, Di Carlo A. Ultrasound morphology of inguinal lymph nodes may not herald an associated pathology. *J Exp Clin Cancer Res* 2012;**31**:88	Diagnostic study
Turner RM, Bell KJ, Morton RL, Hayen A, Francken AB, Howard K, *et al.* Optimizing the frequency of follow-up visits for patients treated for localized primary cutaneous melanoma. *J Clin Oncol* 2011;**29**:4641–6	Less than 80% of participants at stage I
Vensby PH, Schmidt G, Kjær A, Fischer BM. The value of FDG PET/CT for follow-up of patients with melanoma: a retrospective analysis. *Am J Nucl Med Mol Imaging* 2017;**7**:255–62	Less than 80% of participants at stage I
Wolf A, Kvint S, Chachoua A, Pavlick A, Wilson M, Donahue B, *et al.* Toward the complete control of brain metastases using surveillance screening and stereotactic radiosurgery. *J Neurosurg* 2018;**128**:23–31	Advanced stage of disease
Wu CE, Hsieh CH, Chang CJ, Yeh JT, Kuo TT, Yang CH, *et al.* Prognostic factors for Taiwanese patients with cutaneous melanoma undergoing sentinel lymph node biopsy. *J Formos Med Assoc* 2015;**114**:415–21	Prognostic study
Xing Y, Cromwell KD, Cormier JN. Review of diagnostic imaging modalities for the surveillance of melanoma patients. *Dermatol Res Pract* 2012;**2012**:941921	Diagnostic study
Youl PH, Soyer HP, Baade PD, Marshall AL, Finch L, Janda M. Can skin cancer prevention and early detection be improved via mobile phone text messaging? A randomised, attention control trial. *Prev Med* 2015;**71**:50–6	Stage not reported
Zhang G, Cheng Y, Chen G, Tang Y, Ardekani G, Rotte A, *et al.* Loss of tumor suppressors KAI1 and p27 identifies a unique subgroup of primary melanoma patients with poor prognosis. *Oncotarget* 2015;**6**:23026–35	Prognostic study
Zörnig I, Halama N, Lorenzo Bermejo J, Ziegelmeier C, Dickes E, Migdoll A, *et al.* Prognostic significance of spontaneous antibody responses against tumor-associated antigens in malignant melanoma patients. *Int J Cancer* 2015;**136**:138–51	Prognostic study

Appendix 3 The MEDLINE strategy for the systematic review of risk prediction models

MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations (via Ovid)

Date range searched: 1946 to 15 July 2019.

Date of original search: September 2018.

Date of updated search: 16 July 2019.

#	Searches	Results (*n*)
1	(Validat* or Predict* or Rule* or (Predict* and (Outcome* or Risk* or Model*)) or ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* or Prognos*))).mp. or (Decision*.mp. and ((Model* or Clinical*).mp. or Logistic Models/)) or (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model*)).ti,ab. [mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4,600,871
2	Stratification.mp. or ROC Curve/or Discrimination.mp. or Discriminate.mp. or c-statistic.mp. or c statistic.mp. or Area under the curve.mp. or AUC.mp. or Calibration.mp. or Indices.mp. or Algorithm.mp. or Multivariable.mp. or Prognosis.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1,450,061
3	1 or 2	5,285,509
4	*melanoma/	65,466
5	*skin neoplasms/	99,710
6	melanoma.kw.	7937
7	(malignant adj3 melanoma*).ti,ab.	27,244
8	(tumo* adj5 (mole or melanoma)).ti,ab.	12,160
9	4 or 5 or 6 or 7 or 8	149,902
10	(((prognos* or melanoma) adj5 surviv*) or factor*).ti,ab.	3,091,009
11	(metas* or advance* or recur* or relaps* or invasive or second* or disseminat*).ti,ab.	3,355,846
12	(distant metastases or local recurren*).ti,ab.	45,561
13	10 or 11 or 12	5,901,577
14	exp comment/or exp letter/or exp editorial/	1,731,981
15	exp animals/not exp humans/	4,595,710
16	(animal or mouse).mp. or mice.ti,ab. [mp = title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2,072,446
17	exp review/	2,504,343
18	exp case reports/or case report*.ti,ab.	2,103,512
19	or/14-18	10,922,034
20	(3 and 9 and 13) not 19	13,880
21	(201808* or 201809* or 20181* or 2019*).ed. or 2019*.dp.	1,423,830
22	20 and 21	1023

Appendix 4 Systematic review of risk prediction models: excluded studies

Study	Reason(s) for exclusion
Arce PM, Camilon PR, Stokes WA, Nguyen SA, Lentsch EJ. Is sex an independent prognostic factor in cutaneous head and neck melanoma? *Laryngoscope* 2014;**124**:1363–7	One factor
Ariyan C, Brady MS, Gönen M, Busam K, Coit D. Positive nonsentinel node status predicts mortality in patients with cutaneous melanoma. *Ann Surg Oncol* 2009;**16**:186–90	One factor
Avilés-Izquierdo JA, Lázaro-Ochaita P. [Sentinel node biopsy as a prognostic factor in cutaneous melanoma.] *Actas Dermosifiliogr* 2009;**100**:486–92	One factor
Byrom L, Dasgupta P, Youlden D, Baade P, Green A, Khosrotehrani K. Melanoma prognosis differs according to sex: an Australian population study. *Australas J Dermatol* 2014;**55**:9–10	One factor
Cadili A, Dabbs K. Predictors of sentinel lymph node metastasis in melanoma. *Can J Surg* 2010;**53**:32–6	Stage not reported
Callender GG, Egger ME, Burton AL, Scoggins CR, Ross MI, Stromberg AJ, *et al.* Prognostic implications of anatomic location of primary cutaneous melanoma of 1 mm or thicker. *Am J Surg* 2011;**202**:659–64	No validation
Callender GG, Gershenwald JE, Egger ME, Scoggins CR, Martin RC, Schacherer CW, *et al.* A novel and accurate computer model of melanoma prognosis for patients staged by sentinel lymph node biopsy: comparison with the American Joint Committee on Cancer model. *J Am Coll Surg* 2012;**214**:608–17	Excludes stage I
Cascinelli N, Bombardieri E, Bufalino R, Camerini T, Carbone A, Clemente C, *et al.* Sentinel and nonsentinel node status in stage IB and II melanoma patients: two-step prognostic indicators of survival. *J Clin Oncol* 2006;**24**:4464–71	No validation
Chen G, Chen Y, Zhang Z, Martinka M, Li G. Reduced Tip60 expression as a predictive biomarker for advanced melanoma and patient outcome. *Cancer Res* 2011;**71**:2103	One factor
Chen N, Gong J, Chen X, Xu M, Huang Y, Wang L, *et al.* Cytokeratin expression in malignant melanoma: potential application of in-situ hybridization analysis of mRNA. *Melanoma Res* 2009;**19**:87–93	One factor
Cheng Y, Zhou Y. BRAF protein expression as a prognostic marker for thin melanomas. *J Invest Dermatol* 2014;**134**:S131	Stages mixed
Cook RW, Covington KR, Monzon FA. Continued evaluation of a 31-gene expression profile to predict metastasis in an expanded cohort of 782 cutaneous melanoma patients. *Pigment Cell Melanoma Res* 2017;**30**:e73	Stage I and II
Crocetti E, Mangone L, Lo Scocco G, Carli P. Prognostic variables and prognostic groups for malignant melanoma. The information from Cox and classification and regression trees analysis: an Italian population-based study. *Melanoma Res* 2006;**16**:429–33	Stages mixed
Cymerman RM, Wang K, Murzaku EC, Penn LA, Osman I, Shao Y, *et al.* De novo versus nevus-associated melanomas: Differences in associations with prognostic indicators and survival. *J Clin Oncol* 2015;**33**:9025	No validation
Emmett MS, Symonds KE, Rigby H, Cook MG, Price R, Metcalfe C, *et al.* Prediction of melanoma metastasis by the Shields index based on lymphatic vessel density. *BMC Cancer* 2010;**10**:208	Stage not reported
Fang S, Wang Y, Chun YS, Liu H, Ross MI, Gershenwald JE, *et al.* Association of common genetic polymorphisms with melanoma patient IL-12p40 blood levels, risk, and outcomes. *J Invest Dermatol* 2015;**135**:2266–72	Stage I and II
Fang S, Wang Y, Chun YS, Liu H, Ross MI, Gershenwald JE, *et al.* The relationship between blood IL-12p40 level and melanoma progression. *Int J Cancer* 2015;**136**:1874–80	Single prognostic factor
Fang S, Wang Y, Sui D, Liu H, Ross MI, Gershenwald JE, *et al.* C-reactive protein as a marker of melanoma progression. *J Clin Oncol* 2015;**33**:1389–96	No validation

Study	Reason(s) for exclusion
Findeisen P, Zapatka M, Peccerella T, Matzk H, Neumaier M, Schadendorf D, Ugurel S. Serum amyloid A as a prognostic marker in melanoma identified by proteomic profiling. *J Clin Oncol* 2009;**27**:2199–208	Stages mixed
Garnier JP, Letellier S, Cassinat B, Lebbé C, Kerob D, Baccard M, *et al.* Clinical value of combined determination of plasma L-DOPA/tyrosine ratio, S100B, MIA and LDH in melanoma. *Eur J Cancer* 2007;**43**:816–21	No validation
Gerami P, Jewell SS, Pouryazdanparast P, Wayne JD, Haghighat Z, Busam KJ, *et al.* Copy number gains in 11q13 and 8q24 [corrected] are highly linked to prognosis in cutaneous malignant melanoma. *J Mol Diagn* 2011;**13**:352–8	No validation
Gillgren P, Brattström G, Frisell J, Persson JO, Ringborg U, Hansson J. Effect of primary site on prognosis in patients with cutaneous malignant melanoma. A study using a new model to analyse anatomical locations. *Melanoma Res* 2005;**15**:125–32	No validation
Gimotty P, Guerry D, VanBelle P, Montone K, Guerra M, Hwang W, *et al.* Ki67 as a prognostic biomarker for patients with vertical growth phase (VGP) melanomas. *J Clin Oncol* 2009;**27**:9043	Stage I and II
Gomez GV, de Oliveira C, Rinck-Junior JA, de Moraes AM, Lourenço GJ, Lima CS. XPC (A2920C), XPF (T30028C), TP53 (Arg72Pro), and GSTP1 (Ile105Val) polymorphisms in prognosis of cutaneous melanoma. *Tumour Biol* 2016;**37**:3163–71	No validation
Governa M, Dorizzi RM, Gatti S, Tambuscio A, Minic J, Barisoni D. Is increased serum S-100 protein concentration a marker of metastasis in malignant melanoma? A four-year experience report. *Eur J Plast Surg* 2005;**28**:17–20	No validation
Henry L, Lavabre-Bertrand T, Douche T, Uttenweiler-Joseph S, Fabbro-Peray P, Monsarrat B, *et al.* Diagnostic value and prognostic significance of plasmatic proteasome level in patients with melanoma. *Exp Dermatol* 2010;**19**:1054–9	No validation
Henry L, Fabre C, Guiraud I, Bastide S, Fabbro-Peray P, Martinez J, *et al.* Clinical use of p-proteasome in discriminating metastatic melanoma patients: comparative study with LDH, MIA and S100B protein. *Int J Cancer* 2013;**133**:142–8	One factor
Hoon DS, Bostick P, Kuo C, Okamoto T, Wang HJ, Elashoff R, Morton DL. Molecular markers in blood as surrogate prognostic indicators of melanoma recurrence. *Cancer Res* 2000;**60**:2253–7	No validation
Hsueh EC, DeBloom JR, Lee J, Sussman JJ, Covington KR, Middlebrook B, *et al.* Interim analysis of survival in a prospective, multi-center registry cohort of cutaneous melanoma tested with a prognostic 31-gene expression profile test. *J Hematol Oncol* 2017;**10**:152	All stages in analysis by class
Karagiannis P, Villanova F, Josephs DH, Correa I, Van Hemelrijck M, Hobbs C, *et al.* IgG4: a new tool to predict the risk of disease progression in melanoma. *Cancer Immunol Res* 2016;**4**:A009	No validation
Kashani-Sabet M, Nosrati M, Miller JR, Sagebiel RW, Leong SPL, Lesniak A, *et al.* Prospective validation of molecular prognostic markers in cutaneous melanoma: a correlative analysis of E1690. *Clin Cancer Res* 2017;**23**:6888–92	Stages not reported
Kluger HM, Hoyt K, Bacchiocchi A, Mayer T, Kirsch J, Kluger Y, *et al.* Plasma markers for identifying patients with metastatic melanoma. *Clin Cancer Res* 2011;**17**:2417–25	Stage I and II
Lasithiotakis K, Leiter U, Meier F, Eigentler T, Metzler G, Moehrle M, *et al.* Age and sex are significant independent predictors of survival in primary cutaneous melanoma. *Cancer* 2008;**112**:1795–804	Not validated
Leiter U, Buettner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the german dermatological society. *J Clin Oncol* 2004;**22**:3660–7	No validation
Li N, Diao Z, Huang X, Niu Y, Liu T, Liu ZP, *et al.* Increased platelet distribution width predicts poor prognosis in melanoma patients. *Sci Rep* 2017;**7**:2970	Single prognostic factor
Liu J, Li R, Zhou X, Zhang J, Luo R. [Multivariate regression analysis of the biomarkers and clinical characteristics in the prognosis of malignant melanoma.] *Nan Fang Yi Ke Da Xue Xue Bao* 2012;**32**:847–53	Stages mixed

Study	Reason(s) for exclusion
Lyth J, Mikiver R, Nielsen K, Isaksson K, Ingvar C. Prognostic instrument for survival outcome in melanoma patients: based on data from the population-based Swedish Melanoma Register. *Eur J Cancer* 2016;**59**:171–8	No independent validation
Mitchell B, Leone D, Feller K, Menon S, Bondzie P, Yang S, *et al.* Protein expression of the chemokine receptor CXCR4 and its ligand CXCL12 in primary cutaneous melanoma – Biomarkers of potential utility? *Hum Pathol* 2014;**45**:2094–100	No validation
Mocellin S, Pasquali S, Rossi CR, Nitti D. Validation of the prognostic value of lymph node ratio in patients with cutaneous melanoma: a population-based study of 8,177 cases. *Surgery* 2011;**150**:83–90	All stages
Murali R, Shaw HM, Lai K, McCarthy SW, Quinn MJ, Stretch JR, *et al.* Prognostic factors in cutaneous desmoplastic melanoma. *Cancer* 2010;**116**:4130–8	No validation
Murtas D, Piras F, Minerba L, Ugalde J, Floris C, Maxia C, *et al.* Nuclear 8-hydroxy-2'-deoxyguanosine as survival biomarker in patients with cutaneous melanoma. *Oncol Rep* 2010;**23**:329–35	One factor
Naffouje R, Salti GI. The role of microphthalmia transcription factor (Mitf) in prediction of distant metastases in cutaneous melanoma. *J Clin Oncol* 2016;**34**:9564	No validation
Naffouje SA, Naffouje R, Chen J, Salti GI. Validation and enhancement of the clinicopathological melanoma nomogram via incorporation of a molecular marker in the primary tumor. *Anticancer Res* 2016;**36**:6603–10	Stages mixed
Nagore E, Heidenreich B, Garcia-Casado Z, Requena C, Kumar R. TERT promoter mutations in melanoma survival. *Pigment Cell Melanoma Res* 2017;**30**:124	No validation
Nagore Enguídanos E, Oliver Martínez V, Botella Estrada R, Insa Mollá A, Fortea Baixauli JM. [Prognostic factors of localized malignant melanoma: study of 639 patients.] *Med Clin* 2005;**124**:361–7	No validation
Nsengimana J, Laye J, Filia A, Walker C, Jewell R, Van den Oord JJ, *et al.* Independent replication of a melanoma subtype gene signature and evaluation of its prognostic value and biological correlates in a population cohort. *Oncotarget* 2015;**6**:11683–93	Stage not reported
Oliveira C, Rinck JA, Lourenço GJ, Moraes AM, Lima CSP. Polymorphisms in the apoptosis pathway and prognosis in cutaneous melanoma. *J Clin Oncol* 2014;**32**(Suppl. 1):9084	No validation
Ortega-Bernal D, Rangel-Escareno C, Arechaga-Ocampo E, Gonzalez-De La Rosa CH. Biomarkers for staging melanoma, a search at transcriptome level. *Cancer Res* 2018;**78**:2252	No validation
Ostmeier H, Fuchs B, Otto F, Mawick R, Lippold A, Krieg V, Suter L. Prognostic immunohistochemical markers of primary human melanomas. *Br J Dermatol* 2001;**145**:203–9	No validation
Ozao-Choy J, Nelson DW, Hiles J, Stern S, Yoon JL, Sim MS, Faries MB. The prognostic importance of scalp location in primary head and neck melanoma. *J Surg Oncol* 2017;**116**:337–43	No validation
Pacifico MD, Grover R, Richman PI, Daley FM, Buffa F, Wilson GD. CD44v3 levels in primary cutaneous melanoma are predictive of prognosis: assessment by the use of tissue microarray. *Int J Cancer* 2006;**118**:1460–4	No validation
Palmieri G, Ascierto PA, Perrone F, Satriano SM, Ottaiano A, Daponte A, *et al.* Prognostic value of circulating melanoma cells detected by reverse transcriptase-polymerase chain reaction. *J Clin Oncol* 2003;**21**:767–73	No validation
Passos Lima CS, Gomez GVB, Oliveira C, Lourenço GJ, Rinck JA, Moraes AM. XPC, XPF, TP53 and GSTP1 polymorphisms in prognosis of cutaneous melanoma patients. *J Clin Oncol* 2015;**33**:9038	No validation
Pizzichetta MA, Massi D, Mandalà M, Queirolo P, Stanganelli I, De Giorgi V, *et al.* Clinicopathological predictors of recurrence in nodular and superficial spreading cutaneous melanoma: a multivariate analysis of 214 cases. *J Transl Med* 2017;**15**:227	No validation
Ponti G, Pollio A, Cesinaro AM, Pellacani G, Magnoni C, Seidenari S. Value and prognostic significance of mitotic rate in a retrospective series of pT1 cutaneous malignant melanoma patients. *Cancer Epidemiol* 2012;**36**:303–5	No validation

Study	Reason(s) for exclusion
Poukka M, Bykachev A, Siiskonen H, Tyynelä-Korhonen K, Auvinen P, Pasonen-Seppänen S, Sironen R. Decreased expression of hyaluronan synthase 1 and 2 associates with poor prognosis in cutaneous melanoma. *BMC Cancer* 2016;**16**:313	No validation
Ramsden AJ, Grover R, Chana J, Tulley P, Sanders R, Wilson GD. A prospective analysis of c-myc oncoprotein levels as a prognostic marker in malignant melanoma. *J Plast Reconstr Aesthet Surg* 2007;**60**:626–30	One factor
Rangel J, Nosrati M, Torabian S, Shaikh L, Leong SP, Haqq C, *et al.* Osteopontin as a molecular prognostic marker for melanoma. *Cancer* 2008;**112**:144–50	Single prognostic factor
Rangel J, Torabian S, Shaikh L, Nosrati M, Baehner FL, Haqq C, *et al.* Prognostic significance of nuclear receptor coactivator-3 overexpression in primary cutaneous melanoma. *J Clin Oncol* 2006;**24**:4565–9	Single prognostic factor
Rashed H, Bamford M, Flatman K, Teo KWW, Saldanha G. Breslow density as an independent prognostic indicator in cutaneous malignant melanoma. *J Pathol* 2016;**240**:S47	No validation
Reschke M, Mihic-Probst D, van der Horst EH, Knyazev P, Wild PJ, Hutterer M, *et al.* HER3 is a determinant for poor prognosis in melanoma. *Clin Cancer Res* 2008;**14**:5188–97	One factor
Rex J, Paradelo C, Mangas C, Hilari JM, Fernandez-Figueras MT, Fraile M, *et al.* Single-institution experience in the management of patients with clinical stage I and II cutaneous melanoma: results of sentinel lymph node biopsy in 240 cases. *Dermatol Surg* 2005;**31**:1385–93	No validation
Gould Rothberg BE, Berger AJ, Molinaro AM, Subtil A, Krauthammer MO, Camp RL, *et al.* Melanoma prognostic model using tissue microarrays and genetic algorithms. *J Clin Oncol* 2009;**27**:5772–80	Stage II
Rotte A, Bhandaru M, Cheng Y, Sjoestroem C, Martinka M, Li G. Decreased expression of nuclear p300 is associated with disease progression and worse prognosis of melanoma patients. *PLOS ONE* 2013;**8**:e75405	Stage I and II
Rotte A, Martinka M, Li G. MMP2 expression is a prognostic marker for primary melanoma patients. *Cell Oncol* 2012;**35**:207–16	Single prognostic factor
Rowe C, Tang F, Hughes MC, Rodero M, Malt M, Lambie D, *et al.* Prognostic value of nomograms incorporating biomarkers vs. sentinel node status in patients with stage IB and II melanoma. *Australas J Dermatol* 2016;**57**:3	Stages mixed
Roxanis I, Chow J. Cellular cohesion as a prognostic factor in malignant melanoma: a retrospective study with up to 12 years follow-up. *Mod Pathol* 2010;**23**:502–10	One factor
Sabel MS, Liu Y, Griffith KA, He J, Xie X, Lubman DM. Clinical utility of serum autoantibodies detected by protein microarray in melanoma. *Int J Proteomics* 2011;**2011**:413742	No validation
Sanlorenzo M, Ribero S, Osella-Abate S, Zugna D, Marenco F, Macripo G, *et al.* Prognostic differences across sexes in melanoma patients: what has changed from the past? *Melanoma Res* 2014;**24**:568–76	No validation
Schäfer A, Emmert S, Kruppa J, Schubert S, Tzvetkov M, Mössner R, *et al.* No association of vitamin D metabolism-related polymorphisms and melanoma risk as well as melanoma prognosis: a case-control study. *Arch Dermatol Res* 2012;**304**:353–61	No results by stage
Schmidt H, Johansen JS, Sjoegren P, Christensen IJ, Sorensen BS, Fode K, *et al.* Serum YKL-40 predicts relapse-free and overall survival in patients with American Joint Committee on Cancer stage I and II melanoma. *J Clin Oncol* 2006;**24**:798–804	No validation
Shi Q, Liu H, Li C, Wang Y, Liu Z, Amos C, *et al.* Genetic variants in the Wnt pathway genes NFATC1 and PLCB1 predict melanoma survival. *J Invest Dermatol* 2016;**136**:S37	No validation
Shourkaei SMJ, Wani AA, Martinka M, Li G. Prognostic significance of nuclear Sox4 expression in cutaneous melanoma and its role in cell migration. *Cancer Res* 2010;**70**:2243	Single prognostic factor
Silva S, Cox A, Teare D, Bradford J, Brock I, Connley D, *et al.* Copy-number profiles from circulating cell-free DNA as a potential biomarker in melanoma. *Eur J Surg Oncol* 2018;**44**:S26–S7	No validation

Study	Reason(s) for exclusion
Sjoestroem C, Khosravi S, Cheng Y, Safaee Ardekani G, Martinka M, Li G. DLC1 expression is reduced in human cutaneous melanoma and correlates with patient survival. *Mod Pathol* 2014;**27**:1203–11	No validation
Stiegel E, Xiong D, Ya J, Funchain P, Isakov R, Gastman B, Vij A. Prognostic value of sentinel lymph node biopsy according to Breslow thickness for cutaneous melanoma. *J Am Acad Dermatol* 2018;**78**:942–8	No validation
Stokes WA, Lentsch EJ. Age is an independent poor prognostic factor in cutaneous head and neck melanoma. *Laryngoscope* 2014;**124**:462–5	One factor
Straume O, Sviland L, Akslen LA. Loss of nuclear p16 protein expression correlates with increased tumor cell proliferation (Ki-67) and poor prognosis in patients with vertical growth phase melanoma. *Clin Cancer Res* 2000;**6**:1845–53	No validation
Tchernev G, Orfanos CE. Downregulation of cell cycle modulators p21, p27, p53, Rb and proapoptotic Bcl-2-related proteins Bax and Bak in cutaneous melanoma is associated with worse patient prognosis: preliminary findings. *J Cutan Pathol* 2007;**34**:247–56	Stage II
Thies A, Mangold U, Moll I, Schumacher U. PAS-positive loops and networks as a prognostic indicator in cutaneous malignant melanoma. *J Pathol* 2001;**195**:537–42	No validation
Thomas NE, Busam KJ, From L, Kricker A, Armstrong BK, Anton-Culver H, *et al.* Tumor-infiltrating lymphocyte grade in primary melanomas is independently associated with melanoma-specific survival in the population-based genes, environment and melanoma study. *J Clin Oncol* 2013;**31**:4252–9	No validation
Thomas NE, Kricker A, Waxweiler WT, Dillon PM, Busman KJ, From L, *et al.* Comparison of clinicopathologic features and survival of histopathologically amelanotic and pigmented melanomas: a population-based study. *JAMA Dermatol* 2014;**150**:1306–314	No validation
Thompson JF, Soong SJ, Balch CM, Gershenwald JE, Ding S, Coit DG, *et al.* Prognostic significance of mitotic rate in localized primary cutaneous melanoma: an analysis of patients in the multi-institutional American Joint Committee on Cancer melanoma staging database. *J Clin Oncol* 2011;**29**:2199–205	No validation
Väisänen AH, Kallioinen M, Turpeenniemi-Hujanen T. Comparison of the prognostic value of matrix metalloproteinases 2 and 9 in cutaneous melanoma. *Hum Pathol* 2008;**39**:377–85	No validation
van Akkooi ACJ, Rutkowski P, Van Der Ploeg IM, Voit C, Robert C, Hoekstra HJ, *et al.* Excellent long-term survival of patients with minimal sentinel node tumor burden (< 0.1 mm) according to Rotterdam Criteria: a study of the EORTC melanoma group. *Eur J Cancer Supp* 2009;**7**:576–7	No validation
van der Ploeg AP, van Akkooi AC, Rutkowski P, Nowecki ZI, Michej W, Mitra A, *et al.* Prognosis in patients with sentinel node-positive melanoma is accurately defined by the combined Rotterdam tumor load and Dewar topography criteria. *J Clin Oncol* 2011;**29**:2206–14	No validation
Vergilis IJ, Szarek M, Ferrone S, Reynolds SR. Presence and prognostic significance of melanoma-associated antigens CYT-MAA and HMW-MAA in serum of patients with melanoma. *J Invest Dermatol* 2005;**125**:526–31	Excludes stage I
Vollmer RT, Seigler HF. A model for pretest probability of lymph node metastasis from cutaneous Melanoma. *Am J Clin Pathol* 2000;**114**:875–9	No results by stage
Vuylsteke RJ, van Leeuwen PA, Statius Muller MG, Gietema HA, Kragt DR, Meijer S. Clinical outcome of stage I/II melanoma patients after selective sentinel lymph node dissection: long-term follow-up results. *J Clin Oncol* 2003;**21**:1057–65	No validation
Wan X, Liu R, Li Z. The prognostic value of HRAS mRNA expression in cutaneous melanoma. *Biomed Res Int* 2017;**2017**:5356737	No validation
Wang K, Zhang ZW. Expression of miR-203 is decreased and associated with the prognosis of melanoma patients. *Int J Clin Exp Pathol* 2015;**8**:13249–54	Single prognostic factor
Wang Q, Wang X, Liang Q, Wang S, Xiwen L, Pan F, *et al.* Distinct prognostic value of mRNA expression of guanylate-binding protein genes in skin cutaneous melanoma. *Oncol Lett* 2018;**15**:7914–22	Stage not reported
Wang Q, Wang X, Liang Q, Wang S, Liao X, Li D, Pan F. Prognostic value of dynactin mRNA expression in cutaneous melanoma. *Med Sci Monit* 2018;**24**:3752–63	Stage I and II

Study	Reason(s) for exclusion
Weinlich G, Bitterlich W, Mayr V, Fritsch PO, Zelger B. Metallothionein-overexpression as a prognostic factor for progression and survival in melanoma. A prospective study on 520 patients. *Br J Dermatol* 2003;**149**:535–41	Single prognostic factor
Weinlich G, Eisendle K, Hassler E, Baltaci M, Fritsch PO, Zelger B. Metallothionein – overexpression as a highly significant prognostic factor in melanoma: a prospective study on 1270 patients. *Br J Cancer* 2006;**94**:835–41	Single prognostic factor
White RL, Ayers GD, Stell VH, Ding S, Gershenwald JE, Salo JC, *et al.* Factors predictive of the status of sentinel lymph nodes in melanoma patients from a large multicenter database. *Ann Surg Oncol* 2011;**18**:3593–600	No validation
Xing Y, Badgwell BD, Ross MI, Gershenwald JE, Lee JE, Mansfield PF, *et al.* Lymph node ratio predicts disease-specific survival in melanoma patients. *Cancer* 2009;**115**:2505–13	No validation
Yuan H, Liu H, Liu Z, Zhu D, Amos CI, Fang S, *et al.* Genetic variants in Hippo pathway genes YAP1, TEAD1 and TEAD4 are associated with melanoma-specific survival. *Int J Cancer* 2015;**137**:638–45	Stage I/II combined
Yun SJ, Gimotty PA, Hwang WT, Dawson P, Van Belle P, Elder DE, *et al.* High lymphatic vessel density and lymphatic invasion underlie the adverse prognostic effect of radial growth phase regression in melanoma. *Am J Surg Pathol* 2011;**35**:235–42	No validation
Zhang Z, Chen G, Cheng Y, Martinka M, Li G. Prognostic significance of RUNX3 expression in human melanoma. *Cancer* 2011;**117**:2719–27	No results by stage
Zietek M, Donizy P, Leskiewicz M, Kaczorowski M, Kozyra C, Wojnar A, *et al.* ALCAM overexpression in primary tumour predicts shorter overall survival in cutaneous malignant melanoma patients. *Eur J Surg Oncol* 2014;**40**:S149	Stage not reported

Appendix 5 Example search strategy for diagnostic accuracy review

#	Searches	Results (*n*)
Melanoma		
1	exp melanoma/or melanoma*.ti,ab,kw,kf.	124,381
2	amelanotic.ti,ab,kw,kf.	1956
3	((lentigo* or lentigin*) adj3 (tumo?r* or cancer* or maligna* or n?evus)).ti,ab,kw,kf.	1323
4	or/1-3	124,778
5	(ocular or uveal or iris or cornea or eye or choroidal or ciliary or intraocular).ti.	160,371
6	4 not 5	117,550
Diagnostic method		
7	exp ultrasonography, doppler/or exp ultrasonography, interventional/	89,434
8	ultrasound.ti,ab,kw,kf.	224,479
9	ultrason*.ti,ab,kw,kf.	154,906
10	exp image guided biopsy/	4564
11	exp Biopsy, Fine-Needle/	14,316
12	("fine needle biopsy" or FNB).ti,ab,kw,kf.	2199
13	("fine needle aspirat*" or FNA).ti,ab,kw,kf.	29,692
14	("fine needle aspiration cytology" or FNAC).ti,ab,kw,kf.	8227
15	exp Biopsy, Needle/	62,995
16	"core biopsy".ti,ab,kw,kf.	3971
17	or/7-16	455,777
Diagnostic test filter		
18	exp "sensitivity and specificity"/	544,996
19	(sensitivity or specificity or accuracy).tw.	1,228,698
20	((predictive adj3 value$) or (roc adj curve$)).tw.	123,584
21	((false adj positiv$) or false negativ$).tw.	71,980
22	((observer adj variation$) or (likelihood adj3 ratio$)).tw.	15,344
23	likelihood function/	20,943
24	exp mass screening/	119,419
25	diagnosis, differential/or exp Diagnostic errors/	535,465
26	di.xs. or du.fs.	3,322,576
27	or/18-26	4,663,491
Follow up/surveillance terms		
28	follow up.ti,ab,kw,kf.	872,434
29	surveillance.ti,ab,kw,kf.	159,164
30	monitor*.ti,ab,kw,kf.	729,895

#	Searches	Results (*n*)
31	exp Neoplasm Recurrence, Local/di, dg, pc [Diagnosis, Diagnostic Imaging, Prevention & Control]	21,379
32	exp Recurrence/di, dg, pc [Diagnosis, Diagnostic Imaging, Prevention & Control]	8
33	recur*.ti,ab,kw,kf.	538,847
34	exp Neoplasm Metastasis/di, dg, pc [Diagnosis, Diagnostic Imaging, Prevention & Control]	8124
35	metast*.ti,ab,kw,kf.	462,364
36	exp Aftercare/	177,538
37	or/28-36	2,591,386
38	6 and 17 and 27 and 37	928
39	exp animals/not humans.sh.	4,548,295
40	38 not 39	904

Appendix 6 Diagnostic accuracy review: excluded studies

Study	Reason(s) for exclusion
Ben Lakhdar A, Ilie M, Tomasic G, Chami L, Robert C, Vielh P. Benefits of ultrasound-guided fine needle aspirationcytology before lymph node biopsy in melanoma patients. *Virchows Arch* 2011;**459**:S10	Stage not reported (conference abstract)
Blum A, Schlagenhauff B, Stroebel W, Breuninger H, Rassner G, Garbe C. Ultrasound examination of regional lymph nodes significantly improves early detection of locoregional metastases during the follow-up of patients with cutaneous melanoma: results of a prospective study of 1288 patients. *Cancer* 2000;**88**:2534–9	No analysis by disease stage or Breslow depth
Blum A, Schmid-Wendtner MH, Mauss-Kiefer V, Eberle JY, Kuchelmeister C, Dill-Müller D. Ultrasound mapping of lymph node and subcutaneous metastases in patients with cutaneous melanoma: results of a prospective multicenter study. *Dermatology* 2006;**212**:47–52	Not diagnostic accuracy: ultrasonography mapping
Brountzos EN, Panagiotou IE, Bafaloukos DI, Kelekis DA. Ultrasonographic detection of regional lymph node metastases in patients with intermediate or thick malignant melanoma. *Oncol Rep* 2003;**10**:505–10	No analysis by disease stage or Breslow depth
Calvo López MJ, Vallejos Roca E, Muñoz Alcántara I, Navarro Díaz F, García Palacios MV. [Ultrasonographic and power Doppler appearance of locoregional metastases from cutaneous melanoma.] *Radiologia* 2008;**50**:483–8	Stage not reported (Spanish)
Caudron A, Chassine AF, Arnault JP, Dadban A, Chaby G, Lok C. Elastography as a new screening tool for metastatic lymph nodes in patients monitored for melanoma. *Melanoma Res* 2011;**1**:e32–e3	No analysis by disease stage or Breslow depth
Chai CY, Zager JS, Szabunio MM, Marzban SS, Chau A, Rossi RM, *et al.* Preoperative ultrasound is not useful for identifying nodal metastasis in melanoma patients undergoing sentinel node biopsy: preoperative ultrasound in clinically node-negative melanoma. *Ann Surg Oncol* 2012;**19**:1100–6	Pre-operative staging
Dalle S, Paulin C, Lapras V, Balme B, Ronger-Savle S, Thomas L. Fine-needle aspiration biopsy with ultrasound guidance in patients with malignant melanoma and palpable lymph nodes. *Br J Dermatol* 2006;**155**:552–6	No analysis by disease stage or Breslow depth
Fakhry N, Tessonnier L, Cohen F, Gras R, Grob JJ, Giovanni A, *et al.* Management of cervical lymph node recurrence of melanoma of the head and neck. *Rev Laryngol Otol Rhinol* 2009;**130**:211–14	No analysis by disease stage or Breslow depth
Galanzha EI, Menyaev YA, Yadem AC, Sarimollaoglu M, Juratli MA, Nedosekin DA, *et al.* In vivo liquid biopsy using Cytophone platform for photoacoustic detection of circulating tumor cells in patients with melanoma. *Sci Transl Med* 2019;**11**:eaat5857	Advanced stage
Georgieva M, Prantl L, Utpatel K, Wiesinger I, Stroszczynski C, Jung F, Jung EM. Diagnostic performance of ultrasound strain elastography for differentiation of malignant breast lesions. *Clin Hemorheol Microcirc* 2019;**71**:237–47	Insufficient patients
Hayes AJ, Moskovic E, O'Meara K, Smith HG, Pope RJE, Larkin J, Thomas JM. Prospective cohort study of ultrasound surveillance of regional lymph nodes in patients with intermediate-risk cutaneous melanoma. *Br J Surg* 2019;**106**:729–34	Stages I to II

Study	Reason(s) for exclusion
Herceg GH, Bracic I, Kusacic-Kuna S, Mutvar A, Antulov J, Herceg D. Introduction of US-guided FNAC in preoperative staging prior to sentinel lymph node biopsy: benefit for patients with cutaneous melanoma. *Nuklearmedizin* 2014;**53**:A132	Pre-operative staging (conference abstract)
Heřman J, Sedláčková Z, Fürst T, Vachutka J, Salzman R, Vomáčka J, Heřman M. The role of ultrasound and shear-wave elastography in evaluation of cervical lymph nodes. *Biomed Res Int* 2019;**2019**:4318251	Insufficient patients
Hinz T, Hoeller T, Wenzel J, Bieber T, Schmid-Wendtner MH. Real-time tissue elastography as promising diagnostic tool for diagnosis of lymph node metastases in patients with malignant melanoma: a prospective single-center experience. *Dermatology* 2013;**226**:81–90	No analysis by disease stage or Breslow depth
Hinz T, Wilsmann-Theis D, Buchner A, Wenzel J, Wendtner CM, Bieber T, *et al.* High-resolution ultrasound combined with power Doppler sonography can reduce the number of sentinel lymph node biopsies in cutaneous melanoma. *Dermatology* 2011;**222**:180–8	Pre-operative staging
Hocevar M, Bracko M, Pogacnik A, Vidergar-Kralj B, Besic N, Zgajnar J, Music MM. The role of preoperative ultrasonography in reducing the number of sentinel lymph node procedures in melanoma. *Melanoma Res* 2004;**14**:533–6	No analysis by disease stage or Breslow depth
Hofmann U, Szedlak M, Rittgen W, Jung EG, Schadendorf D. Primary staging and follow-up in melanoma patients – monocenter evaluation of methods, costs and patient survival. *Br J Cancer* 2002;**87**:151–7	Stages I–II
Horvatic Herceg G, Bracic I, Kusacic-Kuna S, Herceg D, Mutvar A, Dodig D. Ultrasound and US-guided FNAC can reduce the number of sentinel lymph node biopsies in cutaneous melanoma. *Eur J Nucl Med Mol Imaging* 2012;**39**:S591	Stages I–II (conference abstract)
Kaushik V, Baloch Z, Jones L, Gupta P. Diagnostic utility of ultrasound-guided fine needle aspiration (USG-FNA) in head and neck lesions detected by positron emission tomography (PET) scan. *J Am Soc Cytopathol* 2012;**1**:S72	Stage not reported (conference abstract)
Kiyohara Y, Tsuchiya T, Nakajima M, Adachi M, Shimizu M, Hatano M, *et al.* Evaluation of color and power Doppler images in malignant melanoma. *Ultrasound Med Biol* 2000;**26**:A184	Stage not reported (conference abstract)
Klebl FH, Gelbmann CM, Lammert I, Bogenrieder T, Stolz W, Schölmerich J, Schlottmann K. [Detection of lymph node metastases of malignant melanoma by palpation and ultrasound.] *Med Klin* 2003;**98**:783–7	No analysis by disease stage or Breslow depth (German)
Kunte C, Schuh T, Eberle JY, Baumert J, Konz B, Volkenandt M, *et al.* The use of high-resolution ultrasonography for preoperative detection of metastases in sentinel lymph nodes of patients with cutaneous melanoma. *Dermatol Surg* 2009;**35**:1757–65	Pre-operative staging (German)
Lassau N, Koscielny S, Avril MF, Margulis A, Duvillard P, De Baere T, *et al.* Prognostic value of angiogenesis evaluated with high-frequency and color Doppler sonography for preoperative assessment of melanomas. *AJR Am J Roentgenol* 2002;**178**:1547–51	Pre-operative staging
Lassau N, Chami L, Peronneau P. [Imaging of melanoma: accuracy of ultrasonography before and after contrast injection for diagnostic and early evaluation of treatments.] *Bull Cancer* 2007;**94**:93–8	Stage not reported (French)
Layfield LJ. Diagnosis and work-up of malignant melanoma in the age of fine needle aspiration and molecular testing. *Eur Oncol Haematol* 2014;**10**:58–61	Review
Lean CL, Bourne R, Thompson JF, Scolyer RA, Stretch J, Li LX, *et al.* Rapid detection of metastatic melanoma in lymph nodes using proton magnetic resonance spectroscopy of fine needle aspiration biopsy specimens. *Melanoma Res* 2003;**13**:259–61	Stage not reported

Study	Reason(s) for exclusion
Levang J, Manzoni P, Puyraveau M, Sarlieve P, Puzenat E, Humbert P, *et al.* The value of contrast-enhanced ultrasonography in the detection of liver metastases in the follow-up of patients with stages III and IV melanoma. *Arch Dermatol Res* 2007;**299**:282	Stage III–IV (conference abstract)
Lussier C, Klijanienko J, Vielh P. Fine-needle aspiration of metastatic nonlymphomatous tumors to the major salivary glands: a clinicopathologic study of 40 cases cytologically diagnosed and histologically correlated. *Cancer* 2000;**90**:350–6	Stage not reported
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. *Br J Dermatol* 2005;**152**:66–70	No analysis by disease stage or Breslow depth
Machet L, Vaillant L, Lorette G. Follow-up of excision of cutaneous melanoma: sentinel node biopsy, lymph node ultrasound or clinical surveillance alone? *Ann Dermatol Venereol* 2005;**132**:941–4	Editorial
Marone U, Catalano O, Caracò C, Anniciello AM, Sandomenico F, Di Monta G, *et al.* Can high-resolution ultrasound avoid the sentinel lymph-node biopsy procedure in the staging process of patients with stage I-II cutaneous melanoma? *Ultraschall Med* 2012;**33**:E179–85	Pre-operative staging
Metzger S, Dohmen BM, Breuninger H, Rassner G, Flerlbeck G. Sensitivity and specificity of 18FDG-PET in the staging diagnosis of patients with high-risk melanomas in comparison with sonography and CT. *Z Hautkr* 2000;**75**:465	Stage not reported
Moehrle M, Blum A, Rassner G, Juenger M. Lymph node metastases of cutaneous melanoma: diagnosis by B-scan and color Doppler sonography. *J Am Acad Dermatol* 1999;**41**:703–9	Stage not reported
Molajo A, Powell B. Ultrasound guided FNAC of sentinel nodes in melanoma. Which patients may be suitable? *J Dtsch Dermatol Ges* 2013;**11**:36	Not diagnostic accuracy of ultrasonography (conference abstract)
Murali R, Doubrovsky A, Watson GF, McKenzie PR, Lee CS, McLeod DJ, *et al.* Diagnosis of metastatic melanoma by fine-needle biopsy: analysis of 2,204 cases. *Am J Clin Pathol* 2007;**127**:385–97	Stage not reported
Nasuti JF, Yu G, Boudousquie A, Gupta P. Diagnostic value of lymph node fine needle aspiration cytology: an institutional experience of 387 cases observed over a 5-year period. *Cytopathology* 2000;**11**:18–31	Stage not reported
Nazarian LN, Alexander AA, Kurtz AB, Capuzzi DM, Rawool NM, Gilbert KR, Mastrangelo MJ. Superficial melanoma metastases: appearances on gray-scale and color Doppler sonography. *AJR Am J Roentgenol* 1998;**170**:459–63	Stage not reported
Oehr P, Stegemann G, Steen K, Ruhlmann J. The value of FDG-PET whole body imaging, conventional imaging, and serum S-100 determinations in metastatic malignant melanoma. *Clin Lab* 1999;**45**:523–8	No analysis by disease stage or Breslow depth: ultrasonography results not reported
Ogata D, Uematsu T, Yoshikawa S, Kiyohara Y. Accuracy of real-time ultrasound elastography in the differential diagnosis of lymph nodes in cutaneous malignant melanoma (CMM): a pilot study. *Int J Clin Oncol* 2014;**19**:716–21	Stage not reported
Olmedo D, Brotons-Seguí M, Del Toro C, González M, Requena C, Traves V, *et al.* Use of lymph node ultrasound prior to sentinel lymph node biopsy in 384 patients with melanoma: a cost-effectiveness analysis. *Actas Dermosifiliogr* 2017;**108**:931–8	Stage not reported
Olson MT, Novak A, Kirby J, Shahid H, Boonyaarunnate T, Ali SZ. Cytotechnologist-attended on-site evaluation of adequacy for metastatic disease involving bone and soft tissue. *Acta Cytol* 2013;**57**:550–6	Stage not reported

Study	Reason(s) for exclusion
Olszanski AJ. mutation testing and adjuvant systemic therapy in cutaneous melanoma. *J Natl Compr Canc Netw* 2019;**17**:615–17	Treatment
Oude Ophuis CMC, Koppert LB, de Monyé C, van Deurzen CHM, Koljenović S, van Akkooi ACJ, *et al.* Gamma probe and ultrasound guided fine needle aspiration cytology of the sentinel node (GULF) trial – overview of the literature, pilot and study protocol. *BMC Cancer* 2017;**17**:258	No analysis by disease stage or Breslow depth
Oude Ophuis CMC, Verhoef C, Grünhagen DJ, Siegel P, Schoengen A, Röwert-Huber J, *et al.* Long-term results of ultrasound guided fine needle aspiration cytology in conjunction with sentinel node biopsy support step-wise approach in melanoma. *Eur J Surg Oncol* 2017;**43**:1509–16	Not diagnostic study: protocol
Panagiotou IE, Brountzos EN, Bafaloukos D, Tsavaris N, Mylonakis N, Karabelis A, *et al.* Evaluation of imaging studies at the initial staging and during follow-up of patients with local-regional malignant melanoma. *JBUON* 2001;**6**:411–4	Initial stage not reported: advanced stage
Pánczél G, Liszkay G, Borbola K, Balatoni T, Hunyadi J. [The importance of fine needle aspiration cytology in the management of recurrent and metastatic melanoma.] *Orv Hetil* 2012;**153**:1419–23	Hungarian translation required
Papadopoulos O, Konofaos P, Georgoulakis J, Chrisostomidis C, Tsantoulas Z, Kostopoulos E, *et al.* The role of ThinPrep cytology in the investigation of SLN status in patients with cutaneous melanoma. *Surg Oncol* 2007;**16**:121–9	No analysis by disease stage or Breslow depth
Pilko G, Zgajnar J, Music M, Hocevar M. Lower tumour burden and better overall survival in melanoma patients with regional lymph node metastases and negative preoperative ultrasound. *Radiol Oncol* 2012;**46**:60–8	No relevant outcomes
Porcellato I, Brachelente C, De Paolis L, Menchetti L, Silvestri S, Sforna M, *et al.* FoxP3 and IDO in canine melanocytic tumors. *Vet Pathol* 2019;**56**:189–99	Animals
Proebstle T, Schwurzer-Voit M, Sterry W, Knop J, Voit C. Detection of regional melanoma metastases by ultrasound B-scan, cytology or tyrosinase RT-PCR of fine needle aspirates. *J Invest Dermatol* 1999;**113**:514	Stage not reported (conference abstract)
Radzhabova ZA, Barchuk AS, Kostromina EV, Anisimov VV. [The detection of early regional metastases in patients with skin melanoma by dopplerography.] *Vestn Khir Im I I Grek* 2009;**168**:50–3	Stage not reported (Russian)
Ribero S, Podlipnik S, Osella-Abate S, Sportoletti-Baduel E, Manubens E, Barreiro A, *et al.* Ultrasound-based follow-up does not increase survival in early-stage melanoma patients: a comparative cohort study. *Eur J Cancer* 2017;**85**:59–66	No relevant outcomes
Rodrigues LKE, Leong SPL, Ljung BM, Sagebiel RW, Burnside N, William Hu TL, *et al.* Fine needle aspiration in the diagnosis of metastatic melanoma. *J Am Acad Dermatol* 2000;**42**:735–40	Stage not reported
Rossi CR, Scagnet B, Vecchiato A, Mocellin S, Pilati P, Foletto M, *et al.* Sentinel node biopsy and ultrasound scanning in cutaneous melanoma: clinical and technical considerations. *Eur J Cancer* 2000;**36**:895–900	Pre-operative staging
Rubaltelli L, Beltrame V, Scagliori E, Bezzon E, Frigo AC, Rastrelli M, Stramare R. Potential use of contrast-enhanced ultrasound (CEUS) in the detection of metastatic superficial lymph nodes in melanoma patients. *Ultraschall Med* 2014;**35**:67–71	No analysis by disease stage or Breslow depth
Rubaltelli L, Beltrame V, Tregnaghi A, Scagliori E, Frigo AC, Stramare R. Contrast-enhanced ultrasound for characterizing lymph nodes with focal cortical thickening in patients with cutaneous melanoma. *AJR Am J Roentgenol* 2011;**196**:W8–12	Stages I–II

Study	Reason(s) for exclusion
Rue Nielsen K, Klyver H, Hougaard Chakera A, Nedergaard L, Hesse B, Bachmann Nielsen M. Sentinel node detection in melanomas using contrast-enhanced ultrasound. *Acta Radiol* 2009;**50**:412–17	No relevant outcomes
Saiag P, Bernard M, Beauchet A, Bafounta ML, Bourgault-Villada I, Chagnon S. Ultrasonography using simple diagnostic criteria vs. palpation for the detection of regional lymph node metastases of melanoma. *Arch Dermatol* 2005;**141**:183–9	No analysis by disease stage or Breslow depth
Saiag P, Lebbe C, Basset-Seguin N, Wolkenstein P, Dupin N, Descamps V, *et al.* Role of lymph-node ultrasonography (US) in the follow-up of melanoma patients to detect nodal recurrence after sentinel lymph node biopsy (SNLB): a prospective cohort study. *Melanoma Res* 2010;**20**:e31	Stage not reported (conference abstract)
Saiag P, Lebbe C, Seguin NB, Wolkenstein P, Dupin N, Descamps V, *et al.* Role of lymph-node ultrasonography (US) in the follow-up of melanoma patients to detect nodal recurrence after sentinel lymph node biopsy (SNLB): a prospective cohort study. *J Clin Oncol* 2010;**28**:1	Stage not reported (conference abstract)
Samimi M, Perrinaud A, Naouri M, Maruani A, Perrodeau E, Vaillant L, Machet L. High-resolution ultrasonography assists the differential diagnosis of blue naevi and cutaneous metastases of melanoma. *Br J Dermatol* 2010;**163**:550–6	No analysis by disease stage
Sanki A, Uren RF, Moncrieff M, Tran KL, Scolyer RA, Lin HY, *et al.* Targeted high-resolution ultrasound is not an effective substitute for sentinel lymph node biopsy in patients with primary cutaneous melanoma. *J Clin Oncol* 2009;**27**:5614–9	Pre-operative staging; no analysis by disease stage at recurrence
Sasaki Y, Kanki H, Nagano T, Nishigori C, Fukuoka K, Kumagai S. Assessment of sentinel lymph nodes identified by lymphoscintigraphy, compared histopathology to ultrasonography. *Skin Res* 2008;**7**:586–92	Insufficient patients (Japanese)
Schmid-Wendtner MH, Dill-Müller D, Baumert J, Wagner A, Eberle J, Tilgen W, Plewig G. Lymph node metastases in patients with cutaneous melanoma: improvements in diagnosis by signal-enhanced color Doppler sonography. *Melanoma Res* 2004;**14**:269–76	No analysis by disease stage or Breslow depth
Schmid-Wendtner MH, Paerschke G, Baumert J, Plewig G, Volkenandt M. Value of ultrasonography compared with physical examination for the detection of locoregional metastases in patients with cutaneous melanoma. *Melanoma Res* 2003;**13**:183–8	No analysis by disease stage or Breslow depth
Schmid-Wendtner MH, Partscht K, Korting HC, Volkenandt M. Improved differentiation of benign and malignant lymphadenopathy in patients with cutaneous melanoma by contrast-enhanced color Doppler sonography. *Arch Dermatol* 2002;**138**:491–7	No analysis by disease stage or Breslow depth
Sijan G, Kozarski J, Stefanović D, Lalković M, Milićević S, Stanković G. [Ultrasonographic findings validity in the identification of metastatic regional lymph nodes in patients with cutaneous melanoma.] *Vojnosanit Pregl* 2010;**67**:25–31	Advanced stage (Croatian)
Šijan G, Kozarski J, Stepić N, Milojević S, Stefanović D, Tatomirović Ž, *et al.* Validity of ultrasound-guided aspiration needle biopsy in the diagnosis of micrometastases in sentinel lymph nodes in patients with cutaneous melanoma. *Vojnosanit Pregl* 2016;**73**:934–40	Pre-operative staging of lymph node
Solivetti FM, Desiderio F, Guerrisi A, Bonadies A, Maini CL, Di Filippo S, *et al.* HF ultrasound vs PET-CT and telethermography in the diagnosis of In-transit metastases from melanoma: a prospective study and review of the literature. *J Exp Clin Cancer Res* 2014;**33**:96	Pre-operative staging
Solivetti FM, Di Luca Sidozzi A, Pirozzi G, Coscarella G, Brigida R, Eibenshutz L. Sonographic evaluation of clinically occult in-transit and satellite metastases from cutaneous malignant melanoma. *Radiol Med* 2006;**111**:702–8	No analysis by disease stage or Breslow depth

Study	Reason(s) for exclusion
Solivetti FM, Elia F, Graceffa D, Di Carlo A. Ultrasound morphology of inguinal lymph nodes may not herald an associated pathology. *J Exp Clin Cancer Res* 2012;**31**:88	Pre-operative staging
Solivetti FM, Elia F, Santaguida MG, Guerrisi A, Visca P, Cercato MC, Di Carlo A. The role of ultrasound and ultrasound-guided fine needle aspiration biopsy of lymph nodes in patients with skin tumours. *Radiol Oncol* 2014;**48**:29–34	No analysis by disease stage or Breslow depth
Teng E, Sue GR, Sawh-Martinez R, Nishikawa S, Ariyan S, Natarajan A, Narayan D. Scalp melanoma and in-transit metastases: a retrospective case-controlled study. *Am Surg* 2014;**80**:1272–4	No diagnostic tests reported
Ternov NK, Lambine TL, Wagenblast ALH, Clasen-Linde E, Oturai PS, Klyver H, *et al.* Targeted ultrasound and fine-needle aspiration cytology for sentinel node diagnostics in early-stage melanoma: a validation study. *Melanoma Res* 2018;**28**:319–25	Pre-operative staging
Testori A, Lazzaro G, Baldini F, Tosti G, Mosconi M, Lovati E, *et al.* The role of ultrasound of sentinel nodes in the pre- and post-operative evaluation of stage I melanoma patients. *Melanoma Res* 2005;**15**:191–8	Insufficient diagnostic data at recurrence
Testori A, Rastrelli M, De Fiori E, Soteldo J, Della Vigna P, Trifirò G, *et al.* Radio-guided ultrasound lymph node localization: feasibility of a new technique for localizing and excising nonpalpable lymph nodes ultrasound suspicious for melanoma metastases. *Melanoma Res* 2010;**20**:197–202	Stage not reported
Thompson JF, Haydu LE, Sanki A, Uren RF. Ultrasound assessment of lymph nodes in the management of early-stage melanoma *J Surg Oncol* 2011;**104**:354–60	Review
Tombesi P, Tassinari D, Sartori S. Contrast-enhanced ultrasound for characterizing lymph nodes with focal cortical thickening in patients with cutaneous melanoma. *AJR Am J Roentgenol* 2011;**197**:W371	Letter
Tsimpaki T, Beis E, Othmer V, Grabbe S, Tuttenberg A. The role of preoperative ultrasound as an auxiliary tool for the detection of nodal micro- and macrometastasis in melanoma patients undergoing sentinel lymph node biopsy: a retrospective analysis. *J Dtsch Dermatol Ges* 2017;**15**:57–8	Stage not reported (poster)
Ulrich J, van Akkooi AC, Eggermont AM, Voit CA. [Sonographic criteria for diagnosing sentinel node metastases in melanoma patients.] *Ultraschall Med* 2015;**36**:149–53	No analysis by disease stage or Breslow depth
Uren RF, Howman-Giles R, Thompson JF, Shaw HM, Roberts JM, Bernard E, McCarthy WH. High-resolution ultrasound to diagnose melanoma metastases in patients with clinically palpable lymph nodes. *Australas Radiol* 1999;**43**:148–52	No analysis by disease stage or Breslow depth
Ustün M, Risberg B, Davidson B, Berner A. Cystic change in metastatic lymph nodes: a common diagnostic pitfall in fine-needle aspiration cytology. *Diagn Cytopathol* 2002;**27**:387–92	No analysis by disease stage or Breslow depth
Val-Bernal JF, Martino M, Yllera E, Romay F, Sánchez-Ares M, Nallib IA. Endobronchial ultrasound-guided transbronchial needle aspiration in the diagnosis of hilar and mediastinal lymph node metastases of melanoma. *Turk Patoloji Derg* 2019;**35**:92–101	Insufficient number of stage I patients
van Akkooi ACJ, Siegel P, Gooskens S, Schoengen A, Sterry W, Eggermont AM, *et al.* Use of preoperative ultrasound (US)-guided fine needle aspiration cytology (FNAC) to identify positive sentinel nodes (SN) in melanoma. *J Clin Oncol Conf* 2013;**31**:e20035	Stage I–II (conference abstract)
van Akkooi ACJ, Siegel P, Schoengen A, Roewert-Huber J, Eggermont AM, Voit CA. Long-term results of ultrasound (US)-guided fine needle aspiration cytology (FNAC) in conjunction with sentinel node biopsy (SNB) to support step-wise approach in melanoma. *J Clin Oncol Conf* 2015;**33**:9067	Stage I–II (conference abstract)

Study	Reason(s) for exclusion
Vensby PH, Schmidt G, Kjær A, Fischer BM. The value of FDG PET/CT for follow-up of patients with melanoma: a retrospective analysis. *Am J Nucl Med Mol Imaging* 2017;**7**:255–62	No analysis by disease stage or Breslow depth
Voit C, Kron M, Schäfer G, Schoengen A, Audring H, Lukowsky A, *et al.* Ultrasound-guided fine needle aspiration cytology prior to sentinel lymph node biopsy in melanoma patients. *Ann Surg Oncol* 2006;**13**:1682–9	No analysis by disease stage or Breslow depth
Voit C, Mayer T, Kron M, Schoengen A, Sterry W, Weber L, Proebstle TM. Efficacy of ultrasound B-scan compared with physical examination in follow-up of melanoma patients. *Cancer* 2001;**91**:2409–16	No analysis by disease stage or Breslow depth
Voit C, Mayer T, Proebstle TM, Weber L, Kron M, Krupienski M, *et al.* Ultrasound-guided fine-needle aspiration cytology in the early detection of melanoma metastases. *Cancer* 2000;**90**:186–93	No analysis by disease stage or Breslow depth
Voit C, Van Akkooi ACJ, Schaefer G, Schoengen A, Sterry W, Eggermont AMM. Early ultrasound criteria drive sensitivity for detection of sentinel node metastases in melanoma patients: a prospective study in 800 patients. *Pigment Cell Melanoma Res* 2010;**23**:984	Stages I-II (conference abstract)
Voit CA, Gooskens S, Van Akkooi ACJ, Eggermont AMM. Ultrasound (US) – guided fine needle aspiration cytology (FNAC) of the sentinel node (SN) in 1000 consecutive melanoma patients. *Eur J Cancer* 2013;**2**:S855	Stage I-II (conference abstract)
Voit CA, Oude Ophuis CM, Ulrich J, van Akkooi AC, Eggermont AM. Ultrasound of the sentinel node in melanoma patients: echo-free island is a discriminatory morphologic feature for node positivity. *Melanoma Res* 2016;**26**:267–71	No analysis by disease stage or Breslow depth
Voit CA, van Akkooi AC, Schaefer-Hesterberg G, Schoengen A, Sterry W, Eggermont AM. Correlation of ultrasound criteria for detection of melanoma metastases in the sentinel lymph node (SN) with tumor burden and survival. *J Clin Oncol* 2009;**27**:9015	Initial staging – Rotterdam criteria
Voit CA, van Akkooi ACJ, Catalano O, Eggermont AMM. Pre-SN ultrasound-FNAC can be sensitive for lymph node metastases in melanoma patients if performed with the use of the Berlin criteria. *Ann Surg Oncol* 2017;**24**:661–2	Letter
Voit CA, Van Akkooi ACJ, Siegel P, Schoengen A, Sterry W, Eggermont AMM. High sensitivity rate of ultrasound (US) guided fine needle aspiration cytology (FNAC) using the Berlin morphology criteria for lymph node metastases significantly reduces need for surgical sentinel node (SN) staging in melanoma. *Skin Res Technol* 2012;**19**:e574–e5	Stage I–II (conference abstract)
Voit CA, Van Akkooi ACJ, Siegel P, Sterry W, Schoengen A, Schaefer-Hesterberg G, *et al.* Ultrasound (US)-guided fine-needle aspiration cytology (FNAC) for the prediction of sentinel node (SN) metastases and its effect on the nomogram for melanoma patients. *J Clin Oncol Conf* 2011;**29**:8850	No analysis by disease stage or Breslow depth (conference abstract)
Voit CA, van Akkooi AJ, Schäfer-Hesterberg G, Sterry W, Eggermont AM. The value of preoperative ultrasound (after lymphoscintigraphy) in conjunction with pre-sentinel lymph node biopsy fine-needle aspiration outweighs the usage of ultrasound alone in conjunction with lymphoscintigraphy: the need for an algorithm. *Melanoma Res* 2010;**20**:357–9	Letter

Appendix 7 Targeted economic search

TABLE 22 Summary of targeted economic review protocol

Objectives and research questions	
Primary objective	To review primary health economic evaluation for surveillance strategies for stage I melanoma post primary tumour excision
Studies to include	
Study designs	• Cost-effectiveness analyses • Cost–utility analyses • Cost–benefit analyses • Cost-minimisation analyses • Cost–consequence studies
Population	• Age: adults • Race: any • Disease: stage I cutaneous melanoma
Interventions	• Clinician visits by type • Clinician visit by interval • Prognostic • Diagnostic
Comparators	• Any of the included interventions • No follow-up strategy
Language	Studies with abstracts in the English language, but full-text published in a language other than English, will be evaluated; if local expertise is available, they will be included
Publication time frame	Studies published from start of database to present will be included in this review in order to retrieve evidence from all the available data. No restriction on publication period will be applied
Data sources	
Databases	• MEDLINE • EMBASE • NHS Economic Evaluation Database (NHS EED) • Cost-effectiveness Analysis Registry (CEA registry)
Information to extract (indicative list only. Outcomes to be finalised prior to data extraction after identification of all included studies)	
Study details	• Study name • Year and journal of publication • Type of evaluation • Study objective • Cost year and currency(ies) • Country(ies) • Intervention and comparator details
Population characteristics	• Disease type • Mean/median age
Basic modelling methodologies	• Perspective (health-care payer, societal) • Time horizon • Cycle length • Markov or discrete event structure or other types • Simulation method (cohort, patient level) • Capture second-order (parameter) uncertainty • Discounting • Model assumptions

continued

TABLE 22 Summary of targeted economic review protocol (*continued*)

Objectives and research questions	
Model structure and key data sources/risk equations	• Incorporation of treatment effects • Incorporation of health-related quality of life • Incorporation of resource use and costs (direct and indirect)
Outcomes	• Life-years gained • QALYs gained • ICERs • Details of sensitivity analyses results

Search strategy and number of hits for different searched databases

Dates searched: 29 March 2018 and 18 April 2018.

Number	Search term	Facet	Results (*n*)
NHS Economic Evaluation Database			
1	MELANOMA	Disease	321
2	SKIN TUMO(U)R		1
3	CUTANEOUS TUMO(U)R		0
4	SKIN CANCER		97
5	#1 AND Surveillance	Follow-up	12
6	#1 AND Monitoring		14
7	#1 AND Follow-up		92
8	#1 AND Screening		24
9	#1 AND Management		42
10	#4 AND Surveillance		4
11	#4 AND Monitoring		4
12	#4 AND Follow-up		33
13	#4 AND Screening		19
14	#4 AND Management		14
15	5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14	Final numbers	260
Cost-effectiveness Analysis Registry			
1	MELANOMA	Disease	33
2	SKIN TUMO(U)R		0
3	CUTANEOUS TUMO(U)R		0
4	SKIN CANCER		9
13	or/1-4	Final numbers	42

MEDLINE (via Ovid) without revisions

Date range searched: 1996 to April week 2 2018.

Number	Search term	Facet	Results (*n*)
1	exp Melanoma/	Disease	52,542
2	melanoma$.tw.		62,394
3	(maligna$ adj1 lentigo$).tw.		623
4	(hutchinson$ adj1 (freckle$ or melano$)).tw.		10
5	dubreuilh.tw.		8
6	Melanoma/		508
7	or/1-6		70,232
8	(follow-up or "follow up" or followup).tw.	Follow-up	589,929
9	(check-up*1 or check up*1).tw.		5394
10	surveillance.tw.		105,813
11	exp Aftercare/		110,607
12	(aftercare or after-care).tw.		1974
13	((post-treatment or posttreatment) adj1 evaluation*).tw.		330
14	((post-treatment or posttreatment) adj1 care).tw.		83
15	((post-treatment or posttreatment) adj1 monitoring).tw.		117
16	((post-treatment or posttreatment) adj1 surveillance).tw.		237
17	or/8-16		789,181
18	7 and 17		6269
19	exp models, economic/	Economics	11,826
20	*models, theoretical/		44,201
21	*models, organizational/		5210
22	Markov chains/		11,638
23	monte carlo method/		22,458
24	exp decision theory		8779
25	(Markov* or monte carlo).ti, ab.		34,389
26	econom* model*.ti.ab		2302
27	(decision* adj2 (tree* or analy* or model*)).ti,ab.		13,341
28	19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27		117,975
29	18 and 28	Final Numbers	29
30	limit 29 to yr="2015 -Current"		9

EMBASE

Date range searched: 1996 to 2018 week 16.

Number	Search term	Facet	Results (*n*)
1	exp Melanoma/	Disease	110,054
2	melanoma$.tw.		109,004
3	(maligna$ adj1 lentigo$).tw.		1068
4	(hutchinson$ adj1 (freckle$ or melano$)).tw.		17
5	dubreuilh.tw.		19
6	Melanoma/		917
7	or/1-6		135,649
8	(follow-up or "follow up" or followup).tw.	Follow-up	1,144,022
9	(check-up*1 or check up*1).tw.		9880
10	surveillance.tw.		178,560
11	exp Aftercare/		1,195,286
12	(aftercare or after-care).tw.		3395
13	((post-treatment or posttreatment) adj1 evaluation*).tw.		634
14	((post-treatment or posttreatment) adj1 care).tw.		177
15	((post-treatment or posttreatment) adj1 monitoring).tw.		199
16	((post-treatment or posttreatment) adj1 surveillance).tw.		499
17	or/8-16		1,626,817
18	7 and 17		16,537
19	statistical model/	Economics	142,384
20	exp economic aspect/		1,155,950
21	19 and 20		19,009
22	*theoretical model/		23,005
23	*nonbiological model/		4027
24	stochastic model/		9777
25	decision theory/		1303
26	decision tree/		8931
27	monte carlo method/		32,115
28	(Markov* or monte carlo).ti,ab.		54,952
29	econom* model*.ti,ab.		4413
30	(decision* adj2 (tree* or analy* or model*)).ti,ab.		23,730
31	19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28		1,371,121
32	18 and 31	Final Numbers	756
33	limit 32 to yr="2015 -Current"		289

The total number of studies retrieved by each individual database is provided in *Table 23*.

After reviewing titles and abstracts, 15 studies were fully reviewed to see if any model could be adapted to answer the research question and inform the structure of the model. No study was judged as being suitable for adaptation as the structure of the model would need to be malleable to patient behaviour, such as self-diagnosis and false alarms.

TABLE 23 Total number of studies retrieved

Database	Retrieved (*n*)
NHS Economic Evaluation Database	260
Cost-effectiveness Analysis Registry	42
MEDLINE	9
EMBASE	289
Total	590

Appendix 8 Model structure

A simplified version of the Markov model is shown in *Figure 6*. This appendix presents the detailed structure of this model.

With no recurrence, patients will end up back in the disease-free health cycle for the next cycle (*Figure 21*). However, there is a chance that patients will have a 'false alarm' that results in an 'E Visit: Emergency visit' to their local dermatologist. There is a chance that, after a history and physical examination, a local biopsy would be taken and a false-positive result will ensue, which may result in a SLNB. All patients return to the disease-free health state in the Markov model.

However, if a recurrence does occur (*Figure 22*), it can be stage IA to IV; this can be picked up by a scheduled screening/appointment with a health-care professional that is part of the surveillance strategy or it can be self-diagnosed (opportunistic diagnosis). The clinical pathway consists of local biopsy (B) (*Figure 23*), SLNB (C) (*Figure 24*) and CT (D) (*Figure 25*). If the recurrence is not picked up and the patient survives, their melanoma progresses (E) (*Figure 26*). Patients will re-enter the model in the next cycle, where they go through the same process. If treatment is successful, patients return to the disease-free health state.

Surveillance strategies were compared based on the 'p_screen' variable, with associated sensitivity and specificity, and costs, based on the health-care professional involved in the screening. Each surveillance strategy is captured by a 'follow-up duration' variable and a 'follow-up interval' that contributes to lifetime costs and health outcomes expressed in terms of QALYs.

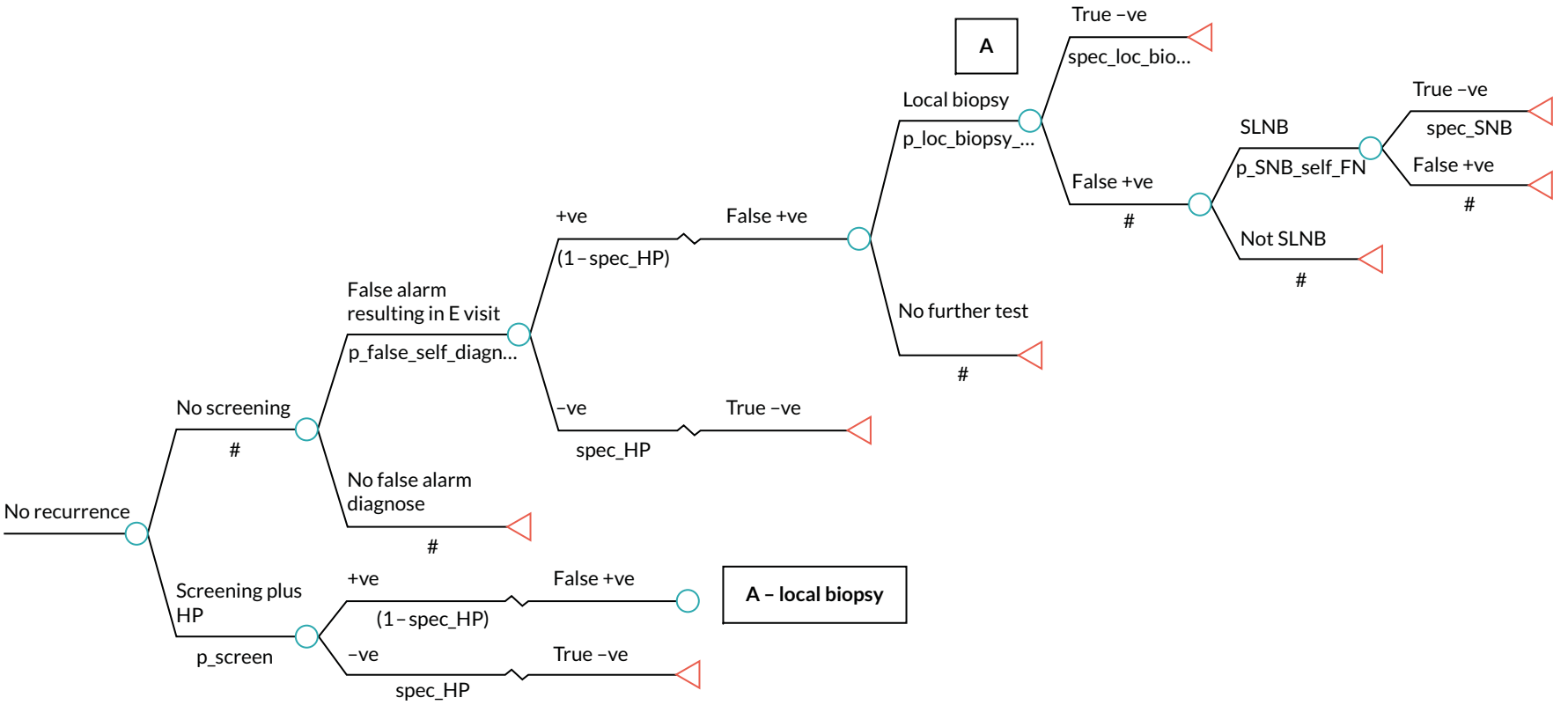


FIGURE 21 Disease-free health state: no recurrence. +ve, positive; –ve, negative; E visit, Emergency visit; FN, False negative; HP, health and physical examination; SNB, sentinel node biopsy; spec, specificity.

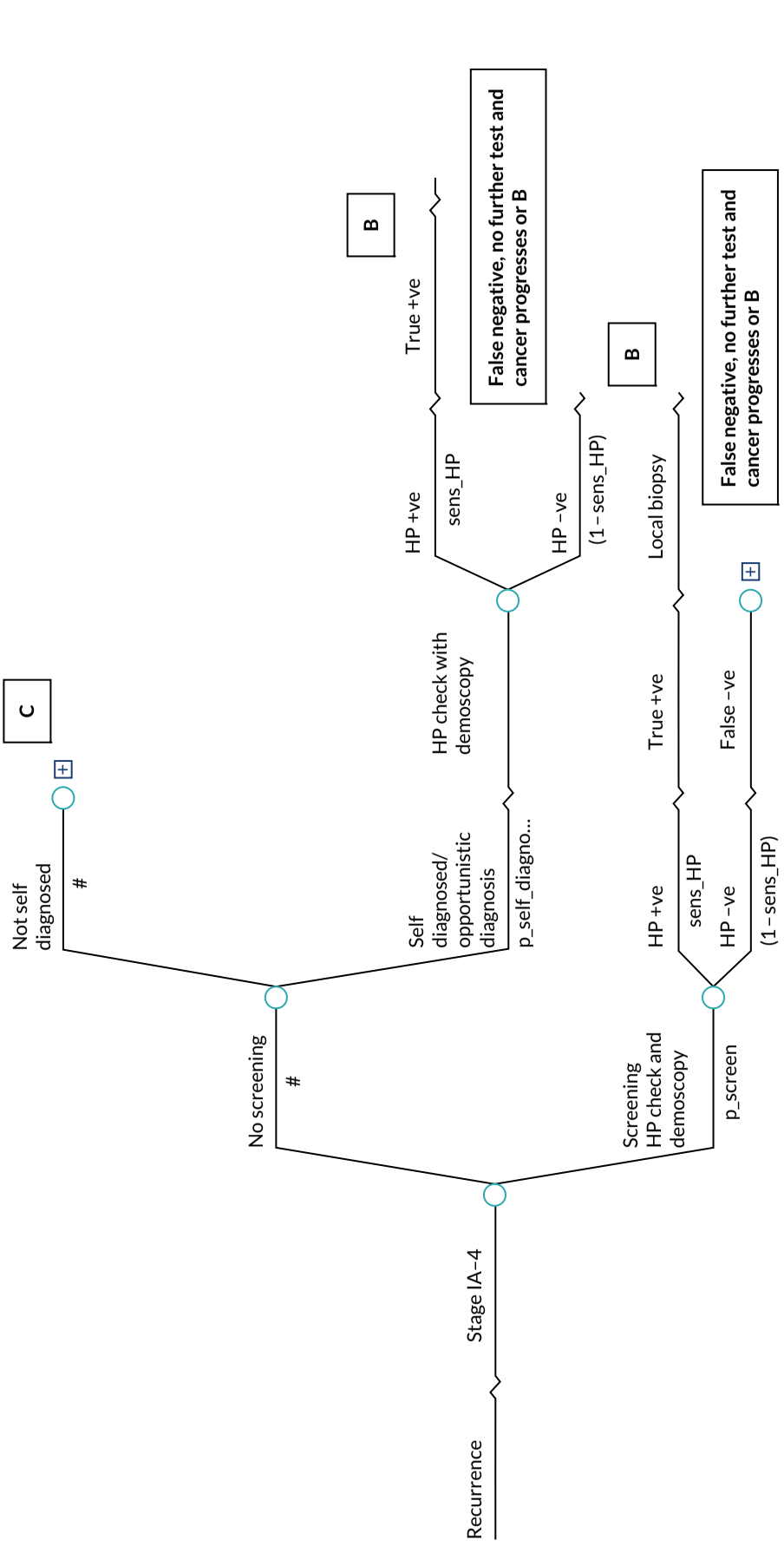


FIGURE 22 Recurrence health state. +ve, positive; –ve, negative; HP, health and physical examination; sens, sensitivity.

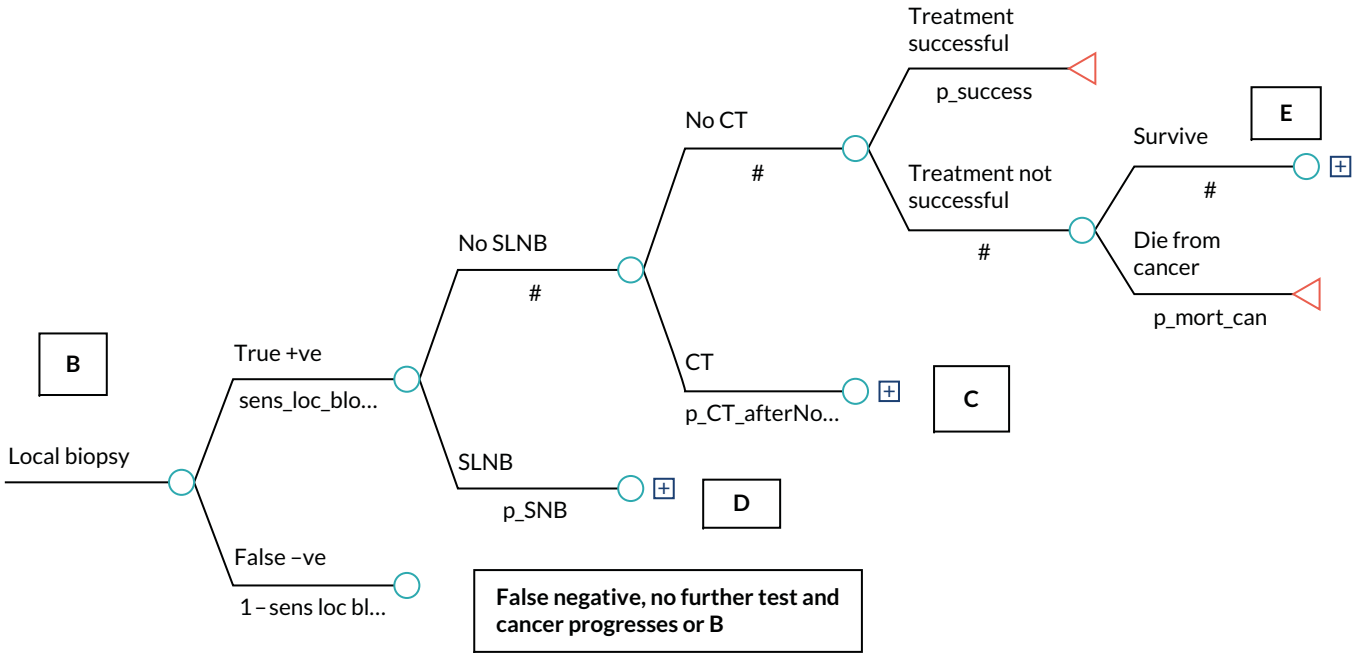


FIGURE 23 Clinical pathway following a local biopsy. +ve, positive; –ve, negative; mort, mortality; sens, sensitivity.

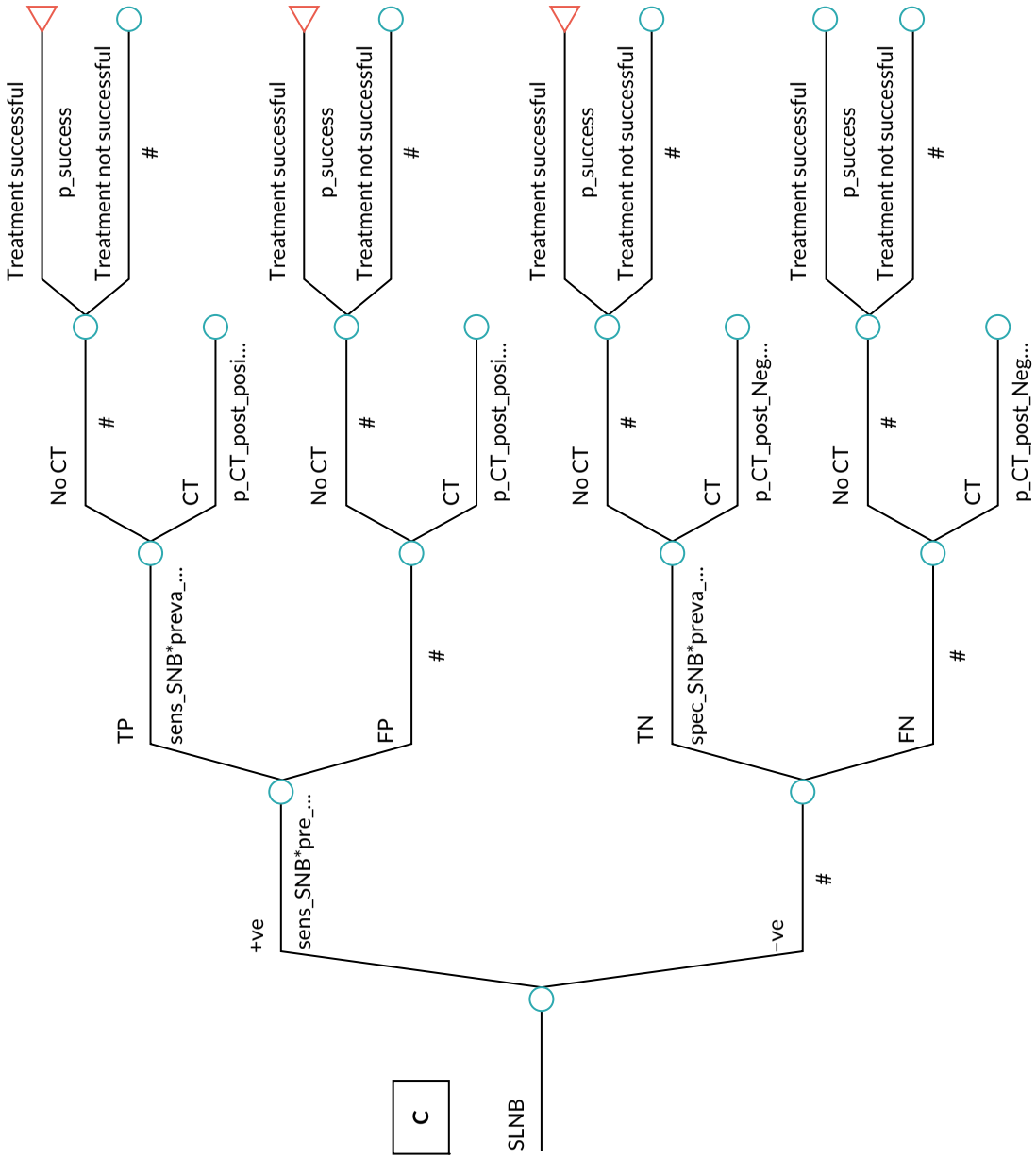


FIGURE 24 Care pathway following a SLNB. +ve, positive; –ve, negative; FN, false negative; FP, false positive; sens, sensitivity; spec, specificity; TN, true negative; TP, true positive.

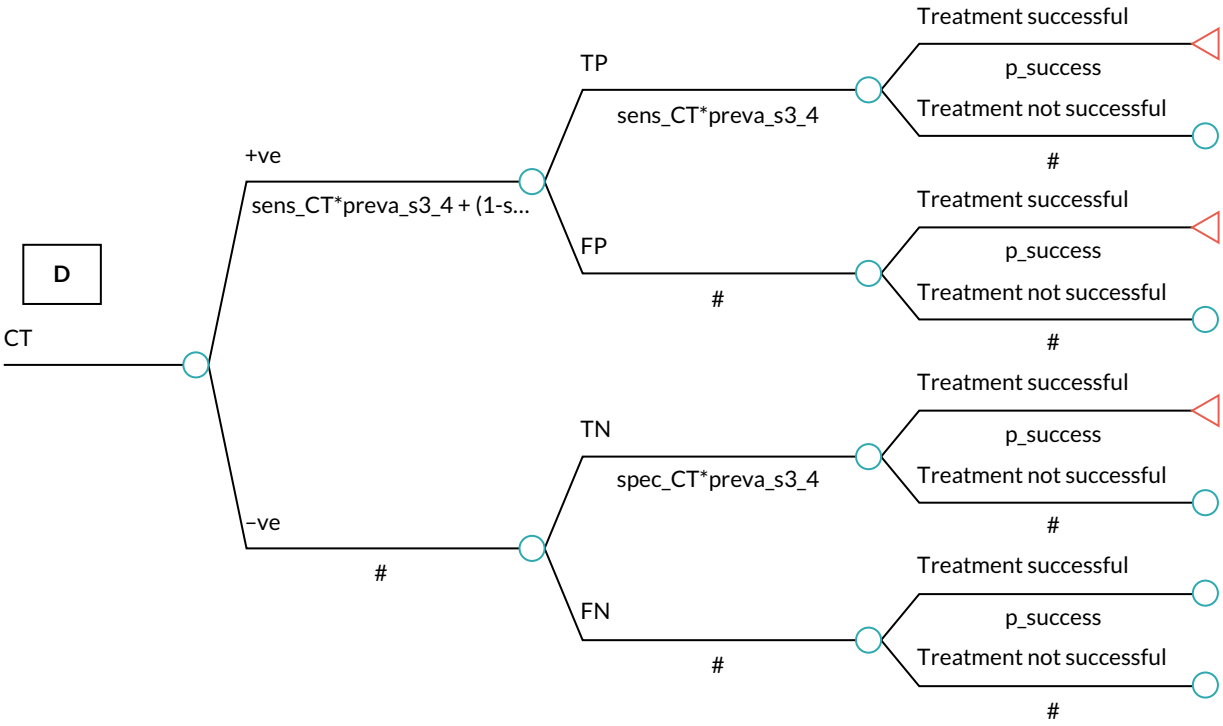


FIGURE 25 Care pathway following CT. +ve, positive; –ve, negative; FN, false negative; FP, false positive; sens, sensitivity; spec, specificity; TN, true negative; TP, true positive.

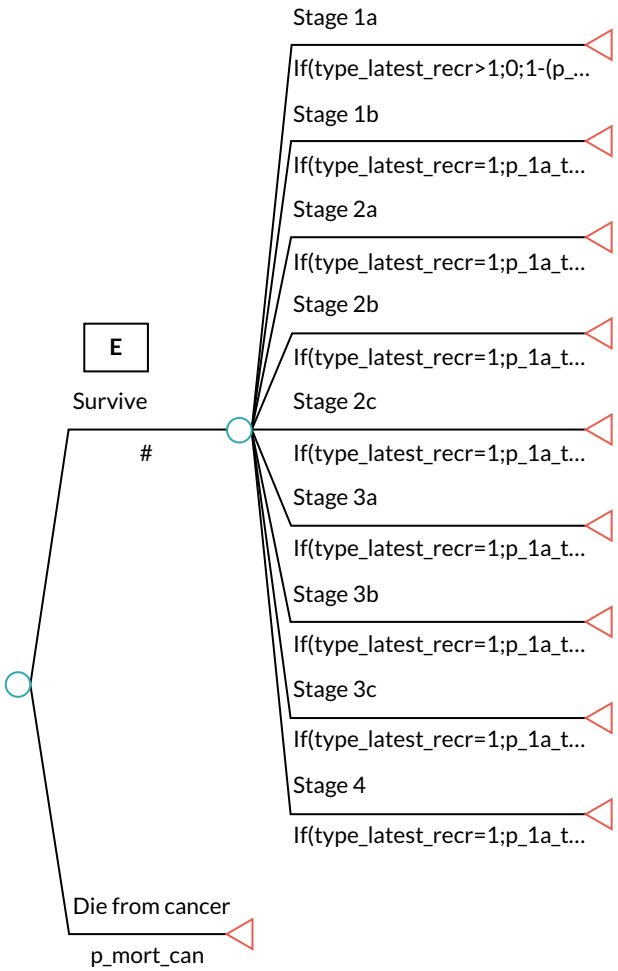


FIGURE 26 Disease pathway for disease progression.

Appendix 9 Clinical parameter values

The clinical parameters used in the model are presented in *Table 24* as fixed values without any distributions assigned.

Monthly probability values of p_false_self_diagnose and p_self_diagnosed were created from the TreeAge function probtoprob, which converts a probability into a rate, multiplies the rate by the given multiplier and converts this back to a probability.

Mortality values

The monthly probability values of mortality from melanoma are presented in *Table 25*. These values are based on calculations from Wilson *et al.*[182] The OR of survival was a function of disease stage at diagnosis. This model assumed that stage IA disease has no impact on overall survival, then the annual probability of death is calculated as the age/sex baseline rate for the general population, adjusted for the OR.

TABLE 24 List of clinical parameters used in the base-case economic model

Parameter	Description	Value	Source
p_CT_afterNo_SNB	Probability of receiving a CT scan after no SNB	0.9	Expert opinion
p_CT_post_Neg_SNB	Probability of receiving a CT scan after negative SNB	0.9	Expert opinion
p_CT_post_posi_SNB	Probability of receiving a CT scan after positive SNB	1	Expert opinion
p_CT_s4	Probability of receiving a CT scan after stage IV diagnosis	1	Expert opinion
p_false_self_diagnose	Probability of a false (diagnosis) after self-diagnosis	probtoprob(0.85; 1/12) 0.1462	Holterhues *et al.*[191] 2016 and expert opinion
p_go_visit_E_SD	Probability of seeking treatment emergently after self-diagnosis	1	Expert opinion
p_loc_biopsy_HP_positive	Probability of having a further test after HP positive	1	Expert opinion
p_no_test_post_negativeHP	Probability of no test after negative HP	1	Expert opinion
p_self_diagnosed	Probability of self-diagnosis of melanoma	probtoprob(0.6; 1/12) 0.073	Damude *et al.*[23] 2016 and expert opinion
p_SNB	Probability of SNB after wider excision	0.95	Expert opinion
p_SNB_self_FN	Probability of SNB after a false negative from a self-diagnosis	0.9	Expert opinion
p_success_s1	Probability of treatment success after stage I	0.95	Expert opinion
p_success_s2	Probability of treatment success after stage II	0.85	Expert opinion
p_success_s3	Probability of treatment success after stage III	0.6	Expert opinion
p_success_s4	Probability of treatment success after stage IV	0.4	Expert opinion

HP, health and physical examination; SNB, sentinel node biopsy.

TABLE 25 Mortality OR by melanoma stage with value in model

Stage	OR	Value in model (monthly)
IA	–	1.700×10^{-5}
IB	4.261	7.243×10^{-5}
IIA	12.250	2.082×10^{-4}
IIB	21.000	3.569×10^{-4}
IIC	41.741	7.091×10^{-4}
IIIA	13.821	2.349×10^{-4}
IIIB	32.667	5.551×10^{-4}
IIIC	67.667	0.0011
IV	312.104	0.0053

The UK age/sex background annual mortality probabilities are presented in *Table 26*. These values were converted to monthly probabilities using the TreeAge function probtoprob.

Recurrence values

Given that recurrence data are not collected at a registry level, the next best alternative would be to use melanoma incidence data from the NCRAS/PHE as a surrogate to find the recurrence probability in England/the UK. However, NCRAS/PHE data are based on the summary stage information (i.e. AJCC stage I–IV) of all patients diagnosed by CCGs in England, as presented in *Table 27*.

TABLE 26 The background age/sex probability of mortality by age/sex in the UK

Age (index)	Male	Female
0	0.004276	0.00349
1	2.71E–04	2.31E–04
2	1.54E–04	1.40E–04
3	1.20E–04	9.70E–05
4	1.02E–04	8.30E–05
5	9.20E–05	6.90E–05
6	7.80E–05	7.30E–05
7	8.30E–05	6.80E–05
8	7.00E–05	6.20E–05
9	7.80E–05	6.00E–05
10	7.80E–05	5.10E–05
11	9.10E–05	7.30E–05
12	9.50E–05	7.40E–05
13	1.06E–04	9.10E–05
14	1.18E–04	1.07E–04

TABLE 26 The background age/sex probability of mortality by age/sex in the UK (*continued*)

Age (index)	Male	Female
15	1.70E–04	1.26E–04
16	2.20E–04	1.51E–04
17	3.02E–04	1.53E–04
18	3.99E–04	2.09E–04
19	4.39E–04	2.00E–04
20	4.67E–04	2.04E–04
21	4.96E–04	2.11E–04
22	4.94E–04	2.04E–04
23	5.29E–04	2.26E–04
24	5.35E–04	2.15E–04
25	6.20E–04	2.42E–04
26	5.73E–04	2.68E–04
27	6.14E–04	2.77E–04
28	6.79E–04	3.09E–04
29	6.84E–04	3.32E–04
30	7.25E–04	3.75E–04
31	7.69E–04	4.01E–04
32	9.27E–04	4.79E–04
33	8.88E–04	5.00E–04
34	9.77E–04	5.08E–04
35	0.001026	5.88E–04
36	0.001161	6.60E–04
37	0.001191	7.30E–04
38	0.001241	7.30E–04
39	0.001393	8.41E–04
40	0.001501	9.30E–04
41	0.001713	9.76E–04
42	0.001811	0.001076
43	0.001987	0.001144
44	0.002096	0.001294
45	0.002202	0.001441
46	0.00237	0.001528
47	0.002741	0.001655
48	0.002797	0.001776
49	0.003107	0.001919
50	0.003402	0.002142
51	0.003501	0.002349

continued

TABLE 26 The background age/sex probability of mortality by age/sex in the UK (*continued*)

Age (index)	Male	Female
52	0.003813	0.002576
53	0.003968	0.002807
54	0.004408	0.003014
55	0.004923	0.00329
56	0.005467	0.003611
57	0.005868	0.003861
58	0.006371	0.004228
59	0.007031	0.004764
60	0.007955	0.005166
61	0.008614	0.00562
62	0.009324	0.006282
63	0.010484	0.006855
64	0.011447	0.007317
65	0.012244	0.007878
66	0.013496	0.008883
67	0.014599	0.009608
68	0.015607	0.010351
69	0.017228	0.011515
70	0.018837	0.01267
71	0.021307	0.01424
72	0.023546	0.015891
73	0.026015	0.017933
74	0.029585	0.019667
75	0.032946	0.022235
76	0.036425	0.025461
77	0.03986	0.027457
78	0.044087	0.03139
79	0.049207	0.03462
80	0.055222	0.039684
81	0.06139	0.044619
82	0.069292	0.051081
83	0.078533	0.058425
84	0.087689	0.06669
85	0.098188	0.074956
86	0.109512	0.08545
87	0.122389	0.096451
88	0.137063	0.109963
89	0.151054	0.123634

TABLE 26 The background age/sex probability of mortality by age/sex in the UK (*continued*)

Age (index)	Male	Female
90	0.168744	0.139389
91	0.186002	0.154579
92	0.201643	0.171622
93	0.220694	0.188373
94	0.243034	0.209568
95	0.26757	0.233693
96	0.28755	0.253124
97	0.305291	0.268228
98	0.313561	0.283968
99	0.348595	0.31482
100	0.385655	0.340913

TABLE 27 Incident cases (*n*) of melanoma (2012–14) based on NCRAS data

Stage					
1	2	3	4	'X' unknown	Total
21,737	6068	1764	794	6560	36,923

Stage 'X' indicates that the full TNM stage group is unknown to the NCRAS (but not necessarily to the treating clinicians). Among other reasons, this may be because of a data quality problem (missing data), because of patient mortality before staging was completed or because it was clinically inappropriate to fully stage the patient. *Table 28* presents the known melanoma diagnosis.

A large registry-based German study of 33,384 patients recorded by the Central Malignant Melanoma Registry of the German Society of Dermatology between 1976 and 2007 with stage I–III initial diagnosis is presented in *Table 29*.[209] These values are then scaled to 100% to include stage IV from NCRAS. Unfortunately, this paper provided rate of recurrence-free survival by summary stage and did not provide information to calculate the follow-up recurrence probability over time per AJCC staging criteria.

Recurrence probabilities over time for AJCC stages IA–IIC were obtained from a different study conducted in Australia by Turner *et al.*[210] The study authors analysed 2298 patient records for the development of recurrence and new primary melanoma up to 10 years. Kaplan–Meier curves in this paper showing time to recurrence for localised melanoma were able to be digitised using WebPlotDigitizer. Various points along the curve were chosen and the co-ordinates of those points were extracted.

TABLE 28 Incident cases of known melanoma diagnosis (2012–14)

Stage, *n* (%)				
1	2	3	4	Total (*n*)
21,737 (72)	6068 (20)	1764 (6)	794 (3)	30,363

TABLE 29 Stage at primary diagnosis

Stage at primary diagnosis	CMMR ($n = 33,384$) (%)	Total per summary stage (%)	Substage proportions (%)	Scaled to 100% (%)
IA	45.3	I = 71.4	45.3/71.4 = 63.4	43.9
IB	26.1		26.1/71.4 = 36.5	25.3
IIA	12.9	II = 23.5	12.9/23.5 = 54.9	12.5
IIB	7.9		7.9/23.5 = 33.6	7.7
IIC	2.7		2.7/23.5 = 11.5	2.6
IIIA	1.0	III = 5.0	1.0/5.0 = 20.0	1.0
IIIB	3.6		3.6/5.0 = 72.0	3.5
IIIC	0.4		0.4/5.0 = 8.0	0.4
IV (NCRAS)	3.0			3.0

CMMR, Central Malignant Melanoma Registry.

These points were then used to calculate the lamda and gamma parameters of the Weibull distribution for AJCC stage IA. The two parameters were then used to calculate the transition probabilities (*tp*) for recurrence using *Equation 4*:

$$tp(t\mu) = 1 - \exp[\lambda(t-\mu)^{\gamma} - \lambda t^{\gamma}], \tag{4}$$

where:

- t is time (measured in terms of the number of cycles; each cycle is equivalent to 1 month)
- λ is the scale parameter, which describes the probability that an individual experiences recurrence, given that he/she is recurrence free during the current time period
- γ is the shape parameter, which describes the hazard function of Weibull function for the survival time
- μ is the length of the Markov cycle.

Moreover, for the calculation of the baseline transition probability, *Equation 5* was used:

$$tp(t\mu) = 1 - \frac{s(t)}{s(t-\mu)}, \tag{5}$$

where μ is the length of the Markov cycle.

Recurrence rates for the remaining melanoma stages were computed as a function of the probability of recurrence of stage IA disease and the distribution of the hazard ratio of each stage up to stage IIC reported in the study.

By using the CIs presented in the paper by Turner *et al.*,[210] the corresponding SEs were calculated (*Table 30*), which, along with the hazard ratios, were used as gamma distributions in the model.

Because it was not possible to source recurrence rates for AJCC stages III and IV, it was assumed that the recurrence rates for these two stages are similar to those of stage IIC. Depending on the stage of the recurrence in the model, the monthly probabilities were selected (*Table 31*).

TABLE 30 Hazard ratios for 5 and 10 years of follow-up

	Follow-up					
	5 years			10 years		
Stage	HR	95% CI	SE	HR	95% CI	SE
IA	1.00			1.00		
IB	2.02	1.54 to 2.65	0.28	2.10	1.15 to 3.86	0.69
IIA	4.32	3.22 to 5.81	0.66	2.43	1.13 to 5.25	1.05
IIB	6.10	4.54 to 8.20	0.93	2.98	1.33 to 6.67	1.36
IIC	7.09	5.15 to 9.76	1.18	3.95	1.59 to 9.84	2.10
IIIA	7.09	5.15 to 9.76	1.18	3.95	1.59 to 9.84	2.10
IIIB	7.09	5.15 to 9.76	1.18	3.95	1.59 to 9.84	2.10
IIIC	7.09	5.15 to 9.76	1.18	3.95	1.59 to 9.84	2.10
IV	7.09	5.15 to 9.76	1.18	3.95	1.59 to 9.84	2.10

Based on data from Turner *et al.*,[210] table 2. Multivariate Cox Proportional Hazard Models for Recurrence or New Primary within first ten years of Follow-Up in 2996 Patients with Localized Melanoma.

TABLE 31 Monthly probability in the model

Stage	Variable name	Probability of recurrence or new primary
IA	p_recur_s1a	0.0022
IB	p_recur_s1b = p_recur_s1a*dist_hr_recu_1b	0.0046
IIA	p_recur_s2a = p_recur_s1a*dist_hr_recu_2a	0.0095
IIB	p_recur_s2b = p_recur_s1a*dist_hr_recu_2b	0.0134
IIC	p_recur_s2c = p_recur_s1a*dist_hr_recu_2c	0.0156
IIIA	p_recur_s3a = p_recur_s1a*dist_hr_recu_2c	0.0156
IIIB	p_recur_s3b = p_recur_s1a*dist_hr_recu_2c	0.0156
IIIC	p_recur_s3c = p_recur_s1a*dist_hr_recu_2c	0.0156
IV	p_recur_4 = p_recur_s1a*dist_hr_recu_2c	0.0156

Diagnostic accuracy

The diagnostic accuracy statistics used in the model are based on the pooled statistics (meta-analysis) taken from a Cochrane systematic review.[170] It was assumed that the diagnostic accuracy of the health-care professional should be based on visual inspection, plus the use of dermoscopy on real patients. The studies of those physicians deemed to have 'high' experience of invasive melanoma or atypical intraepidermal melanocytic variants were selected. The studies identified were then retrieved and physicians classified as either 'surgical oncologists' or 'dermatologists' (*Tables 32* and *33*). For dermatological nurse specialists, only one study in the Cochrane review was from the USA and it was based on 'physician assistant' (*Table 34*). A beta distribution of the sensitivity and specificity was then used in the base-case PSA.

In the clinical pathway, once a health-care professional deems the mole/skin to be suspicious, a local biopsy is taken and sent to a pathologist for confirmation. The sensitivity and specificity of local biopsy

TABLE 32 Sensitivity and specificity of surgical oncologists used in the model

Country	Study (first author and year of publication)	True positive (*n*)	False positive (*n*)	False negative (*n*)	True negative (*n*)	Sensitivity (95% CI)	Specificity (95% CI)
Italy	Bono *et al.*[206] 2002	60	63	6	184	0.91 (0.81 to 0.97)	0.74 (0.69 to 0.80)
Italy	Bono *et al.*[205] 2002	10	42	3	106	0.77 (0.46 to 1.95)	0.72 (0.64 to 0.79)
Pooled						0.886 (0.795 to 0.94); SE 0.036	0.734 (0.688 to 0.775); SE 0.022

TABLE 33 Sensitivity and specificity of dermatologists used in the model

Country	Study (first author and year of publication)	True positive (*n*)	False positive (*n*)	False negative (*n*)	True negative (*n*)	Sensitivity (95% CI)	Specificity (95% CI)
Sweden	Ahnlide *et al.*[194] 2016	34	23	12	240	0.74 (0.59 to 0.86)	0.91 (0.87 to 0.94)
Italy	Bauer *et al.*[195] 2000	33	10	9	263	0.79 (0.63 to 0.90)	0.96 (0.93 to 0.98)
Italy	Carli *et al.*[197] 1994	5	28	0	35	1.00 (0.48 to 1.00)	0.56 (0.42 to 0.68)
Italy	Carli *et al.*[196] 2002	53	9	1	193	0.98 (0.90 to 1.00)	0.96 (0.92 to 0.98)
Austria	Dreiseitl *et al.*[198] 2009	26	121	1	310	0.96 (0.81 to 1.00)	0.72 (0.67 to 0.76)
Turkey	Gokdemir *et al.*[199] 2011	12	25	1	410	0.92 (0.64 to 1.00)	0.94 (0.92 to 0.96)
Italy (Modena)	Guitera *et al.*[204] 2009	68	83	11	33	0.86 (0.76 to 0.93)	0.28 (0.20 to 0.38)
Germany	Haenssle *et al.*[200] 2010	32	146	8	8263	0.80 (0.64 to 0.91)	0.98 (0.98 to 0.99)
Germany	Haenssle *et al.*[237] 2010	47	228	40	2373	0.54 (0.43 to 0.65)	0.91 (0.90 to 0.92)
Spain	Morales-Callaghan *et al.*[201] 2008	4	6	2	188	0.67 (0.22 to 0.96)	0.97 (0.93 to 0.99)
Germany and USA	Nachbar *et al.*[202] 1994	64	11	5	114	0.93 (0.84 to 0.98)	0.91 (0.85 to 0.96)
Austria	Soyer *et al.*[203] 1995	61	17	4	77	0.94 (0.85 to 0.98)	0.82 (0.73 to 0.89)
Pooled						0.875 (0.784 to 0.931); SE 0.037	0.893 (0.792 to 0.949); SE 0.038

TABLE 34 Sensitivity and specificity of dermatological specialist[a] used in the model

Country	Study (first author and year of publication)	True positive (*n*)	False positive (*n*)	False negative (*n*)	True negative (*n*)	Sensitivity (95% CI)	Specificity (95% CI)
USA	Ferris *et al.*[207] 2015	20	21	5	19	0.80 (0.59 to 0.93); SE 0.086	0.47 (0.32 to 0.64); SE 0.081

a Dermatological specialist nurse based on 'physician assistant' with dermoscopy on images.

were derived from a study that aimed to investigate how accurate and reproducible the results are of pathologists' diagnoses of melanocytic skin lesions.[208] The results of the paper indicate that 82.8% (95% CI 81.0% to 84.5%) of melanocytic skin biopsy diagnoses would have their diagnosis verified if reviewed by a consensus reference panel of experienced pathologists. In the model, the sensitivity and specificity were assumed based on this study (*Table 35*).

For regional disease staging, the accompanying systematic review of the clinical evidence for the NICE guideline was the source of data for the staging of melanoma.[16] The sensitivity of SLNB in identifying micrometastatic nodal/regional disease for patients was estimated to be 86.6% (95% CI 84.6% to 88.4%), based on 47 studies with 19,607 data points. Specificity was 100%, as in the review (*Table 36*).

For advanced staging, further diagnostic tests were used, including ultrasonography, CT, PET and a combination of both (PET-CT). The 2015 NICE guideline[16] recommended that CT staging be offered to people with stage III or suspected stage IV melanoma. According to a meta-analysis[147] of staging of distant metastasis, median sensitivity of CT scan was 51% (95% CrI 24% to 76%) and specificity was 69% (95% CrI 30% to 92%). These median estimates, along with the corresponding CrIs, were used in the model as beta distributions.

Health-state utilities

An initial search was conducted (in September 2018) in PubMed, Tufts Cost-Effectiveness Analysis Registry[238] and other relevant sources (e.g. ScHARRHUD, the HERC database of mapping studies[239] and PROSPERO[240]). However, once the most appropriate systematic review and meta-analysis relevant paper was identified, the search was truncated.[216]

Search filters developed and maintained by the Canadian Agency for Drugs and Technologies in Health Information Services Filters Working Group were used.[241]

PubMed

Health utilities/quality of life

"Value of Life" [mh] OR Quality of Life[mh] OR quality of life[tiab] OR Quality-Adjusted Life years[mh] OR quality adjusted life[tiab] OR qaly*[tiab] OR qald*[tiab] OR qale*[tiab] OR qtime*[tiab] OR life year [tiab] OR life years[tiab] OR disability adjusted life[tiab] OR daly*[tiab] OR sf36[tiab] OR sf 36[tiab] OR short form 36[tiab] OR shortform 36[tiab] OR short form36[tiab] OR shortform36[tiab] OR sf6

TABLE 35 Sensitivity and specificity of local biopsy

Local biopsy	Mean	95% CI; SE
Sensitivity (assumed)	0.828	0.810 to 0.845; SE 0.0089
Specificity (assumed)	0.50	SE 0.05

TABLE 36 Sensitivity and specificity of SLNB

Stage	*n* studies (*n* data points)	Prevalence (%)	Sensitivity (95% CI) (%)	Specificity (%)
Any	47 (19,607)	9–41	86.6 (84.6 to 88.4)	100

[tiab] OR sf 6[tiab] OR short form 6[tiab] OR sf6d[tiab] OR sf 6d[tiab] OR short form 6d[tiab] OR sf8[tiab] OR sf 8[tiab] OR short form 8[tiab] OR sf12[tiab] OR sf 12[tiab] OR short form 12[tiab] OR sf16[tiab] OR sf 16[tiab] OR sf20[tiab] OR sf 20[tiab] OR short form 20[tiab] OR hql[tiab] OR hqol[tiab] OR h qol[tiab] OR hrqol[tiab] OR hr qol[tiab] OR hye[tiab] OR hyes[tiab] OR healthy year equivalent* [tiab] OR healthy years equivalent*[tiab] OR pqol[tiab] OR qls[tiab] OR quality of well being[tiab] OR index of wellbeing[tiab] OR qwb[tiab] OR nottingham health profile*[tiab] OR sickness impact profile [tiab] OR health status indicators[mh] OR health utilit*[tiab] OR health status[tiab] OR disutilit*[tiab] OR rosser[tiab] OR willingness to pay[tiab] OR standard gamble*[tiab] OR time trade off[tiab] OR time tradeoff[tiab] OR tto[tiab] OR hui[tiab] OR hui1[tiab] OR hui2[tiab] OR hui3[tiab] OR eq[tiab] OR euroqol[tiab] OR euro qol[tiab] OR eq5d[tiab] OR eq 5d[tiab] OR euroqual[tiab] OR euro qual[tiab] OR duke health profile[tiab] OR functional status questionnaire[tiab] OR dartmouth coop functional health assessment*[tiab] OR (utilit*[tiab] AND (valu*[tiab] OR measur*[tiab] OR health[tiab] OR life[tiab] OR estimat*[tiab] OR elicit*[tiab] OR disease[tiab] OR score*[tiab] OR weight[tiab])) OR (preference*[tiab] AND (valu*[tiab] OR measur*[tiab] OR health[tiab] OR life[tiab] OR estimat*[tiab] OR elicit*[tiab] OR disease[tiab] OR score*[tiab] OR instrument[tiab] OR instruments[tiab])).

AND

Disease

Melanoma OR skin tumor OR skin tumour OR cutaneous tumor OR cutaneous tumour OR skin cancer.

Using this search strategy, > 7000 hits were recorded (*Table 37*). However, once the most appropriate systematic review and meta-analysis relevant paper was identified, the search was truncated.[216]

Treatment costs

In all cases, patients undergo biopsy excision (at which point the disease is staged according to AJCC guidelines), followed by definitive surgery known as wide local excision. Patients with stage IA or stage IB disease undergo no further treatment. Patients with stage IIA or higher disease undergo SLNB and patients with stage IIB or higher disease undergo CT. Patients with a positive SLNB undergo follow-up surgery for lymph node involvement, comprising pre-operative CT and radical lymph node basin dissection. Patients with stage III disease receive systemic therapy. Patients with stage IV disease undergo surgery for removal of localised metastases and combination immunotherapy (*Table 38*). Gamma distributions were assigned to cost values for stage III and IV treatments.

TABLE 37 Search for utility values in PubMed

Search facet	Results (*n*)
Health utilities/quality of life	806,807
Disease	195,615
Final numbers	7676

TABLE 38 Costs included in the pathway

Initial treatment	Stage								
	IA	IB	IIA	IIB	IIC	IIIA	IIIB	IIIC	IV
Biopsy excision	✓	✓	✓	✓	✓	✓	✓	✓	✓
Definitive surgery (wide local excision)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Investigations									
CT				✓	✓	✓	✓	✓	✓
SLNB (carried out at same time as definitive surgery)			✓	✓	✓	✓	✓	✓	✓
Follow-up surgery for positive lymph nodes									
Pre-operative CT						✓	✓	✓	
Radical lymph node dissection						✓	✓	✓	
Metastatic disease									
Surgical removal of localised metastases									✓
Targeted therapy (combination of dabrafenib + trametinib)						✓	✓	✓	
Immunotherapy (combination of ipilimumab + nivolumab)									✓

Appendix 10 Sensitivity analyses: detailed results

Self-diagnosis is an important parameter to capture in the model. In one-way sensitivity analysis, with all values at their base case, there is little difference in terms of NMB; this is because of the low probability of recurrence. At the extreme, where no patient in the model self-diagnosed, strategies 22 (stage IA) and 77 (stage IB) have the highest NMB. When every recurrence is detected by self-diagnosis, there is no benefit in surveillance, as can be seen by the high ICERs. *Tables 39* and *40* report the results of the one-way sensitivity analyses of self-diagnosis for stages IA and IB, respectively.

TABLE 39 Results of the one-way sensitivity analysis of self-diagnosis for stage IA

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
Probability of self-diagnosis per month = 0						
22	9895		21.78			425,637
15	10,133	238	21.78	0.00	244,139	425,418
23	10,456	323	21.78	0.00	73,394	425,184
1 (NICE)	10,590	135	21.78	0.00	Dominated	424,990
19	10,681	226	21.78	0.00	355,202	424,971
16	11,166	485	21.79	0.02	172,490	424,542
21	12,452	1286	21.80	0.00	68,455	423,632
14	12,660	208	21.80	0.00	1,586,856	423,426
7	12,867	207	21.80	0.00	1,872,254	423,221
4	14,607	1740	21.81	0.00	280,098	421,605
Probability of self-diagnosis per month = 0.35						
22	12,611	0	21.89	0.00		425,144
15	12,811	199	21.89	0.00	49,462,070	424,944
23	12,977	166	21.89	0.00	26,478,336	424,778
19	13,176	199	21.89	0.00	5,2041,509	424,579
1 (NICE)	13,210	33	21.89	0.00	Dominated	424,546
16	13,555	379	21.89	0.00	44,940,623	424,201
21	14,273	718	21.89	0.00	31,383,914	423,483
14	14,472	199	21.89	0.00	60,603,316	423,284
7	14,671	199	21.89	0.00	66,783,439	423,085
4	16,186	1515	21.89	0.00	48,874,025	421,570

continued

TABLE 39 Results of the one-way sensitivity analysis of self-diagnosis for stage IA (*continued*)

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
Probability of self-diagnosis per month = 0.70						
22	12,687		21.89			425,100
15	12,886	199	21.89	0.00	256,032,538	424,902
23	13,052	166	21.89	0.00	348,679,760	424,736
19	13,250	199	21.89	0.00	257,377,027	424,537
1 (NICE)	13,283	33	21.89	0.00	209,731,828	424,504
16	13,628	345	21.89	0.00	310,586,952	424,160
21	14,345	717	21.89	0.00	290,721,349	423,442
14	14,544	199	21.89	0.00	280,009,293	423,244
7	14,743	199	21.89	0.00	289,522,912	423,045
4	16,254	1511	21.89	0.00	323,644,663	421,534

TABLE 40 Results of the one-way sensitivity analysis of self-diagnosis for stage IB

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
Probability of self-diagnosis per month = 0						
77	9197	0	21.30	0.00	0	416,868
86	9473	276	21.31	0.01	54,335	416,693
80	9564	91	21.32	0.01	11,511	416,761
83	10,130	566	21.33	0.01	48,071	416,431
23	10,700	570	21.36	0.04	15,888	416,578
15	10,784	84	21.35	−0.01	Dominated	416,200
8	11,529	829	21.36	0.00	Dominated	415,684
2	12,217	1518	21.37	0.00	357,121	415,145
9	12,678	461	21.40	0.03	16,535	415,242
5	12,812	134	21.39	−0.01	Dominated	414,951
1 (NICE)	13,312	634	21.40	0.00	193,317	414,674
25	14,770	1458	21.58	0.18	7967	416,876
18	18,185	3415	21.62	0.03	102,892	414,124
11	18,795	610	21.62	0.00	1,224,590	413,524
4	19,395	600	21.62	0.00	1,748,100	412,931

TABLE 40 Results of the one-way sensitivity analysis of self-diagnosis for stage IB (*continued*)

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
Probability of self-diagnosis per month = 0.35						
77	14,535		21.87			422,792
80	14,721	186	21.87	0.00	4,184,463	422,606
86	14,735	14	21.87	0.00	Dominated	422,592
83	15,114	393	21.87	0.00	5,512,013	422,215
23	15,242	128	21.87	0.00	2,579,499	422,088
15	15,481	238	21.87	0.00	9,204,223	421,850
8	16,060	579	21.87	0.00	6,814,888	421,272
2	16,639	579	21.87	0.00	8,145,629	420,695
9	16,755	116	21.87	0.00	1,908,787	420,580
5	16,980	225	21.87	0.00	21,786,488	420,355
25	17,166	186	21.87	0.00	533,666	420,176
1 (NICE)	17,334	168	21.87	0.00	Dominated	420,002
18	20,094	2928	21.87	0.00	4,847,788	417,260
11	20,673	579	21.87	0.00	7,317,353	416,683
4	21,252	579	21.87	0.00	8,725,665	416,105
Probability of self-diagnosis per month = 0.70						
77	14,718		21.88			422,861
80	14,903	185	21.88	0.00	33,226,479	422,677
86	14,916	13	21.88	0.00	12,594,062	422,663
83	15,292	376	21.88	0.00	32,684,368	422,287
23	15,421	129	21.88	0.00	70,535,481	422,159
15	15,656	235	21.88	0.00	28,997,710	421,924
8	16,231	575	21.88	0.00	35,696,084	421,349
2	16,807	575	21.88	0.00	38,065,537	420,774
9	16,923	116	21.88	0.00	40,943,451	420,658
5	17,146	223	21.88	0.00	35,271,500	420,434
25	17,339	193	21.88	0.00	7,692,907	420,242
1 (NICE)	17,498	159	21.88	0.00	Dominated	420,083
18	20,252	2913	21.88	0.00	31,305,095	417,331
11	20,827	575	21.88	0.00	35,954,371	416,757
4	21,402	575	21.88	0.00	38,343,472	416,182

Recurrence is the driving parameter that needs to be captured in the model. In one-way sensitivity analysis, with all values set at their base-case value, it can be seen that strategies 22 (stage IA) and 77 (stage IB) have the highest NMB (*Tables 41* and *42*).

TABLE 41 Probability of recurrence for stage IA

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
Probability of recurrence per month = 0.00						
22	10,664		21.89			427,152
15	10,863	199	21.89	0.00	Dominated	426,953
23	11,029	365	21.89	0.00	Dominated	426,787
19	11,228	565	21.89	0.00	Dominated	426,588
1 (NICE)	11,261	598	21.89	0.00	Dominated	426,555
16	11,607	943	21.89	0.00	Dominated	426,209
21	12,325	1661	21.89	0.00	Dominated	425,491
14	12,524	1860	21.89	0.00	Dominated	425,292
7	12,724	2060	21.89	0.00	Dominated	425,093
4	14,237	3573	21.89	0.00	Dominated	423,579
Probability of recurrence per month = 0.14						
22	85,732		21.09			336,109
15	87,982	729	21.09	0.00	Dominated	333,864
23	93,327	3777	21.09	0.00	Dominated	330,148
1 (NICE)	93,346	20	21.09	0.00	Dominated	330,132
19	95,599	2273	21.09	0.00	Dominated	327,919
16	101,055	5456	21.09	0.00	Dominated	322,543
21	113,879	12,824	21.10	0.00	5,724,843	309,974
14	115,855	1976	21.10	0.00	8,807,771	308,033
7	117,642	1787	21.10	0.00	Dominated	306,275
4	139,469	21,827	21.10	0.00	7,576,950	284,776
Probability of recurrence per month = 0.29						
22	90,527		21.06			330,643
15	93,823	2127	21.06	0.00	Dominated	327,352
23	97,222	2176	21.06	0.00	Dominated	323,958
1 (NICE)	99,565	1093	21.06	0.00	Dominated	321,617
19	100,517	2045	21.06	0.00	Dominated	320,667
16	107,163	2298	21.06	0.00	Dominated	314,030
21	120,590	12,033	21.06	0.00	11,407,236	300,633
14	123,570	1433	21.06	0.00	Dominated	297,655
7	126,330	1148	21.06	0.00	Dominated	294,900
4	152,887	24,879	21.06	0.00	14,663,184	268,386
Probability of recurrence per month = 0.86						
22	93,856	0	21.03			326,837
15	97,741	2967	21.03	0.00	Dominated	322,953
23	100,982	2288	21.03	0.00	Dominated	319,715
19	104,867	2905	21.03	0.00	Dominated	315,831
1 (NICE)	105,506	3543	21.03	0.00	Dominated	315,193
16	112,244	2443	21.04	0.00	Dominated	308,458

TABLE 41 Probability of recurrence for stage IA (*continued*)

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
21	126,228	12,904	21.04	0.00	31,148,004	294,486
14	130,113	2683	21.04	0.00	Dominated	290,602
7	133,992	2643	21.04	0.00	Dominated	286,725
4	163,485	28,223	21.04	0.00	38,672,722	257,250
Probability of recurrence per month = 1.0						
22	94,101		21.03			326,556
15	98,004	3002	21.03	0.00	Dominated	322,655
23	101,258	2321	21.03	0.00	Dominated	319,403
19	105,161	2941	21.03	0.00	Dominated	315,502
1 (NICE)	105,809	3590	21.03	0.00	Dominated	314,854
16	112,570	2479	21.03	0.00	Dominated	308,096
21	126,614	12,987	21.03	0.00	35,499,571	294,062
14	130,517	2725	21.03	0.00	Dominated	290,161
7	134,419	2692	21.03	0.00	Dominated	286,260
4	164,041	28,377	21.03	0.00	43,776,631	256,655

TABLE 42 Probability of recurrence for stage IB

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
Probability of recurrence per month = 0.00						
77	10,664		21.89			427,152
80	10,850	186	21.89	0.00	Dominated	426,966
86	10,863	199	21.89	0.00	Dominated	426,953
83	11,241	577	21.89	0.00	Dominated	426,575
23	11,370	706	21.89	0.00	Dominated	426,446
15	11,607	943	21.89	0.00	Dominated	426,209
8	12,184	1520	21.89	0.00	Dominated	425,632
2	12,762	2098	21.89	0.00	Dominated	425,055
9	12,878	2214	21.89	0.00	Dominated	424,938
5	13,103	2439	21.89	0.00	Dominated	424,714
25	13,293	2630	21.89	0.00	Dominated	424,523
1 (NICE)	13,456	2792	21.89	0.00	Dominated	424,361
18	16,215	5551	21.89	0.00	Dominated	421,601
11	16,792	6128	21.89	0.00	Dominated	421,024
4	17,370	6706	21.89	0.00	Dominated	420,446

continued

TABLE 42 Probability of recurrence for stage IB (*continued*)

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
Probability of recurrence per month = 0.29						
77	83,849		19.45			305,105
86	87,146	2208	19.45	0.00	Dominated	301,823
80	87,253	2314	19.45	0.00	Dominated	301,718
83	93,929	5546	19.45	0.00	10,154,431	295,072
23	96,771	1610	19.45	0.00	22,053,492	292,253
15	100,479	3708	19.45	0.00	7,168,015	288,555
8	109,631	9152	19.45	0.01	4,262,455	279,446
2	118,106	8475	19.46	0.00	4,264,035	271,011
9	122,044	3938	19.46	0.00	2,612,543	267,103
5	124,337	2293	19.46	0.00	7,351,939	264,816
1 (NICE)	130,510	6173	19.46	0.00	3,689,587	258,677
25	131,498	988	19.47	0.01	89,416	257,910
18	182,194	50,697	19.49	0.02	2,364,881	207,642
11	191,345	9151	19.49	0.00	4,258,565	198,535
4	199,818	8473	19.50	0.01	4,260,292	190,102
Probability of recurrence per month = 0.43						
77	85,331		19.42			303,055
80	88,845	2542	19.42	0.00	Dominated	299,552
86	89,018	2715	19.42	0.00	Dominated	299,379
83	96,132	6279	19.42	0.00	13,633,220	292,289
23	98,669	1448	19.42	0.00	Dominated	289,767
15	102,998	5778	19.42	0.00	13,211,340	285,446
8	113,330	10,332	19.42	0.00	6,288,715	275,147
2	123,180	9850	19.43	0.01	6,261,490	265,328
9	126,352	3172	19.43	0.00	3,522,177	262,174
5	129,613	3261	19.43	0.00	8,854,734	258,920
25	134,481	4867	19.44	0.01	580,263	254,221
1 (NICE)	136,201	1721	19.43	−0.01	Dominated	252,356
18	188,584	54,104	19.45	0.02	3,527,071	200,424
11	198,923	10,338	19.45	0.00	6,283,914	190,119
4	208,778	9855	19.45	0.00	6,256,797	180,295
Probability of recurrence per month = 0.86						
77	86,839		19.39			300,966
80	90,460	2770	19.39	0.00	Dominated	297,351
86	90,725	3034	19.39	0.00	Dominated	297,087
83	98,096	6486	19.39	0.00	24,721,481	289,728

TABLE 42 Probability of recurrence for stage IB (*continued*)

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
23	100,582	1536	19.39	0.00	Dominated	287,250
15	105,210	6164	19.39	0.00	24,988,057	282,627
8	116,463	11,253	19.39	0.00	12,137,137	271,392
2	127,699	11,237	19.39	0.00	12,000,331	260,175
9	129,954	2254	19.39	0.00	5,864,334	257,928
5	134,327	4373	19.39	0.00	15,127,386	253,560
25	137,448	3121	19.40	0.01	744,399	250,523
1 (NICE)	141,191	3743	19.40	0.00	Dominated	246,709
18	193,797	56,349	19.41	0.01	6,909,043	194,337
11	205,054	11,257	19.41	0.00	12,132,357	183,099
4	216,295	11,241	19.41	0.00	11,995,711	171,876
Probability of recurrence per month = 1.00						
77	87,059		19.39			300,662
80	90,696	2803	19.39	0.00	Dominated	297,031
86	90,961	3069	19.39	0.00	Dominated	296,765
83	98,364	6536	19.39	0.00	28,314,236	289,373
23	100,861	1567	19.39	0.00	Dominated	286,883
15	105,509	6215	19.39	0.00	28,631,462	282,240
8	116,812	11,304	19.39	0.00	14,008,860	270,952
2	128,117	11,304	19.39	0.00	13,827,533	259,664
9	130,361	2245	19.39	0.00	6,806,023	257,426
5	134,772	4411	19.39	0.00	17,294,387	253,020
25	137,881	3108	19.39	0.00	860,915	249,984
1 (NICE)	141,666	3785	19.39	0.00	Dominated	246,138
18	194,464	56,584	19.40	0.01	8,009,114	193,542
11	205,772	11,307	19.40	0.00	14,004,157	182,250
4	217,080	11,308	19.40	0.00	13,823,009	170,959

Hypothetical prognostic test

The hypothetical prognostic test produced the results in *Table 43* in terms of cost, effectiveness, ICERs and NMB.

TABLE 43 Results of hypothetical prognostic test in surveillance of patients with stage IA and patients with stage IB

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (QALY)	Incremental effectiveness (QALY)	ICER (£/QALY)	NMB (£)
Stage IA						
22	8456		14.72			285,985
Prognostic	8618	162	14.74	0.02	12,344	286,086
1 (NICE)	9277	658	14.74	0.00	146,685	285,517
Stage IB						
77	9457		14.56			281,823
Prognostic	10,099	642	14.57	0.01	77,015	281,348
23	10,235	136	14.58	0.01	11,273	281,453
1 (NICE)	12,606	2371	14.60	0.02	158,111	279,347

EME
HS&DR
HTA
PGfAR
PHR

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

Published by the NIHR Journals Library