Supplementary Online Content

Marchetti MA, Coit DG, Dusza SW, et al. Performance of gene expression profile tests for prognosis in patients with localized cutaneous melanoma: a systematic review and meta-analysis. *JAMA Dermatol.* Published online July 29, 2020. doi:10.1001/jamadermatol.2020.1731

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. Trial Protocol

Version 1: 8.17.2019
Version 2: 1.16.2020 (Amendment 1)

Summary and rationale of changes in Amendment 1 (after extraction of data, before data analysis):

1. **Due to sparsity of studies reporting on distant metastases and the sparsity of data for this outcome, we changed the primary outcome after data extraction (but before data analysis) to prediction of melanoma recurrence. We changed the prediction of distant metastasis to a secondary outcome.**
2. **Due to incomplete data reporting by studies regarding time-dependent discrimination indices (sensitivity and specificity), we sought to report the overall proportion of patients with a melanoma event classified by the index GEP test as high risk, stratified by melanoma stage, and the proportion of patients without a melanoma event classified by the index GEP test as low risk, stratified by stage.**
3. **We clarified additional exclusion criteria for study selection, including duplicate publications, abstracts later published as articles, and studies with fewer than 50 participants.**
4. **We added effect estimates (hazard ratios) of index GEP test scores and survival outcomes (i.e., RFS, DMFS, MSS, and OS) as a secondary outcome.**
5. **To best assess the level of evidence (as required by journals) we identified an adaptation of GRADE proposed by Huguet A et al and recommended by the Cochrane Prognosis Methods Group.**
6. **We added a cited reference search of included studies to identify any other potentially relevant articles.**
7. **We performed decision curve analysis to examine the clinical value of the GEP test for our primary outcome.**

Title	Performance of gene expression profile-based tests for predicting clinical outcome in localized cutaneous melanoma: a systematic review and meta-analysis
Contact person	Michael A. Marchetti, MD; Dermatology Service, Department of Medicine; Memorial Sloan Kettering Cancer Center, New York, New York, United States marchetm@mskcc.org; 646-888-6016
Background	
Description of the health condition and context	The health condition under consideration in this review is localized invasive cutaneous melanoma. At present, prognostic estimates of clinical outcome are derived from the American Joint

	Committee on Cancer (AJCC) melanoma staging system, 8th edition, which inform management guidelines and clinical decision-making.[1] Patients with localized disease are followed using heterogeneous, provider-specific, surveillance protocols (i.e., variable frequency and intensity of follow-up visits and imaging tests). Recently, gene expression profile (GEP)-based prognostic tests for cutaneous melanoma have become commercially available in the United States and Europe; these tests are performed using primary cutaneous melanoma tumor tissue (formalin-fixed, paraffin-embedded). This tissue is evaluated by reverse transcription polymerase chain reaction (RT-PCR) to classify patients into high or low risk for recurrence, metastasis, or death. Health care providers may use these test results to inform their surveillance protocols for detection of metastatic disease.
Description of the prognostic / predictive model(s) / factor(s)	A prognostic factor (tumor marker prognostic factor) based on gene expression profiling of primary cutaneous melanoma tumor tissue.
Health outcomes	GEP-based prognostic tests aim to predict clinical outcomes for patients with localized invasive cutaneous melanoma. For example, DecisionDx-Melanoma (Castle Biosciences, Inc.) has been reported to predict 5-year metastasis-free survival (time from diagnosis to any regional or distant metastasis), as well as 5-year recurrence-free survival (time from diagnosis to any local, regional, or distant recurrence), 5-year distant metastasis-free survival (time from diagnosis to any distant metastasis), and 5-year melanoma-specific survival (time from diagnosis to death documented as resulting from melanoma).
Why it is important to do this review	Multiple studies have reported the overall performance of GEP-based prognostic tests for predicting clinical outcome in CM but a systematic review of their performance, particularly by AJCC Stage, is lacking. Health care providers are actively using GEP-based prognostic test results throughout the Untied States and Europe to make patient care decisions, but the settings in which GEP-based prognostic tests have clinical utility have not yet been formally established. There is preliminary evidence that the clinical validity and performance of GEP-based prognostic tests

	may vary across the risk spectrum of CM.[2] Many published studies aggregate all participants when calculating performance measures, often including with heterogenous staging characteristics. A stage-specific review of clinical validity is needed to best inform clinical utility.
Objectives	
Primary objectives	What is the clinical validity of commercially available gene expression profile (GEP)-based prognostic tests for localized cutaneous melanoma? Target population – The population of interest comprises patients diagnosed with invasive cutaneous melanoma that is localized to the primary site (Stage I/II melanoma). We will examine this population in summary and stratified by staging characteristics (e.g., T1, Stage, Stage II, etc.), if possible. **Intervention** – GEP-based prognostic tests for cutaneous melanoma **Comparator** – not applicable; although we may report the performance of multiple GEP-based prognostic tests, we do not aim to formally compare their performance using statistical measures. **Outcome** –The performance of the index GEP in predicting recurrence will be the primary outcome. **Timing** – As we expect significant heterogeneity in reported time periods by study, we will not define a time period for inclusion or exclusion and we will report data independently. The prognostic factor (tumor marker prognostic factor) is based on gene expression profiling of primary cutaneous melanoma tumor tissue and predictions are based on date of initial diagnosis. The intended timing of using the test is at the moment of diagnosis. **Setting** – An intended use of GEP-based prognostic tests is to improve risk prediction of melanoma metastasis in patients with localized disease (Stage I/II), thereby informing management decisions (e.g., type and frequency of surveillance, adjuvant therapy).
Secondary objectives	Secondary outcomes include the performance of GEP-based tests in predicting distant metastasis, melanoma death, and death from

	any cause in the same population, timing, and setting as outlined above.
Investigation of sources of heterogeneity between studies	We will extract key characteristics to evaluate clinical and methodological diversity and statistical heterogeneity between studies.
Methods	
Criteria for considering studies for this review	
Types of studies	Inclusion criteria: Any observational or experimental study reporting the performance of a GEP-based prognostic test for localized CM in humans. Identified conference abstracts will be included if adequate data is available or is provided by the study authors upon request. Exclusion criteria: case reports, review articles, animal studies.
Targeted population	Inclusion criteria: Participants with localized invasive CM without any patient characteristic restrictions. We will include hospital- and/or community-based study settings. Exclusion criteria: Participants with melanoma in situ (Stage 0 melanoma) or non-localized melanoma (Stage III or IV melanoma). We will exclude duplicate publications, abstracts later published as articles, and studies with fewer than 50 study participants. If studies include subsets of relevant participants we will include them in our review if they report outcomes specific to our intended study population.
Types of prognostic / predictive factor(s) or model(s)	We will include all commercially available GEP-based prognostic tests for localized CM. To our knowledge, these include: a. DecisionDx-Melanoma, Castle Biosciences, Inc., United States. This test classifies patients into high risk (Class 2) or low risk (Class 1) for metastasis, using a 31-gene signature. b. MelaGenix, NeraCare GmbH, Germany.

	This test classifies patients into high risk (high-risk score) or low risk (low-risk score) for disease relapse, using an eight-gene signature. We will only include studies in which models are externally validated (i.e., external validation of prediction model using independent data). We will exclude studies that exclusively report model development.
Types of outcomes to be predicted	Primary outcome: Melanoma recurrence (or relapse) Secondary outcomes: 1. Melanoma distant metastasis 2. Death from melanoma 3. Death from any cause
Search methods for identification of studies	
Electronic searches	Search Methodology: Comprehensive searches will be conducted in three electronic databases: 1) PubMed/MEDLINE (NLM) 2) EMBASE (Elsevier) 3) Web of Science (Clarivate Analytics) The literature search strategy was developed in PubMed/MEDLINE and then translated to the other databases. A combination of relevant keywords and subject headings were used; Medical Subject Headings (MeSH) in PubMed/MEDLINE and EMTREE in EMBASE. An equivalent keyword search strategy will be used in Web of Science. • No date range restrictions • English language restriction • Results will be limited to Humans (animal studies excluded) in PubMed/MEDLINE and EMBASE. (Web of Science does not have this functionality.) • MEDLINE records will be excluded from EMBASE results set Publication type limits: • Case reports will be excluded in PubMed/MEDLINE and EMBASE. (Web of Science does not have this functionality.)

- Conference abstracts and Review articles will be excluded from all three databases. (PubMed/MEDLINE does not index conference abstracts so not excluded using a search command.) (Conference abstracts, review articles, and non-English results excluded from the main search results sets will be collected from each database and stored separately in Endnote by source database, allowing the research team to review separately.)

Three concepts will make up the search strategy:

1) Gene Expression Profiling
2) Prognostic/Predicting
3) Cutaneous melanoma

Each of the three concepts will be searched upon individually, using the Boolean operator OR to combine synonyms. The individual concept searches will then be combined using the Boolean operator AND.

Database search results will be managed using the Endnote citation management program (Clarivate Analytics). Citations records will then transferred to the Covidence systematic review software (Veritas Health Innovation).

The searches will be re-run just before the final analyses and any further studies retrieved for inclusion.

<u>PubMed Search:</u>

("Gene Expression Profiling"[Mesh] OR “gene expression profile” OR “gene expression profiles” OR “gene expression profiling” OR “DecisionDx” OR “Decision-Dx” OR “Castle Biosciences” OR “NeraCare” OR “Melagenix” OR "31-GEP" OR “nine-gene” OR “9-gene” OR “eight-gene” OR “8-gene”) AND (Predict* OR "Predictive Value of Tests"[Mesh] OR prognos* OR "Prognosis"[Mesh] OR “validation” OR “validate” OR "Validation Studies as Topic"[Mesh] OR "Validation Studies" [Publication Type]) AND (“melanoma” OR "Melanoma"[Mesh] OR "Melanoma, Cutaneous Malignant" [Supplementary Concept])

NOT (animals[mh] NOT humans [mh]) NOT (Case Reports[ptyp]) NOT (Review[ptyp])
AND English[lang]

<u>Embase Search:</u>

	('gene expression profiling'/exp OR 'gene expression profile' OR 'gene expression profiles' OR 'gene expression profiling' OR 'decisiondx' OR 'decision-dx' OR 'castle biosciences' OR 'neracare' OR 'melagenix' OR '31-gep' OR 'nine-gene' OR '9-gene' OR 'eight-gene' OR '8-gene') AND (predict* OR 'predictive value'/exp OR 'predictive validity'/exp OR prognos* OR 'prognosis'/exp OR 'validation' OR 'validate' OR 'validation study'/exp) AND ('melanoma' OR 'melanoma'/exp OR 'cutaneous melanoma'/exp) NOT ('animal'/de NOT 'human'/de) NOT 'case report'/de NOT 'review'/it NOT 'conference abstract'/it AND english:la NOT [medline]/lim **Web of Science Search:** ((“gene expression profile” OR “gene expression profiles” OR “gene expression profiling” OR “DecisionDx” OR “Decision-Dx” OR “Castle Biosciences” OR “NeraCare” OR “Melagenix” OR "31-GEP" OR “nine-gene” OR “9-gene” OR “eight-gene” OR “8-gene”) AND (Predict* OR prognos* OR “validation” OR “validate”) AND (“melanoma”)) Refined by: [excluding] DOCUMENT TYPES: (REVIEW OR MEETING ABSTRACT) AND LANGUAGES: (ENGLISH) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.
Searching other resources	Grey literature searches will be conducted for additional relevant articles on the following websites: https://www.castlebiosciences.com and https://melagenix.com. A cited reference search of the studies identified for inclusion will be performed using Scopus, Web of Science, and Google Scholar.
Data collection	
Selection of studies	Titles and abstracts of studies retrieved using the search strategy and those from additional sources will be screened independently by two review authors to identify studies that potentially meet the inclusion criteria outlined above. Any disagreement between them over the eligibility of particular studies will be resolved through discussion with a third reviewer.

	The full text of these potentially eligible studies will be retrieved and independently assessed for eligibility by two review team members through application for inclusion and exclusion criteria. Any disagreement between them over the eligibility of particular studies will be resolved through discussion with a third reviewer.
Data extraction and management	A standardized, pre-piloted form will be used to extract data from the included studies for assessment of study quality and evidence synthesis. This form was adapted from the CHARMS-PF checklist (checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies).[3] Extracted information will include: funding source of study and any potential conflicts of interest, aim of study, prespecified hypotheses, study design, participant sampling technique, methods and sources of case ascertainment, study inclusion criteria, study exclusion criteria, study start and end dates, type of GEP test and pre-specified cut-points, participant selection (# of potentially eligible cases, # of cases excluded due to eligibility criteria, # of cases meeting inclusion criteria, # of cases meeting inclusion criteria that were excluded, final # of participants), # of participants included in study not previously published elsewhere, participant demographics and baseline characteristics (i.e., age, sex, race/ethnicity, AJCC Stage, Breslow thickness, mitotic index, ulceration, anatomic site, date range of melanoma diagnosis), treatments received by participants, follow-up times, study outcomes including definitions, outcomes determined blinded or unblinded to prognostic factors under review, missing data and how handled, sample size calculations, analysis methodology (i.e., modeling method, modeling assumptions checked, selection or exclusion of factors during multivariable modeling, methods of handling continuous variables), absolute effect measures (i.e., # of true positives, false positives, true negatives, and false negatives) and rates (95% CIs) of the GEP-test for DMFS, RFS, MSCC and OS in summary and by melanoma stage of disease; unadjusted and adjusted prognostic effect estimates (i.e., hazard ratios) and set of adjustment factors used, information for assessment of NCCN Tumor Marker Category Trial Design; information for assessment of the risk of bias, including use of the REMARK (Reporting Recommendations for Tumor Marker Prognostic Studies) guideline and QUIPS (Quality in Prognosis Studies) tool. Two review authors will extract data independently and discrepancies will be identified and resolved through discussion and by a third reviewer if necessary. Any missing data will be requested from study authors.

Assessment of risk of bias in included studies	Two review authors will independently assess the risk of bias in included studies using the QUIPS (Quality in Prognosis Studies) tool,[4] which measures 6 domains for risk of bias: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting. The risk of bias for each domain will be rated as high, moderate, or low, for study participants, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting. Two review authors will assess the risk of reporting bias using the REMARK (Reporting Recommendations for Tumor Marker Prognostic Studies) checklist.[5] Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary.
Measures of association or predictive performance measures to be extracted	We will extract absolute effect measures (i.e., sensitivity and specificity), in summary and by CM stage. If these are not explicitly reported but the underlying raw data (i.e., true positives, false positives, true negatives, and false negatives) needed to calculate absolute effect measures are available, we will extract raw data. Furthermore, we will extract the risk ratio, odds ratio, hazard ratio (preferred), and/or survival point-estimates, in summary and by stage/substage, as available. During data extraction we identified that sensitivity and specificity were unable to be precisely calculated using provided data. Here we define sensitivity and specificity as the predictive accuracy of the GEP test at a particular cross-sectional timepoint (i.e., 5-year). Thus, a patient is defined as having a melanoma event if the event occurred before the specified timepoint; a patient is defined as not having had a melanoma event if they have had follow-up beyond the specified timepoint without an event. As this data was uncommonly reported, we instead sought to report: (i) the proportion of patients in a study with a melanoma recurrence classified by the index GEP test as high risk, stratified by AJCC stage (i.e., stage I and stage II), and (ii) the proportion of

	patients in a study without a melanoma recurrence classified by the index GEP test as low risk, stratified by stage.
Assessment of heterogeneity	Difference in study design, patient or disease characteristics between studies can be a hinderance to providing a reasonable summary estimate of the outcome measure of the combined studies. For all of the studies meeting inclusion criteria, we will abstract information on studies (geographic location of enrollment, study setting, etc.), demographics (age and sex of participants, etc.) and disease characteristics (anatomic location of primary lesion, Breslow thickness, ulceration, mitotic index, etc.). We will estimate the degree of heterogeneity present by calculating the I^2 statistic for all studies included in the meta-analysis. If significant heterogeneity is present, we will explore the use of meta-regression to explain the heterogeneity in the treatment effects by one or more of these study, patient or disease characteristics. Performing any meta-regression is incumbent on the availability and granularity of the data presented in the individual studies.
Assessment of reporting deficiencies	Reporting biases can lead to an exaggeration of effect estimate since studies with positive results are more likely to be published than null findings. We will create funnel plots to visually evaluate any systematic differences between smaller and larger studies and the observed effect sizes. We will also estimate Egger's test to evaluate the significance of the publication bias.
Data synthesis	We will provide a narrative synthesis of the findings from the included studies, structured around the prognostic performance of the test for predicting clinical outcome, in aggregate and stratified by AJCC Melanoma Stage. We anticipate there will be limited scope for meta-analysis because of the range of different follow-up durations and variability in definitions of clinical outcomes (e.g., metastasis-free survival vs. distant metastasis-free survival). However, where studies have used the same type of intervention, follow-up duration, and outcome measure, we will pool the results using a random-effects meta-analysis.

	The quality of evidence of each index GEP test will be assessed using an adaptation of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) proposed by Huguet A et al[6,7] for systematic reviews of prognostic factor research and recommended by the Cochrane Prognosis Methods Group.[8] This framework considers phase of investigation, study limitations, inconsistency, indirectness, imprecision, publication bias, moderate/large effect size, and exposure-response gradient to assign an overall quality rating (high quality, moderate quality, low quality, very low quality). The clinical effects of index GEP tests in the management of stage I and stage II melanoma patients will be described and compared via decision curve analysis.[9] A predictive marker, model, or test is considered to have clinical value if it has the highest net benefit across the range of thresholds for which an individual would be designated at high risk. Briefly, the net benefit of a model is the difference between the proportion of true positives and the proportion of false positives weighted by the odds of the selected threshold for high risk designation.[10] At any given threshold, the model with the higher net benefit is the preferred model.

eMethods. Supplementary Methods

Search Methodology:

Comprehensive searches were conducted in three databases: (initial 7/28/19, update 12/12/19)

1) PubMed/MEDLINE (NLM)

2) EMBASE (Elsevier)

3) Web of Science (Clarivate Analytics)

The literature search strategy was developed in PubMed/MEDLINE and translated for EMBASE and Web of Science. Relevant keywords and subject headings (when available: MeSH in PubMed/MEDLINE and EMTREE in EMBASE) were used.

Search Limits:

- No date range restrictions
- English language restriction
- Results were limited to Humans (animal studies excluded) in PubMed/MEDLINE and EMBASE. (Web of Science does not have this functionality.)
- MEDLINE records were excluded from EMBASE results set
- Case reports were excluded in PubMed/MEDLINE and EMBASE. (Web of Science does not have this functionality.)

Three concepts made up the search strategy:

1) Gene Expression Profiling

2) Prognostic/Predicting

3) Cutaneous melanoma

Each of these concepts was searched on individually (merging synonyms with Boolean operator OR) and then added all together (using the Boolean operator AND). Database search results were harvested in the Endnote citation management program (Clarivate Analytics). Citations records were then transferred to the Covidence systematic review software (Veritas Health Innovation).

The database search strategies were re-run just before the final analyses and a cited reference search was conducted on the articles identified for inclusion using Web of Science (Clarivate Analytics), Scopus (Elsevier), and Google Scholar (https://scholar.google.com/). A grey literature search of the websites https://www.castlebiosciences.com and https://melagenix.com revealed no additional relevant studies. The final included studies were decided on by discussion between authors, with full agreement required before inclusion. See eTable1.

Data extraction: Data were extracted by M.M. and E.B., independently, using a piloted form adapted from a version of the CHARMS-PF checklist.[3] The extracted data included source of data, participant characteristics, participant recruitment methods, outcomes, index and comparator prognostic factors, missing data, analyses, results, and interpretation/discussion. Study authors were contacted to obtain unpublished stage-specific data relevant to the study outcomes. We were able to extract stage-specific results from Keller et al, Greenhaw et al, and Zager et al directly from review of the manuscript. The authors of Podlipnik et al and Hsueh et al were contacted for a breakdown of stage I and stage II results. Podlipnik et al provided those results to us. Hsueh et al did not provide those results to us.

Assessment of individual study risk of bias: The Quality in Prognosis Studies (QUIPS) tool considers bias across six domains (study participation, study attrition, prognostic factor measurement, adjustment for other prognostic factors, outcome measurement, and statistical analysis and reporting).[4] A study satisfying low risk of bias in all six domains was designated as low overall risk of bias. A study with a high risk of bias in one or more domains was designated as high overall risk of bias, which adhered to the Cochrane risk of bias assessment recommendations.[11] The quality of individual study reporting was assessed independently using the reporting recommendations for tumor marker prognostic studies (REMARK) checklist.[12] Any disagreements between authors were resolved through discussion. No disagreements required resolution by a third reviewer. Reviewers were not blinded to study authors, institution, or journal of publication due to feasibility.

Prognostic factor level of evidence: The quality of evidence of each index GEP test was assessed as high, moderate, low, or very low quality using an adaptation of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) proposed for systematic reviews of prognostic factor research[6,7] and recommended by the Cochrane Prognosis Methods Group.[8] This framework considers phase of investigation, study limitations, inconsistency, indirectness, imprecision, publication bias, moderate/large effect size, and exposure-response gradient to assign an overall quality rating (high quality, moderate quality, low quality, very low quality). Quality of evidence assessment was done by M.M. and E.B. independently for each survival outcome by disease stage and was guided specifically to conform with the recommendations of Huguet et al (outlined in Table 4 of that manuscript).[6] Disagreements were

resolved through discussion without need for a third reviewer. Reviewers were not blinded to study authors, institution, or journal of publication due to feasibility.

Outcomes: Time-based sensitivity and specificity were extracted at a particular cross-sectional timepoint when feasible. In this analysis, a patient was defined as having a melanoma event if the event occurred before the specified timepoint; a patient was defined as not having had a melanoma event if they had follow-up until the specified timepoint without an event. Therefore, sensitivity was the proportion of patients with an event before the specified timepoint who were classified as high-risk by GEP testing; specificity was the proportion of patients with follow-up until the specified timepoint who have not had an event and were classified as low-risk by GEP testing.

eTable 1. Database Search Strategies
PubMed/MEDLINE (NLM)
(("Gene Expression Profiling"[Mesh] OR "gene expression profile" OR "gene expression profiles" OR "gene expression profiling" OR "DecisionDx" OR "Decision-Dx" OR "Castle Biosciences" OR "NeraCare" OR "Melagenix" OR "31-GEP" OR "nine-gene" OR "9-gene" OR "eight-gene" OR "8-gene") **AND** (Predict* OR "Predictive Value of Tests"[Mesh] OR prognos* OR "Prognosis"[Mesh] OR "validation" OR "validate" OR "Validation Studies as Topic"[Mesh] OR "Validation Studies" [Publication Type]) **AND** ("melanoma" OR "Melanoma"[Mesh] OR "Melanoma, Cutaneous Malignant" [Supplementary Concept])) **NOT** (animals[mh] NOT humans [mh]) **NOT** (Case Reports[ptyp]) **AND** English[lang]
EMBASE (Elsevier)
(('gene expression profiling'/exp OR 'gene expression profile' OR 'gene expression profiles' OR 'gene expression profiling' OR 'decisiondx' OR 'decision-dx' OR 'castle biosciences' OR 'neracare' OR 'melagenix' OR '31-gep' OR 'nine-gene' OR '9-gene' OR 'eight-gene' OR '8-gene') **AND** (predict* OR 'predictive value'/exp OR 'predictive validity'/exp OR prognos* OR 'prognosis'/exp OR 'validation' OR 'validate' OR 'validation study'/exp) **AND** ('melanoma' OR 'melanoma'/exp OR 'cutaneous melanoma'/exp)) **NOT** ('animal'/de NOT 'human'/de) **NOT** ('case report'/de) **NOT** ([medline]/lim) **AND** (english:la)
Web of Science (Clarivate Analytics)
(("gene expression profile" OR "gene expression profiles" OR "gene expression profiling" OR "DecisionDx" OR "Decision-Dx" OR "Castle Biosciences" OR "NeraCare" OR "Melagenix" OR "31-GEP" OR "nine-gene" OR "9-gene" OR "eight-gene" OR "8-gene") **AND** (Predict* OR prognos* OR "validation" OR "validate") **AND** ("melanoma")) **AND** LANGUAGES: (ENGLISH)

eTable 2. Key considerations in study risk of bias assessment (QUIPS tool)

	Study participation	Study attrition	Prognostic factor measurement	Outcome Measurement	Study confounding	Statistical analysis and reporting
Hsueh et al, 2017	✓ Participant description* ✓ Sampling frame hospital and community based ✓ Multi-center ✕ Unclear participation rate of population of interest ✕ Melanoma characteristics† ✕ Melanoma diagnosis date range ✕ Recruitment discontinued prior to accrual of pre-specified sample size	✓ Adequate response rate ✕ Unclear rate and reasons of participants lost to follow-up ✕ Unclear attempts to collect information on participants lost to follow-up	✓ Method and setting of PF measurement same for all participants ✓ Clear description of PF ✕ Participants without successful GEP test not described	✕ Not blinded to GEP ✕ Short follow-up ✕ Method and setting of outcome measurement not same for all participants	✓ AJCC confounders** ✕ Additional confounders‡ ✕ Unclear if any participants had multiple primary melanomas ✕ Unclear differences in follow-up intensity by GEP risk score ✕ Unclear if differences in systemic therapy between GEP groups	✓ Reported distributions of available key variables ✓ Relation of key variables to GEP ✓ KM survival curves ✓ MV analysis (AJCC variables) ✕ UV analysis ✕ Full MV analysis with all variables, regardless of significance ✕ Time-based sensitivity and specificity analyses (when reported)
Greenhaw et al, 2018	✓ Participant description* ✕ Unclear participation rate of population of interest ✕ Melanoma characteristics† ✕ Melanoma diagnosis date range ✕ Restricted sampling frame ✕ Retrospective ✕ Unclear pre-specified sample size calculation ✕ Single-center	✓ Telephone calls to patients followed by other physicians ✕ Unclear response rate ✕ Unclear rate and reasons of participants lost to follow-up	✓ Method and setting of PF measurement same for all participants ✓ Clear description of PF ✕ Participants without successful GEP test not described	✕ Not blinded to GEP ✕ Short follow-up ✕ Follow-up rule-based by GEP test ✕ Method and setting of outcome measurement not same for all participants	✕ AJCC or additional confounders** ‡ ✕ Unclear if any participants had multiple primary melanomas ✕ Greater follow-up intensity in GEP high risk test scores ✕ Unclear if differences in systemic therapy between GEP groups	✓ Reported distributions of available key variables ✓ Relation of key variables to GEP ✓ KM survival curves ✕ UV analysis ✕ MV analysis ✕ Full MV analysis with all variables, regardless of significance ✕ Time-based sensitivity and specificity analyses (when reported)
Zager et al, 2018	✓ Melanoma diagnosis date range ✓ Sampling frame hospital and community based ✓ Multi-center ✕ Unclear participation rate of population of interest ✕ Participant description* ✕ Melanoma characteristics† ✕ Retrospective ✕ Unclear pre-specified sample size calculation	✓ Adequate response rate ✓ Adequate follow-up	✓ Method and setting of PF measurement same for all participants ✓ Clear description of PF ✕ Participants without successful GEP test not described	✓ Good follow-up ✓ Blinded to GEP ✕ Method and setting of outcome measurement not same for all participants	✓ AJCC confounders** ✕ Additional confounders‡ ✕ Unclear if any participants had multiple primary melanomas ✕ Unclear if differences in systemic therapy between GEP groups	✓ Reported distributions of available key variables ✓ KM survival curves ✓ UV analysis (AJCC variables) ✓ MV analysis (AJCC variables) ✕ Relation of key variables to GEP ✕ Full UV and MV analyses with all variables, regardless of significance ✕ Time-based sensitivity and specificity analyses (when reported)
Keller et al, 2019	✓ Participant description* ✓ Melanoma diagnosis date range ✕ Unclear participation rate of population of interest ✕ Melanoma characteristics† ✕ Restricted sampling frame ✕ Unclear pre-specified sample size calculation ✕ Single-center	✓ Adequate response rate ✕ Unclear rate and reasons of participants lost to follow-up ✕ Unclear attempts to collect information on participants lost to follow-up	✓ Method and setting of PF measurement same for all participants ✓ Clear description of PF ✕ Inadequate proportion of potentially eligible participants had successful PF measurement (<95%) ✕ Initial GEP testing may have preceded commercial availability	✓ Rule-based by stage of disease ✕ Not blinded to GEP ✕ Short follow-up	✓ AJCC confounders** ✕ Additional confounders‡ ✕ Unclear if any participants had multiple primary melanomas ✕ Unclear if differences in systemic therapy between GEP groups	✓ Reported distributions of available key variables ✓ KM survival curves ✓ Relation of key variables to GEP ✓ UV analysis ✓ MV analysis ✕ Full UV and MV analyses with all variables, regardless of significance ✕ Time-based sensitivity and specificity analyses (when reported)
Podlipnik et al, 2019	✓ Participant description* ✓ Melanoma characteristics† ✓ Melanoma diagnosis date range ✓ Multi-center ✕ Unclear participation rate of population of interest ✕ Restricted sampling frame ✕ Unclear pre-specified sample size calculation ✕ Single-center	✓ Adequate response rate ✕ Unclear rate and reasons of participants lost to follow-up ✕ Unclear attempts to collect information on participants lost to follow-up	✓ Method and setting of PF measurement same for all participants ✓ Clear description of PF ✓ Adequate proportion of study participants had successful PF measurement (>95%)	✕ Not blinded to GEP ✕ Short follow-up ✕ Method and setting of outcome measurement not same for all participants	✓ AJCC confounders** ✕ Additional confounders‡ ✕ Unclear if any participants had multiple primary melanomas ✕ Unclear differences in follow-up intensity by GEP risk score	✓ Reported distributions of available key variables ✓ Relation of key variables to GEP ✓ KM survival curves ✓ UV analysis (AJCC variables) ✓ MV analysis (AJCC variables) ✕ Full UV and MV analyses with all variables, regardless of significance
Koelblinger et al, 2018	✕ Unclear participation rate of population of interest ✕ Participant description* ✕ Melanoma characteristics† ✕ Melanoma diagnosis date range ✕ Restricted sampling frame ✕ Retrospective ✕ Unclear pre-specified sample size calculation ✕ Unclear # of contributing centers	✓ Adequate response rate ✕ Unclear rate and reasons of participants lost to follow-up ✕ Unclear attempts to collect information on participants lost to follow-up	✓ Method and setting of PF measurement same for all participants ✓ Clear description of PF ✕ Participants without successful GEP test not described	✓ Blinded to GEP ✕ Short follow-up ✕ Method and setting of outcome measurement not same for all participants	✕ AJCC or additional confounders** ‡ ✕ Unclear if any participants had multiple primary melanomas ✕ Unclear if differences in systemic therapy between GEP groups	✕ Reported distributions of available key variables ✕ Relation of key variables to GEP ✕ KM survival curves ✕ UV analysis ✕ MV analysis ✕ Full UV and MV analyses with all variables, regardless of significance ✕ Time-based sensitivity and specificity analyses (when reported)
Amaral et al, 2020	✓ Clear and adequate participation rate of population of interest ✓ Participant description* ✓ Melanoma diagnosis date range ✕ Melanoma characteristics† ✕ Restricted sampling frame ✕ Retrospective ✕ Unclear pre-specified sample size calculation	✓ Adequate response rate ✕ Unclear rate and reasons of participants lost to follow-up ✕ Unclear attempts to collect information on participants lost to follow-up	✓ Method and setting of PF measurement same for all participants ✓ Clear description of PF ✓ Adequate proportion of study participants had successful PF measurement (>95%)	✓ Blinded to GEP ✓ Moderate follow-up (stage II disease) ✓ Rule-based by stage of disease	✓ Re-verification of melanoma diagnosis ✓ AJCC confounders** ✕ Additional confounders‡ ✕ Unclear if any participants had multiple primary melanomas ✕ Unclear if differences in systemic therapy between GEP groups	✓ Reported distributions of available key variables ✓ KM survival curves ✓ UV analysis (AJCC variables) ✓ MV analysis (AJCC variables) ✕ Unclear relation of key variables to GEP (no statistical testing) ✕ Full UV and MV analyses with all variables, regardless of significance ✕ No MV analysis reported for RFS and DMFS

*adequate reporting of age and sex; †adequate reporting of Breslow thickness, ulceration, mitotic rate, and anatomic site

**Breslow thickness, ulceration, sentinel lymph node (when relevant), or AJCC stage; ‡age, sex, mitotic rate, anatomic site, melanoma subtype

PF = prognostic factor; GEP = gene expression profile; KM = Kaplan-Meier; UV = univariable; MV = multivariable; AJCC = American Joint Committee on Cancer; RFS = recurrence-free survival; DMFS = distant metastasis-free survival

eTable 3. Performance of index gene expression profile tests in predicting melanoma distant metastasis by study and stage of disease

	Index GEP test	Source	Patients, No.	Events, No.	Observed DMFS Rates (year)		Association Between GEP High Score and Event	Proportion of Events Classified as High Risk‡	Proportion of Non-Events Classified as Low Risk‡	Proportion of High Risk Patients with Event‡	Proportion of Low Risk Patients without Event‡
					GEP Low Score	GEP High Score					
Stage I	Decision Dx	Hsueh et al, 2017	-	-	-	-	-	-	-	-	-
		Greenhaw et al, 2018	-	-	-	-	-	-	-	-	-
		Zager et al, 2018	264	13	97% (5-y)	90% (5-y)	-	31%	86%	10%	96%
		Keller et al, 2019	96	0	-	-	-	N/A	95%	0%	100%
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	-	-	-	-	-	-	-	-	-
		Amaral et al, 2020	-	-	-	-	-	-	-	-	-
Stage II	Decision Dx	Hsueh et al, 2017	-	-	-	-	-	-	-	-	-
		Greenhaw et al, 2018	-	-	-	-	-	-	-	-	-
		Zager et al, 2018	93	30	90% (5-y)	63% (5-y)	-	87%	44%	43%	88%
		Keller et al, 2019	40	9	-	-	-	89%	48%	33%	94%
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	-	-	-	-	-	-	-	-	-
		Amaral et al, 2020	245	47	89% (5-y) 89% (10-y)	70% (5-y) 63% (10-y)	-	81%	41%	25%	90%
Stage I+II	Decision Dx	Hsueh et al, 2017	282	6	100% (1.5-y)	93% (1.5-y)	-	83%	81%	9%	>99%
		Greenhaw et al, 2018	-	-	-	-	-	-	-	-	-
		Zager et al, 2018	357	43	-	-	-	70%	78%	30%	95%
		Keller et al, 2019	136	9	-	-	-	89%	83%	28%	99%
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	-	-	-	-	-	-	-	-	-
		Amaral et al, 2020	-	-	-	-	-	-	-	-	-

*Koelblinger et al and Amaral et al used different GEP score cut-offs.

‡ Unless indicated by a particular cross-sectional follow-up time (i.e., 3- or 5-year), reported proportions were calculated using the raw number of high-score GEP patients with an event, number of high-score GEP patients without an event, number of low-score GEP patients with an event, and number of low-score GEP patients without an event.

GEP = gene expression profile; DMFS = distant metastasis-free survival; y = year

eTable 4. Performance of index gene expression profile tests in predicting death from melanoma by study and stage of disease

	Index GEP test	Source	Patients, No.	Events, No.	Observed MSS Rates (year)		Association Between GEP High Score and Event	Proportion of Events Classified as High Risk‡	Proportion of Non-Events Classified as Low Risk‡	Proportion of High Risk Patients with Event‡	Proportion of Low Risk Patients without Event‡
					GEP Low Score	GEP High Score					
Stage I	Decision Dx	Hsueh et al, 2017	-	-	-	-	-	-	-	-	-
		Greenhaw et al, 2018	-	-	-	-	-	-	-	-	-
		Zager et al, 2018	264	3	99% (5-y)	97% (5-y)	-	33%	85%	3%	99%
		Keller et al, 2019	-	-	-	-	-	-	-	-	-
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	88	9	-	-	-	44%	77%	18%	92%
		Amaral et al, 2020	-	-	-	-	-	-	-	-	-
Stage II	Decision Dx	Hsueh et al, 2017	-	-	-	-	-	-	-	-	-
		Greenhaw et al, 2018	-	-	-	-	-	-	-	-	-
		Zager et al, 2018	93	8	100% (5-y)	87% (5-y)	-	100%	38%	13%	100%
		Keller et al, 2019	-	-	-	-	-	-	-	-	-
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	-		-	-	-				
		Amaral et al, 2020	245	32	92% (5-y) 92% (10-y)	82% (5-y) 67% (10-y)	HR 1.55† (95% CI 1.13-2.13, p=0.006)	84%	40%	18%	95%
Stage I+II	Decision Dx	Hsueh et al, 2017	-	-	-	-	-	-	-	-	-
		Greenhaw et al, 2018	256	6	99% (3-y) 99% (5-y)	86% (3-y) 79% (5-y)	-	83% 80% (3-y) 83% (5-y)	85% 79% (3-y) 70% (5-y)	12%	>99%
		Zager et al, 2018	357	11	-	-	-	82%	74%	9%	99%
		Keller et al, 2019	-	-	-	-	-	-	-	-	-
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	-	-	-	-	-	-	-	-	-
		Amaral et al, 2020	-	-	-	-	-	-	-	-	-

*Koelblinger et al and Amaral et al use different GEP score cut-offs.

‡Unless indicated by a particular cross-sectional follow-up time (i.e., 3- or 5-year), reported proportions were calculated using the raw number of high-score GEP patients with an event, number of high-score GEP patients without an event, number of low-score GEP patients with an event, and number of low-score GEP patients without an event. If indicated by a cross-sectional follow-up time, these estimates represent the sensitivity and specificity of the test.

†Multivariate hazard ratio adjusted for Breslow thickness and age

GEP = gene expression profile; MSS = melanoma-specific survival; HR = hazard ratio; y = year

eTable 5. Performance of index gene expression profile tests in predicting death from any cause by study and stage of disease

	Index GEP test	Source	Patients, No.	Events, No.	Observed OS Rates (year)		Association Between GEP High Score and Event	Proportion of Events Classified as High Risk‡	Proportion of Non-Events Classified as Low Risk‡	Proportion of High Risk Patients with Event‡	Proportion of Low Risk Patients without Event‡
					GEP Low Score	GEP High Score					
Stage I	Decision Dx	Hsueh et al, 2017	-	-	-	-	-	-	-	-	-
		Greenhaw et al, 2018	-	-	-	-	-	-	-	-	-
		Zager et al, 2018	-	-	-	-	-	-	-	-	-
		Keller et al, 2019	-	-	-	-	-	-	-	-	-
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	-	-	-	-	-	-	-	-	-
		Amaral et al, 2020	-	-	-	-	-	-	-	-	-
Stage II	Decision Dx	Hsueh et al, 2017	-	-	-	-	-	-	-	-	-
		Greenhaw et al, 2018	-	-	-	-	-	-	-	-	-
		Zager et al, 2018	-	-	-	-	-	-	-	-	-
		Keller et al, 2019	-	-	-	-	-	-	-	-	-
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	-	-	-	-	-	-	-	-	-
		Amaral et al, 2020	-	-	-	-	-	-	-	-	-
Stage I+II	Decision Dx	Hsueh et al, 2017	282	10	98% (1.5y)	92% (1.5y)	-	70%	81%	12%	99%
		Greenhaw et al, 2018	-	-	-	-	-	-	-	-	-
		Zager et al, 2018	-	-	-	-	-	-	-	-	-
		Keller et al, 2019	-	-	-	-	-	-	-	-	-
		Podlipnik et al, 2019	-	-	-	-	-	-	-	-	-
	MelaGenix*	Koelblinger et al, 2018	-	-	-	-	-	-	-	-	-
		Amaral et al, 2020	-	-	-	-	-	-	-	-	-

*Koelblinger et al and Amaral et al use different GEP score cut-offs.

‡ Reported proportions were calculated using the raw number of high-score GEP patients with an event, number of high-score GEP patients without an event, number of low-score GEP patients with an event, and number of low-score GEP patients without an event.

GEP = gene expression profile; OS = overall survival; y = year

eTables 6-9. Adapted Grading of Recommendations Assessment, Development, and Evaluation (GRADE) table for systematic reviews of prognostic studies

eTable 6

Outcome: Melanoma Recurrence																			
Potential prognostic factor	Disease stage	Number of participants	Number of studies	Number of cohorts	Univariate			Multivariate			Investigation Phase	GRADE factors†							
					+	0	-	+	0	-		Study limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Moderate/large effect size	Dose effect	Overall quality
DecisionDx	Stage I	623	4	4							2	×	Unclear	✓	×	×	Unclear	×	+
	Stage II	212	4	4							2	×	✓	✓	✓	×	Unclear	×	++
	Stage I+II	1117	5	5*	2			1			2	×	✓	×	✓	×	✓	Unclear	++
MelaGenix	Stage I	88	1	1							2	×	N/A	×	×	×	Unclear	Unclear	+
	Stage II	245	1	1							2	×	N/A	✓	✓	×	Unclear	Unclear	++
	Stage I+II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

eTable7

B. Outcome: Melanoma Distant Metastasis																			
Potential prognostic factor	Disease stage	Number of participants	Number of studies	Number of cohorts	Univariate			Multivariate			Investigation Phase	GRADE factors†							
					+	0	-	+	0	-		Study limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Moderate/large effect size	Dose effect	Overall quality
DecisionDx	Stage I	360	2	2							2	×	Unclear	×	×	×	Unclear	×	+
	Stage II	133	2	2							2	×	✓	✓	×	×	Unclear	✓	+
	Stage I+II	775	3	3							2	×	✓	×	✓	×	Unclear	Unclear	++
MelaGenix	Stage I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Stage II	245	1	1							2	×	N/A	✓	✓	×	Unclear	Unclear	++
	Stage I+II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

eTable8

C. Outcome: Death from Melanoma																			
Potential prognostic factor	Disease stage	Number of participants	Number of studies	Number of cohorts	Univariate			Multivariate			Investigation Phase	GRADE factors†							
					+	0	-	+	0	-		Study limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Moderate/large effect size	Dose effect	Overall quality
DecisionDx	Stage I	264	1	1							2	×	N/A	×	×	×	Unclear	×	+
	Stage II	93	1	1							2	×	N/A	✓	×	×	Unclear	✓	+
	Stage I+II	613	2	2							2	×	✓	×	×	×	Unclear	Unclear	+
MelaGenix	Stage I	88	1	1							2	×	N/A	×	×	×	Unclear	Unclear	+
	Stage II	245	1	1				1			2	×	N/A	✓	✓	✓	×	Unclear	++
	Stage I+II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

eTable9

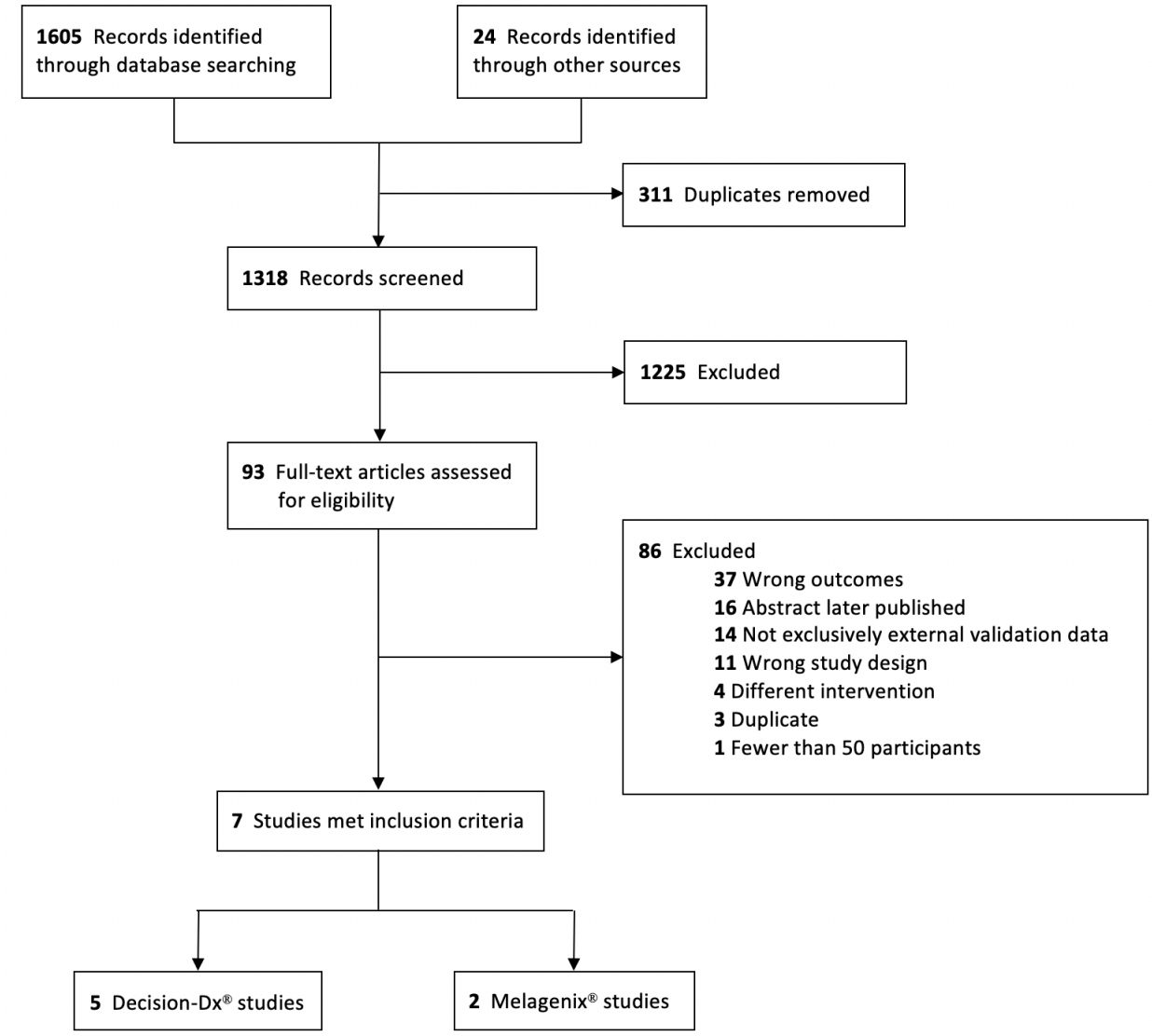
D. Outcome: Death from any cause																			
Potential prognostic factor	Disease stage	Number of participants	Number of studies	Number of cohorts	Univariate			Multivariate			Investigation Phase	GRADE factors†							
					+	0	-	+	0	-		Study limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Moderate/large effect size	Dose effect	Overall quality
DecisionDx	Stage I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Stage II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Stage I+II	282	1	1							1	×	N/A	✓	×	×	Unclear	Unclear	+
MelaGenix	Stage I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Stage II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Stage I+II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Phase, phase of investigation. For uni- and multivariate analyses: +, number of significant effects with a positive value; 0, number of non-significant effects; -, number of significant effects with a negative value. For GRADE factors: ✓, no serious limitations; ×, serious limitations (or not present for moderate/large effect size, dose effect); unclear, unable to rate item based on available information. For overall quality of evidence: +, very low; ++, low; +++, moderate; ++++, high.

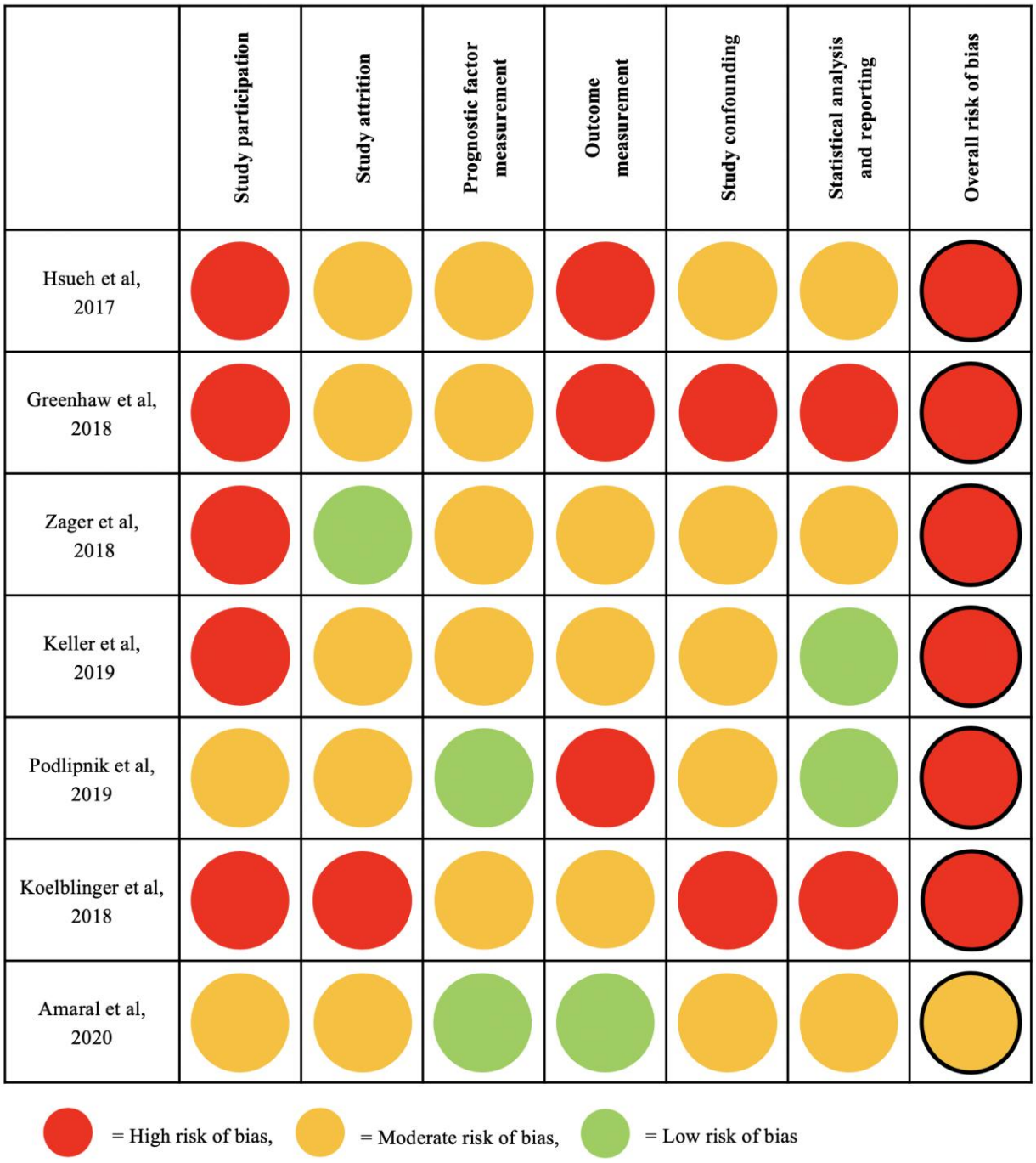
*2 cohorts partially overlapping

†Study limitations refers to assessment by QUIPS tool. Inconsistency refers to unexplained heterogeneity in results across studies. Indirectness refers to a participant population, prognostic factor, and/or outcomes in a primary study that do not fully represent the question defined by the systematic review. Imprecision refers to uncertainty in the relationship between the prognostic factor and its associated risk or predictive value. Publication bias refers to when the published evidence is restricted to only a portion of the studies or analyses conducted on the topic. Effect size refers to moderate or large effect (i.e., odds ratio or hazard ratio). Dose effect refers to when elevated levels of the prognostic factor lead to a larger effect size over lower levels of the factor.

eFigure 1. Diagram of the Study Selection for the Systematic Review

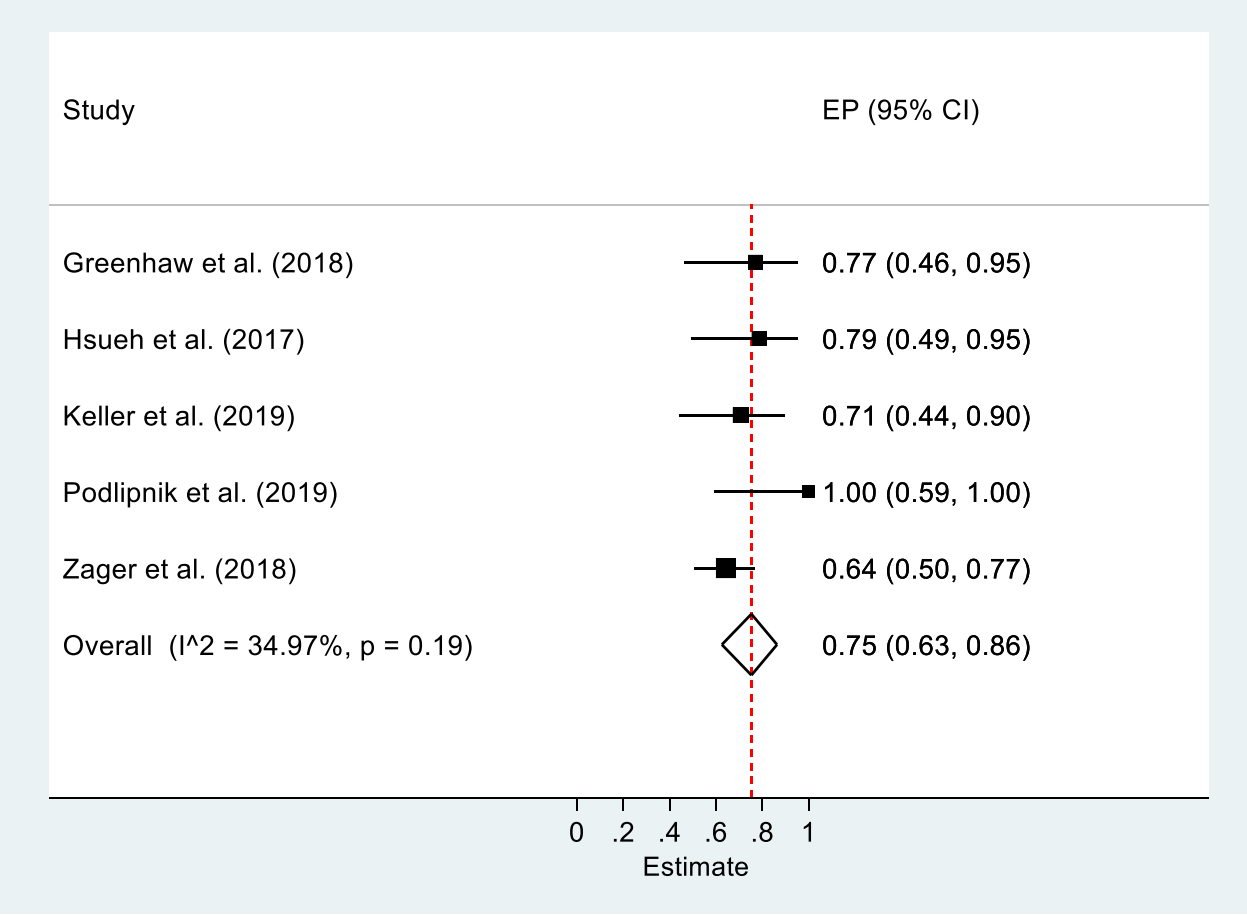


eFigure 2. Risk of bias assessment using the QUIPS tool



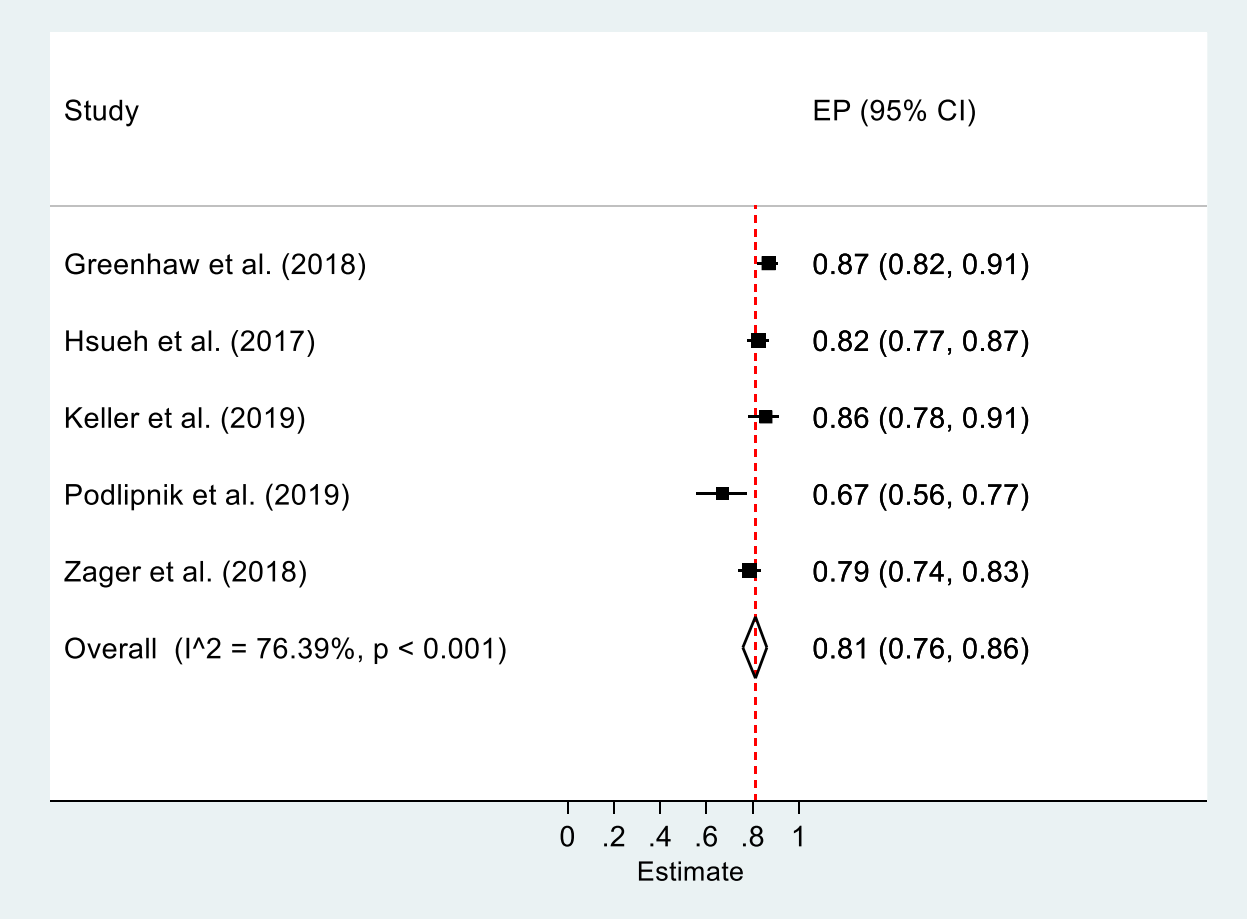
A study satisfying low risk of bias in all six domains was designated as low overall risk of bias. A study with a high risk of bias in one or more domains was designated as high overall risk of bias.

eFigure 3. Forest plot of the proportion of stage I-II patients with a melanoma recurrence correctly classified as high risk by DecisionDx-Melanoma®



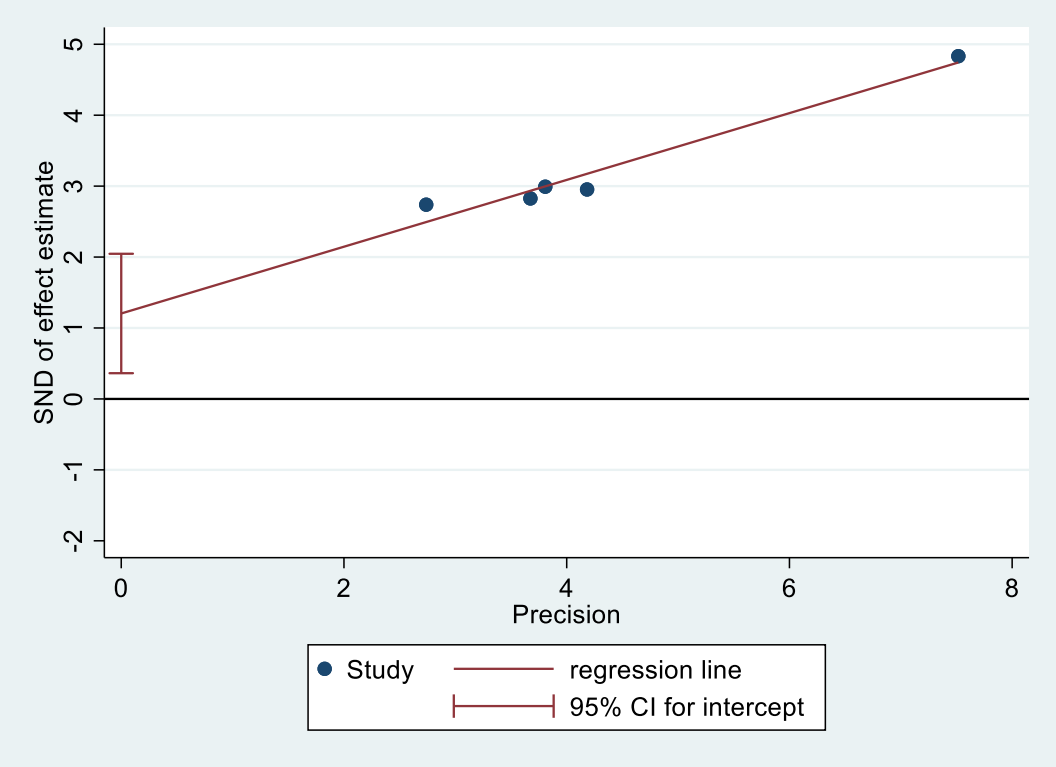
EP = estimated proportion; CI= confidence interval

eFigure 4. Forest plot of the proportion of stage I-II patients without a melanoma recurrence correctly classified as low risk by DecisionDx-Melanoma®



EP = estimated proportion; CI= confidence interval

eFigure 5. Egger's publication bias plot of the standardized effect estimate* for stage I+II disease by the precision of the estimate

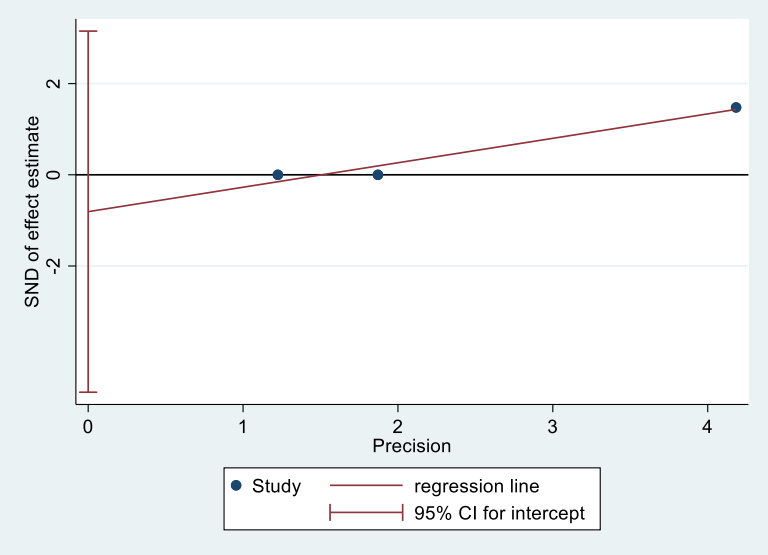


*the proportion of patients with a melanoma recurrence correctly classified as high risk by DecisionDx-Melanoma®

Effect	Coefficient	Std. Err.	t-statistic	p-value	95% CI	
					Lower	Upper
slope	0.470763	0.056515	8.33	0.004	0.290907	0.65062
bias	1.20449	0.264477	4.55	0.02	0.362808	2.046173

p-value of 0.02 suggests that there is evidence of a small study effect

eFigure 6. Egger's publication bias plot of the standardized effect estimate* for stage I disease by the precision of the estimate

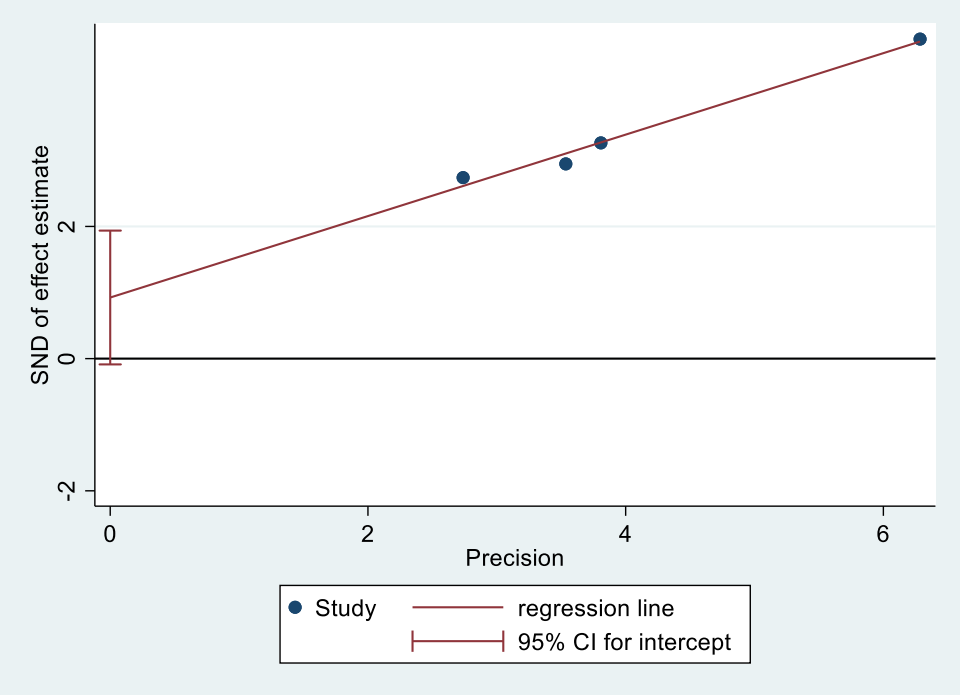


*the proportion of patients with a melanoma recurrence correctly classified as high risk by DecisionDx-Melanoma®

Effect	Coefficient	Std. Err.	t-statistic	p-value	95% CI	
					Lower	Upper
slope	0.5360556	0.1138058	4.71	0.133	-0.9099837	1.982095
bias	-0.8084741	0.3116699	-2.59	0.234	-4.768616	3.151668

p-value of 0.234 suggests that there is little evidence of a small study effect

eFigure 7. Egger's publication bias plot of the standardized effect estimate* for stage II disease by the precision of the estimate

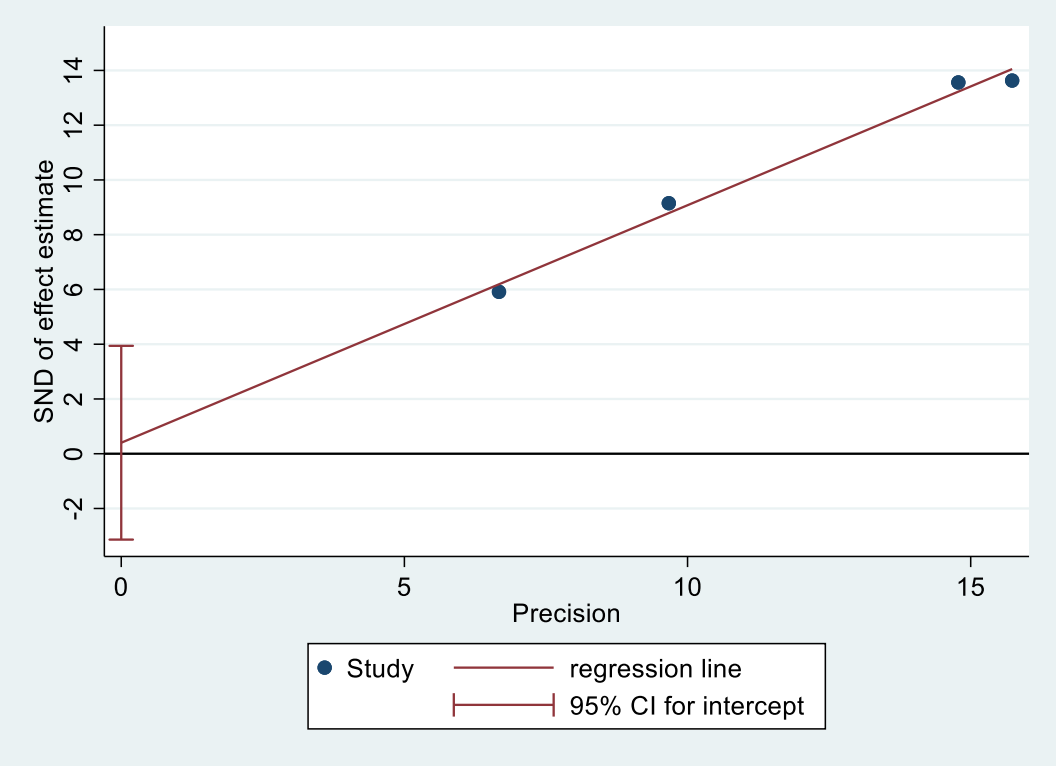


*the proportion of patients with a melanoma recurrence correctly classified as high risk by DecisionDx-Melanoma®

Effect	Coefficient	Std. Err.	t-statistic	p-value	95% CI	
					Lower	Upper
slope	0.616158	0.0546867	11.27	0.008	0.38086	0.8514561
bias	0.9246788	0.2352165	3.93	0.059	-0.0873763	1.936734

p-value of 0.059 suggests that there is little evidence of a small study effect

eFigure 8. Egger's publication bias plot of the standardized effect estimate* for stage I disease by the precision of the estimate

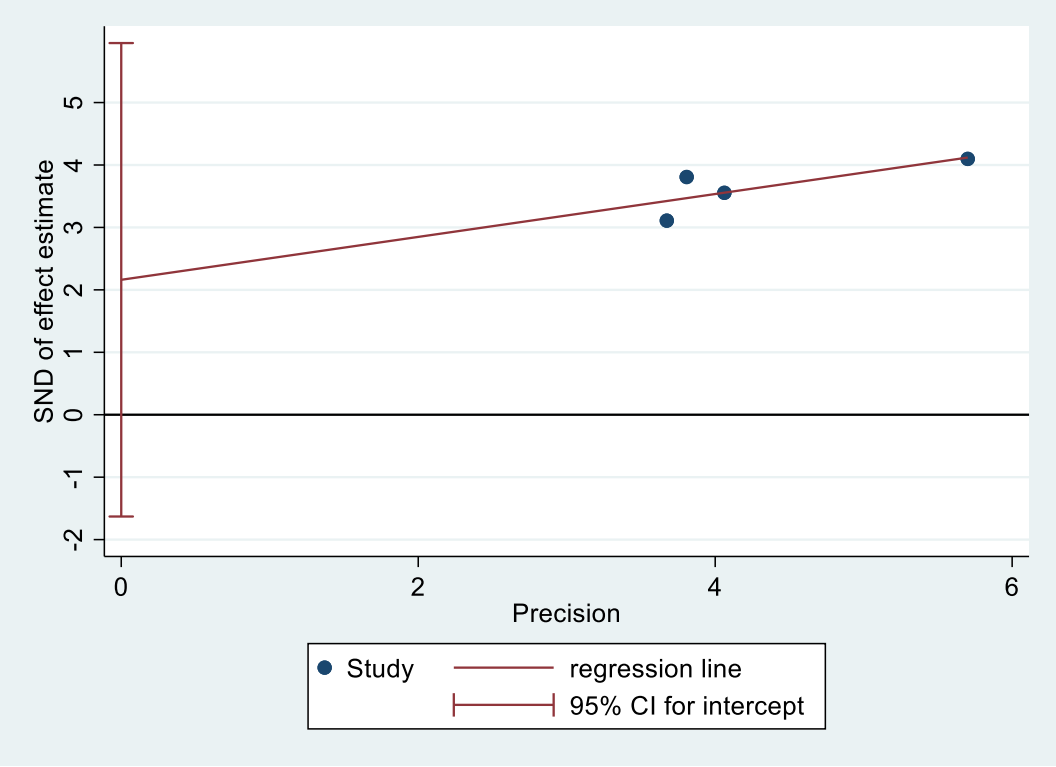


*the proportion of patients without a melanoma recurrence correctly classified as low risk by DecisionDx-Melanoma®

Effect	Coefficient	Std. Err.	t-statistic	p-value	95% CI	
					Lower	Upper
slope	0.8675729	0.0669244	12.96	0.006	0.5796203	1.155526
bias	0.4011196	0.8223812	0.49	0.674	-3.137301	3.93954

p-value of 0.674 suggests that there is no evidence of a small study effect

eFigure 9. Egger's publication bias plot of the standardized effect estimate* for stage II disease by the precision of the estimate

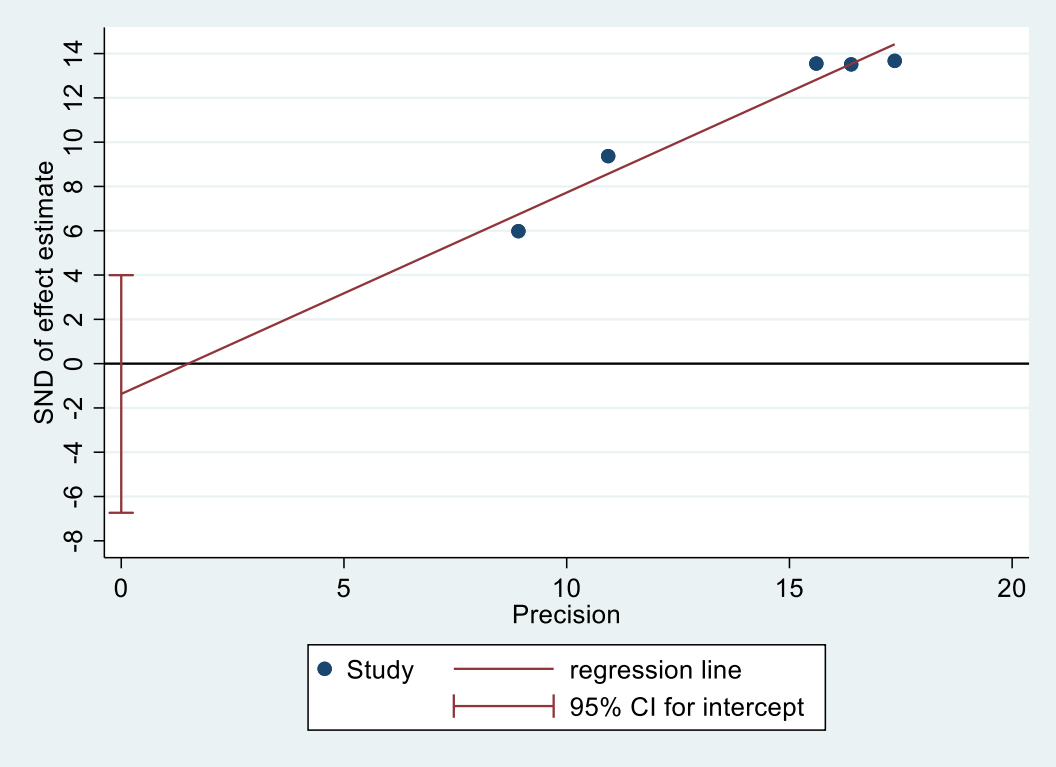


*the proportion of patients without a melanoma recurrence correctly classified as low risk by DecisionDx-Melanoma®

Effect	Coefficient	Std. Err.	t-statistic	p-value	95% CI	
					Lower	Upper
slope	0.3434312	0.2008493	1.71	0.229	-0.5207536	1.207616
bias	2.161538	0.8812227	2.45	0.134	-1.630058	5.953133

p-value of 0.134 suggests that there is little evidence of a small study effect

eFigure 10. Egger's publication bias plot of the standardized effect estimate* for stage I+II disease by the precision of the estimate



*the proportion of patients without a melanoma recurrence correctly classified as low risk by DecisionDx-Melanoma®

Effect	Coefficient	Std. Err.	t-statistic	p-value	95% CI	
					Lower	Upper
slope	0.9094245	0.118446	7.68	0.005	0.5324765	1.286373
bias	-1.36973	1.685516	-0.81	0.476	-6.733794	3.994335

p-value of 0.476 suggests that there is no evidence of a small study effect

eResults. Supplementary Results

DecisionDx-Melanoma®: No stage-specific univariate or multivariate analyses showing an association between the GEP test and a survival outcome were identified; however, using data extracted from Zager et al, we estimated recurrence-free survival (RFS) univariate hazard ratios of 4.01 (95% CI: 1.5-11.5) for stage I disease and 2.5 (95% CI: 1.1-5.5) for stage II disease through five years of follow-up.[13]

Two studies reported an effect estimate for recurrence free survival in mixed stage I/II patients: Greenhaw et al[14] reported a univariate odds ratio of 22.0 (95% CI: 5.7-84.2) and Podlipnik et al[15] reported a multivariate hazard ratio of 18.8 (95% CI: 1.81-2549.8).

The sensitivity and specificity for predicting recurrence at a specified timepoint was estimated from one study.[14] The 3-year and 5-year sensitivity in mixed stage I/II patients was 78% and 73%, respectively; the 3-year and 5-year specificity in mixed stage I/II patients was 79% and 70%, respectively.

MelaGenix®: We identified no univariate or multivariate analyses showing an association between a high-risk GEP result and a survival outcome for stage I patients. Amaral et al[16] reported a multivariate hazards ratio of 1.55 (95% CI 1.13-2.13) with melanoma-specific survival in stage II patients, adjusted for thickness and age. Multivariate analysis for RFS or DMFS was not reported.

eReferences

1. Gershenwald JE, Scolyer RA, Hess KR, et al. Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA: a cancer journal for clinicians.* 2017;67(6):472-492.
2. Marchetti MA, Bartlett EK, Dusza SW, Bichakjian CK. Use of a prognostic gene expression profile test for T1 cutaneous melanoma: Will it help or harm patients? *Journal of the American Academy of Dermatology.* 2019;80(6):e161-e162.
3. Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ (Clinical research ed).* 2019;364:k4597.
4. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of internal medicine.* 2013;158(4):280-286.
5. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC medicine.* 2012;10:51.
6. Huguet A, Hayden JA, Stinson J, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Systematic reviews.* 2013;2:71.
7. Hayden JA, Cote P, Steenstra IA, Bombardier C. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *Journal of clinical epidemiology.* 2008;61(6):552-560.
8. Cochrane Prognosis Methods Group. Tools. Copyright © 2020 The Cochrane Collaboration. https://methods.cochrane.org/prognosis/tools.
9. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ (Clinical research ed).* 2016;352:i6.
10. Vickers AJ. Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. *The American statistician.* 2008;62(4):314-320.
11. Table 8.7.a: Possible approach for summary assessments of the risk of bias for each important outcome (across domains) within and across studies. https://handbook-5-1.cochrane.org/chapter_8/table_8_7_a_possible_approach_for_summary_assessments_of_the.htm. Accessed April 24, 2020.
12. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS medicine.* 2012;9(5):e1001216.
13. Zager JS, Gastman BR, Leachman S, et al. Performance of a prognostic 31-gene expression profile in an independent cohort of 523 cutaneous melanoma patients. *Bmc Cancer.* 2018;18:11.
14. Greenhaw BN, Zitelli JA, Brodland DG. Estimation of Prognosis in Invasive Cutaneous Melanoma: An Independent Study of the Accuracy of a Gene Expression Profile Test. *Dermatol Surg.* 2018;44(12):1494-1500.
15. Podlipnik S, Carrera C, Boada A, et al. Early outcome of a 31-gene expression profile test in 86 AJCC stage IB-II melanoma patients. A prospective multicentre cohort study. *J Eur Acad Dermatol Venereol.* 2019;33(5):857-862.

16. Amaral TMS, Hoffmann MC, Sinnberg T, et al. Clinical validation of a prognostic 11-gene expression profiling score in prospectively collected FFPE tissue of patients with AJCC v8 stage II cutaneous melanoma. *European journal of cancer (Oxford, England : 1990).* 2020;125:38-45.