Title: Gene Expression Profiling for Cutaneous Melanoma

State and Federal mandates and health plan member contract language, including specific provisions/exclusions, take precedence over Medical Policy and must be considered first in determining eligibility for coverage. To verify a member’s benefits, contact Blue Cross and Blue Shield of Kansas Customer Service.

The BCBSKS Medical Policies contained herein are for informational purposes and apply only to members who have health insurance through BCBSKS or who are covered by a self-insured group plan administered by BCBSKS. Medical Policy for FEP members is subject to FEP medical policy which may differ from BCBSKS Medical Policy.

The medical policies do not constitute medical advice or medical care. Treating health care providers are independent contractors and are neither employees nor agents of Blue Cross and Blue Shield of Kansas and are solely responsible for diagnosis, treatment and medical advice.

If your patient is covered under a different Blue Cross and Blue Shield plan, please refer to the Medical Policies of that plan.

<table>
<thead>
<tr>
<th>Populations</th>
<th>Interventions</th>
<th>Comparators</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals: • With suspicious pigmented lesions (based on ABCDE and/or ugly duckling criteria) being considered for biopsy</td>
<td>Interventions of interest are: • Gene expression profiling with the DermTech Pigmented Lesion Assay to determine which lesions should proceed to biopsy</td>
<td>Comparators of interest are: • Dermatology exam and dermoscopy</td>
<td>Relevant outcomes include: • Overall survival • Disease-specific survival • Test accuracy • Test validity • Resource utilization</td>
</tr>
<tr>
<td>Individuals: • Who have melanocytic lesions with indeterminate histopathologic features</td>
<td>Interventions of interest are: • Gene expression profiling with the myPath Melanoma test added to histopathology to aid in diagnosis of melanoma</td>
<td>Comparators of interest are: • Histopathology alone</td>
<td>Relevant outcomes include: • Overall survival • Disease-specific survival • Test accuracy • Test validity • Change in disease status • Treatment-related morbidity</td>
</tr>
<tr>
<td>Populations</td>
<td>Interventions</td>
<td>Comparators</td>
<td>Outcomes</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Individuals:</td>
<td>Interventions of interest are:</td>
<td>Comparators of interest are:</td>
<td>Relevant outcomes include:</td>
</tr>
<tr>
<td>• With American Joint Committee on Cancer</td>
<td>• Gene expression profiling with the DecisionDx-Melanoma test to inform</td>
<td>• Sentinel lymph node biopsy</td>
<td>• Overall survival</td>
</tr>
<tr>
<td>stage I or II cutaneous melanoma</td>
<td>management decisions regarding enhanced surveillance</td>
<td>• Prognostic tools</td>
<td>• Disease-specific survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Test accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Test validity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Change in disease status</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Resource utilization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Treatment-related morbidity</td>
</tr>
<tr>
<td>Individuals:</td>
<td>Interventions of interest are:</td>
<td>Comparators of interest are:</td>
<td>Relevant outcomes include:</td>
</tr>
<tr>
<td>• With American Joint Committee on Cancer</td>
<td>• Gene expression profiling with the DecisionDx-Melanoma test to inform</td>
<td>• Sentinel lymph node biopsy</td>
<td>• Overall survival</td>
</tr>
<tr>
<td>stage I or II cutaneous melanoma</td>
<td>management decisions regarding adjuvant therapy</td>
<td>• Prognostic tools</td>
<td>• Disease-specific survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Test validity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Change in disease status</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Resource utilization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Treatment-related morbidity</td>
</tr>
<tr>
<td>Individuals:</td>
<td>Interventions of interest are:</td>
<td>Comparators of interest are:</td>
<td>Relevant outcomes include:</td>
</tr>
<tr>
<td>• With cutaneous melanoma</td>
<td>• Gene expression profiling with the DecisionDx-Melanoma test to identify</td>
<td>• Sentinel lymph node biopsy</td>
<td>• Overall survival</td>
</tr>
<tr>
<td>• with clinically negative sentinel node</td>
<td>patients who can avoid sentinel lymph node biopsy</td>
<td>• Prognostic tools</td>
<td>• Disease-specific survival</td>
</tr>
<tr>
<td>basins who are being considered for sentinel lymph node biopsy</td>
<td></td>
<td></td>
<td>• Test validity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Change in disease status</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Resource utilization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Treatment-related morbidity</td>
</tr>
</tbody>
</table>

**DESCRIPTION**

Laboratory tests have been developed that detect the expression of different genes in pigmented lesions or melanoma tumor tissue. Test results may help providers and patients decide whether to biopsy suspicious pigmented lesions, aid in diagnosis lesions with indeterminate histopathologic lesions or determine whether to perform sentinel lymph node biopsy in patients diagnosed with stage I or II cutaneous melanoma. This report summarizes the evidence of 3 tests.

**OBJECTIVE**

The objective of this policy is to determine whether gene expression profiling improves the net health outcome in individuals with lesions suspicious for melanoma or with melanoma.

**BACKGROUND**

**Cutaneous Melanoma**

Cutaneous melanoma accounts for more than 90% of cases of melanoma. For many decades, melanoma incidence was rapidly increasing in the United States. However, recent estimates have suggested the rise may be slowing. In 2018, more than 90,000 new cases of melanoma are expected to be diagnosed, and more than 9000 people are expected to die of melanoma.
Risk Factors
Exposure to solar ultraviolet radiation is a major risk factor for melanoma. Most melanomas occur on the sun-exposed skin, particularly those areas most susceptible to sunburn. Likewise, features that are associated with an individual’s sensitivity to sunlight, such as light skin pigmentation, red or blond hair, blue or green eyes, freckling tendency, and poor tanning ability are well-known risk factors for melanoma.3,4 There is also a strong association between high total body nevus counts and melanoma.5

Several genes appear to contribute to melanoma predisposition such as tumor suppressor gene CDKN2A, melanocortin-1 receptor (MC1R) gene, and BAP1 variants.6-8 Individuals with either familial or sporadic melanoma have a 2 to 3 times increased risk of developing a subsequent primary melanoma.9 Several occupational exposures and lifestyle factors, such as body mass index and smoking, have been evaluated as possible risk factors for melanoma.10

Gene Expression Profiling
Gene expression profiling measures the activity of thousands of genes simultaneously and creates a snapshot of cellular function. Data for gene expression profiles are generated by several molecular technologies including DNA microarrays that measures activity relative to previously identified genes and RNA-Seq that directly sequences and quantifies RNA molecules. Clinical applications of gene expression profiling include disease diagnosis, disease classification, prediction of drug response, and prognosis.

REGULATORY STATUS
Clinical laboratories may develop and validate tests in-house and market them as a laboratory service; laboratory-developed tests must meet the general regulatory standards of the Clinical Laboratory Improvement Amendments. The Pigmented Lesion Assay, myPath Melanoma, and DecisionDx-Melanoma tests are available under the auspices of the Clinical Laboratory Improvement Amendments. Laboratories that offer laboratory-developed tests must be licensed by the Clinical Laboratory Improvement Amendments for high-complexity testing. To date, the U.S. Food and Drug Administration has chosen not to require any regulatory review of this test.
**POLICY**

A. Gene expression testing, including, but not limited to, the Pigmented Lesion Assay in the evaluation of patients with suspicious pigmented lesions is considered experimental / investigational.

B. Gene expression testing, including, but not limited to, the myPath Melanoma test in the evaluation of patients with melanocytic lesions with indeterminate histopathologic features is considered experimental / investigational.

C. Gene expression testing, including, but not limited to, DecisionDx-Melanoma in the evaluation of patients with cutaneous melanoma is considered experimental / investigational for all indications.

**Policy Guidelines**

1. **GENETIC COUNSELING:** Experts recommend formal genetic counseling for patients who are at risk for inherited disorders and who wish to undergo genetic testing. Interpreting the results of genetic tests and understanding risk factors can be difficult for some patients; genetic counseling helps individuals understand the impact of genetic testing, including the possible effects the test results could have on the individual or their family members. It should be noted that genetic counseling may alter the utilization of genetic testing substantially and may reduce inappropriate testing; further, genetic counseling should be performed by an individual with experience and expertise in genetic medicine and genetic testing methods.

**RATIONALE**

The literature review was performed through November 1, 2018.

Evidence reviews assess whether a medical test is clinically useful. A useful test provides information to make a clinical management decision that improves the net health outcome. That is, the balance of benefits and harms is better when the test is used to manage the condition than when another test or no test is used to manage the condition.

The first step in assessing a medical test is to formulate the clinical context and purpose of the test. The test must be technically reliable, clinically valid, and clinically useful for that purpose. Evidence reviews assess the evidence on whether a test is clinically valid and clinically useful. Technical reliability is outside the scope of these reviews, and credible information on technical reliability is available from other sources.

**Gene Expression Profiling to Guide Initial Biopsy Decisions**

**Clinical Context and Test Purpose**

Primary care providers evaluate suspicious pigmented lesions to determine who should be referred to dermatology. Factors considered include both a patient’s risk for melanoma as well as
a visual examination of the lesion. The visual examination assesses whether the lesion has features suggestive of melanoma.

Criteria for features suggestive of melanoma have been developed. One checklist is the ABCDE checklist\textsuperscript{11}:

- **A**symmetry;
- **B**order irregularities;
- **C**olor variegation;
- **D**iameter ≥6 mm;
- **E**volution.

Another criteria commonly used is the “ugly duckling” sign.\textsuperscript{12} An ugly duckling is a nevus that is obviously different from others in a given patient. Primary care providers generally have a low threshold for referral to dermatology.

Melanoma is difficult to diagnose based on visual examination, and the criterion standard for diagnosis is histopathology. There is a low threshold for excisional biopsy of suspicious lesions for histopathologic examination due to the procedure’s ease and low-risk as well as the high probability of missing melanoma. However, the yield of biopsy is fairly low. The number of biopsies performed to yield one melanoma diagnosis has been estimated to be about 15 for U.S. dermatologists.\textsuperscript{13} Therefore a test that could accurately identify those lesions not needing a biopsy (ie, a rule-out test for biopsy) could be clinically useful.

The purpose of GEP in patients who have suspicious pigmented lesions being considered for biopsy is to inform a decision about whether to biopsy.

The question addressed in this section of the evidence review is: Does GEP improve the net health outcome in individuals with suspicious pigmented lesions?

The following PICOTS were used to select literature to inform this review.

**Patients**
The relevant population of interest are patients with suspicious pigmented lesions being considered for referral for biopsy, specifically those lesions meeting one or more ABCDE criteria.

**Interventions**
The test being considered is the DermTech Pigmented Lesion Assay (PLA). The PLA test measures expression of six genes (PRAME, LINC00518, CMIP, B2M, ACTB, PPIA). The PRAME (PReferentially expressed Antigen in MElanoma) gene encodes an antigen that is preferentially expressed in human melanomas, and that is not expressed in normal tissues (except testis).\textsuperscript{14} LINC00518 (Long Intergenic Non-protein Coding RNA518) is a regulatory RNA molecule. The other four genes provide normalization values.\textsuperscript{15} The feasibility of a test like PLA was first described in Wachsman et al (2011) and Gerami et al (2014).\textsuperscript{16,17} and development of the specific PLA test was described in Gerami et al (2017).\textsuperscript{18}

The test is performed on skin samples of lesions at least 5 mm in diameter obtained via noninvasive, proprietary adhesive patch biopsies of a stratum corneum specimen. The test does
not work on the palms of hands, soles of feet, nails, or mucous membranes, and it should not be used on bleeding or ulcerated lesions. The PLA test report includes two results. The first result is called the PLA MAGE (Melanoma Associated Gene Expression), which indicates low-risk (neither PRAME nor LINC00518 expression was detected), moderate-risk (expression of either PRAME or LINC00518 was detected), or high-risk (expression of both PRAME and LINC00518 was detected). The second result is as an algorithmic PLA score that ranges from 0 to 100, with higher scores indicating higher suspicion of malignant disease.

It is not clear whether the PLA test is meant to be used as a replacement, triage, or add-on test with respect to dermoscopy. The PLA sample report states that for low-risk lesions, physicians should “consider surveillance,” while for moderate- and high-risk lesions, physicians should “recommend a biopsy.” It does not state whether lesions with negative results should be further evaluated with dermoscopy or other techniques to confirm the lesion should not be biopsied. Therefore, this evidence review evaluates the test as a replacement for dermoscopy. As mentioned previously, there is a low threshold for biopsy of suspicious lesions. As such, tests that can rule-out need for biopsy could be useful and thus sensitivity and negative predictive value are the performance characteristics of most interest.

Comparators

After a referral from primary care to dermatology settings, dermatologists use visual examination as well as tools such as dermoscopy to make decisions regarding biopsy of suspicious lesions. A meta-analysis of 9 studies (8487 lesions with 375 melanomas) compared dermoscopy with visual examination alone for the diagnosis of melanoma; it reported that, for clinicians with training in dermoscopy, adding dermoscopy to visual examination increased the sensitivity from 71% to 90%. The specificity numerically increased from 80% to 90%, but the difference was not statistically significant. Although dermoscopy is noninvasive and may aid in decision making regarding biopsy, it is only used by approximately 50% to 80% of dermatologists in the U. S. due to lack of training, interest, or time required for the examination.

The reference standard for diagnosis of melanoma is histopathology.

Outcomes

The beneficial outcomes of a true-positive test result are appropriate biopsy and diagnosis of melanoma. The beneficial outcome of a true-negative test result is potentially avoiding unnecessary biopsy.

The harmful outcome of a false-positive result is having an unnecessary biopsy. The harmful outcome of a false-negative result is potential delay in diagnosis and treatment.

The timeframe of interest for calculating performance characteristics is time to biopsy result. Patients who forgo biopsy based on test results could miss or delay diagnosis of cancer. Longer follow-up would be necessary to determine the effects on overall survival (OS).

Technically Reliable

Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and
unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

**Clinically Valid**
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Determining whether a test can guide biopsy decisions is not based only on its sensitivity and specificity, but also on how the accuracy of the existing pathway for making biopsy decisions is changed by the test. Therefore, the appropriate design for evaluating performance characteristics depends on the role of the new test in the pathway for making biopsy decisions. New tests may be used as replacements for existing tests, to triage who proceeds for existing tests or add-on tests after existing tests. For replacement tests, the diagnostic accuracy of both tests should be concurrently compared, preferably in a paired design (ie, patients receive both tests), and all patients receive the reference standard. For a triage test, a paired design is also needed, with the reference standard being performed preferably on all patients but at least for all discordant results. For an add-on test, the included patients can be limited to those who were negative after existing tests with verification of the reference standard in patients who are positive on the new test.22

**Study Selection Criteria**
For the evaluation of clinical validity of the PLA test, studies that meet the following eligibility criteria were considered:
- Reported on a validation cohort that was independent of the development cohort;
- Reported on the accuracy of the marketed version of the technology;
- Included a suitable reference standard (histopathology);
- Patient/sample clinical characteristics were described;
- Patient/sample selection criteria were described.

Studies were excluded from the evaluation of the clinical validity of the PLA test because they reported results of the development cohort,17 they did not use the marketed version of the test,16,17 did not include the reference standard test on PLA negative patients,23 did not adequately describe the patient characteristics,24 or did not adequately describe patient selection criteria.24

The validation cohort from the Gerami et al (2017) publication was included.18 The study characteristics are described in Table 1. The report stated that included lesions were selected by dermatologists experienced in pigmented lesion management from 28 sites in the U. S., Europe, and Australia; therefore, the samples were likely not consecutive or random. Information regarding the previous testing was not provided. The flow of potential and included samples was not clear, and whether the samples were all independent or, multiple samples from the same patient were not described. Diagnosis of melanoma was based on consensus among a primary reader and three expert dermatopathologists. The report did not state whether the histopathologic diagnosis was blinded to the results of the PLA test but did state the diagnosis was “routinely” assessed. Interpretation of the PLA result does not depend on a reader, so it is blinded to histopathologic results. In 11% of cases originally selected, a consensus diagnosis was not reached, and these samples were not included in the training or validation cohorts. Dates of data collection were not reported. Sex and anatomic location of biopsy were reported, but other
clinical characteristics (eg, risk factors for melanoma, presenting symptoms) were not. Study results are shown in Table 2. The study training cohort included 157 samples with 80 melanomas and 77 non-melanomas. The study validation cohort included 398 samples with 87 melanomas and 311 non-melanomas. Study relevance, design, and conduct gaps are in Tables 3 and 4.

**Section Summary: PLA Clinical Validity**

Multiple high-quality studies are needed to establish the clinical validity of a test. The PLA test has one clinical validity study with many methodologic and reporting limitations. Therefore, performance characteristics are not well-characterized. Also, the test has not been compared with dermoscopy, another tool frequently used to make biopsy decisions.

<table>
<thead>
<tr>
<th>Study</th>
<th>Initial N</th>
<th>Final N</th>
<th>Excluded Samples</th>
<th>Melanoma Prevalence</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerami et al (2017)</td>
<td>398</td>
<td>398</td>
<td>Before allocation to training and validation cohorts, 11% of original samples excluded due to lack of consensus diagnosis</td>
<td>22%</td>
<td>91 (83 to 96)</td>
<td>69 (64 to 74)</td>
<td>45 (38 to 53)c</td>
<td>96 (93 to 98)c</td>
</tr>
</tbody>
</table>

Dx: diagnosis; PCR: polymerase chain reaction; PLA: Pigmented Lesion Assay.

<table>
<thead>
<tr>
<th>Study</th>
<th>Initial N</th>
<th>Final N</th>
<th>Excluded Samples</th>
<th>Melanoma Prevalence</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerami et al (2017)</td>
<td>398a</td>
<td>398</td>
<td>Before allocation to training and validation cohorts, 11% of original samples excluded due to lack of consensus diagnosis</td>
<td>22%</td>
<td>91 (83 to 96)</td>
<td>69 (64 to 74)</td>
<td>45 (38 to 53)c</td>
<td>96 (93 to 98)c</td>
</tr>
</tbody>
</table>

a 398 samples were included in the validation cohort; the number of independent patients is unclear.
b Values are percentages with 95% confidence interval.
c Confidence intervals provided in the report; calculated from data provided.
Table 3. Clinical Validity Study Relevance Gaps of the PLA Test

<table>
<thead>
<tr>
<th>Study</th>
<th>Population&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Intervention&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Comparator&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Outcomes&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Duration of Follow-Up&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerami et al (2017)&lt;sup&gt;18, 19&lt;/sup&gt;</td>
<td>3. Study population characteristics not adequately described</td>
<td>3. No comparison to dermoscopy</td>
<td>3. Predictive values were not reported but were calculated based on data provided</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.

PLA: Pigmented Lesion Assay.

<sup>a</sup> Population key: 1. Intended use population unclear; 2. Clinical context is unclear; 3. Study population is unclear; 4. Study population not representative of intended use.

<sup>b</sup> Intervention key: 1. Classification thresholds not defined; 2. Version used unclear; 3. Not intervention of interest.

<sup>c</sup> Comparator key: 1. Classification thresholds not defined; 2. Not compared to credible reference standard; 3. Not compared to other tests in use for same purpose.

<sup>d</sup> Outcomes key: 1. Study does not directly assess a key health outcome; 2. Evidence chain or decision model not explicated; 3. Key clinical validity outcomes not reported (sensitivity, specificity and predictive values); 4. Reclassification of diagnostic or risk categories not reported; 5. Adverse events of the test not described (excluding minor discomforts and inconvenience of venipuncture or noninvasive tests).

<sup>e</sup> Follow-Up key: 1. Follow-up duration not sufficient with respect to natural history of disease (true-positives, true-negatives, false-positives, false-negatives cannot be determined).

Table 4. Clinical Validity Study Design and Conduct Gaps of the PLA Test

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Blinding&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Delivery of Test&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Selective Reporting&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Completeness of Follow-Up&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Statistical&lt;sup&gt;f&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerami et al (2017)&lt;sup&gt;18, 19&lt;/sup&gt;</td>
<td>1,2. Not clear what criteria used to select samples but it does not appear to have been random or consecutive</td>
<td>1. Blinding of histopathology readers not described</td>
<td>1. Patch biopsy administered before surgical biopsy but timing between procedures not described</td>
<td>1. No registration reported</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.

PLA: Pigmented Lesion Assay.

<sup>a</sup> Selection key: 1. Selection not described; 2. Selection not random or consecutive (ie, convenience).

<sup>b</sup> Blinding key: 1. Not blinded to results of reference or other comparator tests.

<sup>c</sup> Test Delivery key: 1. Timing of delivery of index or reference test not described; 2. Timing of index and comparator tests not same; 3. Procedure for interpreting tests not described; 4. Expertise of evaluators not described.

<sup>d</sup> Selective Reporting key: 1. Not registered; 2. Evidence of selective reporting; 3. Evidence of selective publication.

<sup>e</sup> Follow-Up key: 1. Inadequate description of indeterminate and missing samples; 2. High number of samples excluded; 3. High loss to follow-up or missing data.

<sup>f</sup> Statistical key: 1. Confidence intervals and/or p values not reported; 2. Comparison to other tests not reported.

Clinically Useful

A test is clinically useful if the results inform management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence

Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials (RCTs).

No direct evidence of clinical utility was identified.
Chain of Evidence
Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility through a chain of evidence.

A decision-impact study by Ferris et al (2017) assessed the potential impact of PLA on physicians’ biopsy decisions in patients. Forty-five dermatologists evaluated 60 clinical and dermoscopic images of atypical pigmented lesions (8 melanoma, 52 nonmelanoma). In the first round, dermatologists did not have PLA test results and, in the second round, dermatologists had access to PLA test results with the order of cases being scrambled. The dermatologists were asked whether the lesions should be biopsied after each round. Therefore, the corresponding number of biopsy decisions should be $45 \times 60 \times 2 = 5400$. Data were collected in 2014 and 2015. Results were reported for 4680 decisions with no description of the disposition of the remaining decisions. Of the 4680 reported decisions, 750 correct biopsy decisions were made without PLA results while 1331 were made with PLA results and 1590 incorrect biopsy decisions were made without PLA results while 1009 incorrect biopsy decisions were made with PLA results.

Section Summary: Clinically Useful
There is no direct evidence of clinical utility. A chain of evidence for clinical utility cannot be constructed due to lack of robust evidence of clinical validity.

GEP for Diagnosing Lesions with Indeterminate Histopathology
Clinical Context and Test Purpose
The diagnosis of melanoma was described in the previous section. The diagnosis of melanoma is histopathologic and when the histopathologic diagnosis is straightforward, ancillary methods such as comparative genomic hybridization, florescence in situ hybridization (FISH), and GEP are not recommended. Therefore, the usefulness of an ancillary test is its ability to predict biologic behavior (metastasis) of lesions that are indeterminate by histopathology.

The purpose of GEP in patients whose melanocytic lesion is indeterminate after histopathology is to aid in the diagnosis of melanoma and decisions regarding treatment and surveillance.

The question addressed in this section of the evidence review is: Does GEP improve the net health outcome in individuals with indeterminate melanocytic lesions?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest are patients whose melanocytic lesion is indeterminate based on clinical and histopathologic features.

Interventions
The test being considered is the Myriad myPath Melanoma test. The myPath test measures expression of 23 genes using quantitative reverse-transcription polymerase chain reaction. Fourteen genes are involved in melanoma pathogenesis and are grouped into three components related to cell differentiation, cell signaling, and the immune response, and nine housekeeper genes are also included. The test is performed on five standard tissue sections from an existing formalin-fixed, paraffin-embedded biopsy specimen.
The myPath test report includes an algorithmic myPath score ranging from -16.7 to 11.1, with higher, positive scores indicating higher suspicion of malignant disease. The myPath report also classifies these scores: -16.7 to -2.1 are "benign"; -2.0 to -0.1 are "indeterminate"; and 0.0 to +11.1 are "malignant". Development of the test has been described by Clarke et al (2015). The myPath test is meant as an add-on test to standard histopathology. Studies have evaluated the performance characteristics of the test when histopathology is used as the reference standard, but are not the focus of this evidence review given that the test's potential usefulness is in evaluation of indeterminate lesions.

No recommendations for treatment or surveillance are given on the report.

Comparators
The reference standard for diagnosis of melanoma is histopathology. However, in cases of indeterminate histopathology, long-term follow-up is needed to evaluate the clinical outcome, specifically metastasis.

Comparative genomic hybridization and FISH are also used to diagnosis indeterminate lesions although neither has been fully validated. FISH has been evaluated as a tool to aid in the diagnosis of lesions that are indeterminate, following histopathology in two studies that included histologically ambiguous lesions and a clinical, long-term follow-up. One study reported by Gaiser et al (2010) included 22 melanocytic lesions (12 indeterminate) followed for a mean of 65 months (range, 10-156 months) and reported a FISH sensitivity of 60% and a specificity of 50% for development of metastases during follow-up. A second study, reported by Vergier et al (2011), included 90 indeterminate melanocytic lesions of which 69 had no recurrence for at least 5 years of follow-up (mean, 9 years; range, 5-19 years) and 21 lesions that exhibited metastases. The sensitivity and specificity rates of the histopathologic review combined with FISH for the clinical outcome were 76% and 90%, respectively.

Outcomes
The beneficial outcomes of a true-positive test result are a diagnosis of melanoma and corresponding appropriate treatment and surveillance. The beneficial outcome of a true-negative test result is avoiding unnecessary surgery.

The harmful outcome of a false-positive result is having an unnecessary surgery and surveillance. The harmful outcome of a false-negative result is a delay in diagnosis and treatment.

The National Comprehensive Cancer Network guidelines state that even in the presence of node metastasis, indeterminate neoplasms can demonstrate benign biologic behavior, making it difficult to define a fully malignant lesion and also states that events in the group of indeterminate lesions tend to occur late. Therefore, the guidelines suggest that long-term follow-up is necessary to validate a test for this purpose.

Recurrence and metastases can occur may years after treatment of melanoma. In the two studies evaluating long-term outcomes of FISH (described above), the mean follow-up was approximately 5.5 and nine years. In Vergier et al (2011), metastases in the FISH-negative group generally occurred by 5 years.
For this section of the review, at least five years of event-free follow-up is required to confirm negative tests. The event of interest is metastasis.

**Technically Reliable**
Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

**Clinically Valid**
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

**Study Selection Criteria**
For the evaluation of clinical validity of the myPath test, studies that meet the following eligibility criteria were considered:

- Reported on a validation cohort that was independent of the development cohort;
- Reported on the accuracy of the marketed version of the technology;
- Included a suitable reference standard (clinical outcome with at least five years of follow-up for negatives);
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described.

Studies were excluded from the evaluation of the clinical validity of the myPath test because authors did not use the specified reference standard of long-term (at least five years) follow-up and/or did not adequately describe patient characteristics.

The Ko et al (2017) clinical validity study met selection criteria. The study characteristics are described in Table 5. In Ko et al (2017), archived melanocytic neoplasms were submitted for myPath testing from university clinics in the U. S. and United Kingdom with additional samples acquired from Avaden BioSciences. Stage I, II, and III primary cutaneous melanomas that produced distant metastases subsequent to the diagnosis and benign lesions with clinical follow-up and no evidence of recurrence of metastases were included. For benign samples, a disease-free time of at least five years was recommended. Information on the previous testing was not provided. It is not clear if any of the samples originally had indeterminate histopathology results. Dates of data collection were not reported. Sex, age, Breslow depth, and anatomic location were described; presenting symptoms were not reported. A total of 293 samples were submitted; of these 53 did not meet inclusion criteria and 58 (24% of those tested) failed to produce a valid test score. An additional seven samples with indeterminate results were excluded from the calculations of performance characteristics.

Study results are shown in Table 6. Study relevance, design, and conduct gaps are in Tables 7 and 8.
### Table 5. Clinical Validity Study Characteristics of the myPath Test for Predicting Metastasis

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Population</th>
<th>Design</th>
<th>Reference Standard</th>
<th>Threshold Score for Positive myPath Test</th>
<th>Timing of Reference and myPath Tests</th>
<th>Blinding of Assessors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ko et al (2017)</td>
<td>· Primary cutaneous melanomas or benign melanocytic nevi · Mean age, 53 y · 55% of samples from men</td>
<td>Retrospective · Not consecutive or randomly selected</td>
<td>· Positive: malignant lesions that produced distant metastases · Negative Event-free follow-up, recommended 5 y (median, 6.2 y)</td>
<td>· Scores from 0.0 to 11.1 (ie, &quot;malignant&quot;)</td>
<td>· Final clinical diagnosis established before myPath test · Length of time between biopsy and myPath test unclear</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Table 6. Clinical Validity Study Results of the myPath Test for Predicting Metastasis

<table>
<thead>
<tr>
<th>Study</th>
<th>Initial N</th>
<th>Final N</th>
<th>Excluded Samples</th>
<th>Melanoma Prevalence</th>
<th>Sensitivity&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Specificity&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PPV&lt;sup&gt;a&lt;/sup&gt;</th>
<th>NPV&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ko et al (2017)</td>
<td>240</td>
<td>175</td>
<td>· 58 failed to produce test result · 7 with indeterminate results</td>
<td>54</td>
<td>94 (87 to 98)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>96 (89 to 99)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>97 (91 to 99)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>93 (85 to 97)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

NPV: negative predictive value; PPV: positive predictive value.
<sup>a</sup> Values are percentages with 95% confidence interval.
<sup>b</sup> Confidence intervals not provided in the report; calculated from data provided.

### Table 7. Clinical Validity Study Relevance Gaps of the myPath Test

<table>
<thead>
<tr>
<th>Study</th>
<th>Population&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Intervention&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Comparator&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Outcomes&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Duration of Follow-Up&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ko et al (2017)</td>
<td>4. Study population is not limited to lesions that are indeterminate following histopathology</td>
<td>None noted</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.
<sup>a</sup> Population key: 1. Intended use population unclear; 2. Clinical context is unclear; 3. Study population is unclear; 4. Study population not representative of intended use.
<sup>b</sup> Intervention key: 1. Classification thresholds not defined; 2. Version used unclear; 3. Not intervention of interest.
<sup>c</sup> Comparator key: 1. Classification thresholds not defined; 2. Not compared to credible reference standard; 3. Not compared to other tests in use for same purpose.
<sup>d</sup> Outcomes key: 1. Study does not directly assess a key health outcome; 2. Evidence chain or decision model not explicated; 3. Key clinical validity outcomes not reported (sensitivity, specificity and predictive values); 4. Reclassification of diagnostic or risk categories not reported; 5. Adverse events of the test not described (excluding minor discomforts and inconvenience of venipuncture or noninvasive tests).
<sup>e</sup> Follow-Up key: 1. Follow-up duration not sufficient with respect to natural history of disease (true positives, true negatives, false positives, false negatives cannot be determined).

### Table 8. Clinical Validity Study Design and Conduct Gaps of the myPath Test

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Blinding&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Delivery of Test&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Selective Reporting&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Completeness of Follow-Up&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Statistical&lt;sup&gt;f&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ko et al (2017)</td>
<td>2. Samples not consecutive or random</td>
<td>1. Unclear how much time elapsed between</td>
<td>1. No registration reported</td>
<td>2. More than 25% of samples tested did not produce results or produced indeterminate results</td>
<td>1. CIs for sensitivity and specificity not reported but were calculated based on data provided. NPV,</td>
<td></td>
</tr>
</tbody>
</table>
Table 1: Key Methodologic Parameters and Performance Characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection</th>
<th>Blinding</th>
<th>Delivery of Test</th>
<th>Selective Reporting</th>
<th>Completeness of Follow-Up</th>
<th>Statistical</th>
</tr>
</thead>
<tbody>
<tr>
<td>biopsy and myPath test</td>
<td>PPV were not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment.

CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

a Selection key: 1. Selection not described; 2. Selection not random or consecutive (ie, convenience).

b Blinding key: 1. Not blinded to results of reference or other comparator tests.

c Test Delivery key: 1. Timing of delivery of index or reference test not described; 2. Timing of index and comparator tests not same; 3. Procedure for interpreting tests not described; 4. Expertise of evaluators not described.


e Follow-Up key: 1. Inadequate description of indeterminate and missing samples; 2. High number of samples excluded; 3. High loss to follow-up or missing data.

f Statistical key: 1. Confidence intervals and/or p values not reported; 2. Comparison to other tests not reported.

Section Summary: Clinically Valid

Multiple high-quality studies are needed to establish the clinical validity of a test. The myPath test has one clinical validity study including long-term follow-up for metastasis as the reference standard. However, it is not clear whether the study population included lesions that were indeterminate following histopathology and the study had other methodologic and reporting limitations. Therefore, performance characteristics are not well-characterized.

Clinically Useful

A test is clinically useful if the results inform management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence

Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

No direct evidence of clinical utility was identified.

Chain of Evidence

Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

Two decision-impact studies assessed the potential impact of myPath on physicians’ treatment decisions in patients with diagnostically challenging lesions.35,36 Given the lack of health outcomes, it is not known whether any treatment changes were clinically appropriate.

Section Summary: Clinically Useful

There is no direct evidence of clinical utility. A chain of evidence for clinical utility cannot be constructed due to lack of robust evidence of clinical validity.

GEP to Guide Management Decisions in Melanoma

Clinical Context and Test Purpose

Many treatments and surveillance decisions are determined by a patient’s prognostic stage group based the American Joint Committee on Cancer tumor, node, metastasis staging system.37 The prognostic groups are as follows: stage I, T1a through T2a primary melanomas without evidence...
of regional or distant metastases; stage II, T2b through T4b primary melanomas without evidence of lymphatic disease or distant metastases; stage III: pathologically documented involvement of regional lymph nodes or in transit or satellite metastases (N1 to N3); stage IV: distant metastases. Patients may also undergo sentinel lymph node biopsy to gain more definitive information about the status of the regional nodes.

Wide local excision is the definitive surgical treatment of melanoma. Following surgery, patients with American Joint Committee on Cancer stage I or II (node-negative) melanoma do not generally receive adjuvant therapy. Patients with higher risk melanoma receive adjuvant immunotherapy or targeted therapy. Ipilimumab has been shown to prolong recurrence-free survival by approximately 25% compared with placebo at a median of 5.3 years in patients with resected, stage III disease. Nivolumab has been shown to further prolong survival compared with ipilimumab by approximately 35% at 18 months. For patients who are BRAF V600 variant-positive with stage III melanoma, the combination of dabrafenib plus trametinib has been estimated to prolong relapse-free survival by approximately 50% over 3 years.

Patients with stage I and II disease should undergo an annual routine physical and dermatologic examination. However, follow-up strategies and intervals have not been standardized or tested, and there is no consensus. These patients typically do not receive surveillance imaging. Patients with stage III melanoma may be managed with more frequent follow-up and imaging surveillance following therapy.

The purpose of GEP in patients with melanoma is to identify low and high-risk patients classified as stage I or II according to the AJCC criteria. Current guidelines do not recommend adjuvant therapy or imaging surveillance for AJCC stage I or II patients following surgery. Patients initially staged as I or II who have positive lymph nodes following sentinel lymph node biopsy (SLNB) are then eligible to be treated with adjuvant therapy as stage III patients.

At least three uses for the test have been suggested. The manufacturer’s website has suggested that physicians can use DecisionDx-Melanoma information to “consider upstaging” patients for “active systemic surveillance or referral to medical oncology for consideration of systemic drug therapy or clinical trials.” Similarly, in one clinical validity study (described below), the authors stated that “high-risk patients with stage I and II disease may benefit from adjuvant therapy and/or enhanced imaging protocols to allow for early detection of metastasis.” In another clinical validity study, the authors concluded that the test’s “role in consideration of patients for adjuvant therapy should be examined prospectively.” This use of the test would be as a replacement for SLNB since SLNB is currently used to identify patients clinically diagnosed as stage I and II who have node involvement.

However, the use of the test reviewed for the Medicare population (see Supplemental Information) is to select patients at low-risk of being lymph node-positive who can avoid an SLNB (ie, a triage test for SLNB).

The question addressed in this section of the evidence review is: Does GEP improve the net health outcome in individuals with AJCC stage I or II melanoma?

The following PICOTS were used to select literature to inform this review.
**Patients**  
To select patients for adjuvant therapy and/or enhanced surveillance, the relevant population of interest are patients with AJCC stage I/II cutaneous melanoma.

To select patients who can avoid SLNB, the relevant population of interest are patients with AJCC stage I or II cutaneous melanoma who are being considered for SLNB.

**Interventions**  
The test being considered is the Castle Biosciences DecisionDx-Melanoma test. The DecisionDx test measures expression of 31 genes using quantitative reverse-transcription polymerase chain reaction. The test includes 28 prognostic gene targets and 3 endogenous control genes. The test is performed on standard tissue sections from an existing formalin-fixed, paraffin-embedded biopsy or wide local excision specimen.

Development of the test was described in Gerami et al (2015).42 To develop the DecisionDx-Melanoma gene panel, Gerami et al (2015) conducted a meta-analysis of published studies that identified differential gene expression in metastatic vs nonmetastatic primary cutaneous melanoma. Of 54 identified genes, investigators selected 20 for further polymerase chain reaction analysis based on chromosomal location. Five genes from Castle Biosciences’ DecisionDx-UM gene panel were added based on analysis of metastatic and nonmetastatic primary cutaneous melanoma, and two probes of the BRCA1-associated protein 1 gene, BAP1, which has been associated with the metastatic potential of uveal melanoma, also were added. Finally, four genes with minimal variation in expression level between metastatic and nonmetastatic primary cutaneous melanoma were added as controls. Patients had a minimum follow-up of five years unless there was a well-documented metastatic event, including positive SLNB. Information about treatments received was not provided.

The DecisionDx test report provides two results: a class and a probability score. The class stratifies tumors as low-risk (class 1) or high-risk (class 2), with subclassifications within each class (A or B) based on how close the probability score is to the threshold between class 1 and class 2. The probability score ranges from zero to one and appears to be the risk of recurrence within five years.

**Comparators**  
Treatment and surveillance recommendations are based on AJCC staging. SLNB may be used to get more definitive information about the status of the regional nodes compared with a physical examination. The American Society of Clinical Oncology and National Comprehensive Cancer Network have similar but not identical recommendations regarding which patients should undergo SLNB based on thickness and other high-risk features.

SLNB has a low rate of complications; in the Sunbelt Melanoma Trial, a prospective multi-institutional study of SLNB for melanoma reported by Wrightson et al (2003), less than 5% of the 2120 patients developed major or minor complications associated with SLNB.44

Online tools are available to predict prognosis based on the AJCC guidelines. The original AJCC tool was developed by Soong et al (n.d.).45 Callender et al (2012) incorporated SLNB results into a revised tool (http://www.melanomacalculator.com/).46
Outcomes
Regarding selecting patients for adjuvant therapy and/or enhanced surveillance:
A negative DecisionDx (class 1) test result would not change outcomes. Per guidelines, the patients would not receive adjuvant therapy or enhanced surveillance, just as without the DecisionDx test. A positive DecisionDx (class 2) test result would indicate that a patient might benefit from adjuvant therapy or enhanced surveillance. Therefore, the potential beneficial outcomes of a true positive result are additional treatment and surveillance and potentially prolonged survival. The potential harmful outcomes of a false-positive result are unnecessary adverse effects and burdens of adjuvant therapy and enhanced surveillance.

Regarding selecting patients who can avoid SLNB:
For patients meeting guideline-recommended criteria for SLNB, a positive DecisionDx (class 2) test result would not change outcomes. The patients would proceed to SLNB, as they would have without the DecisionDx test, and treatment and imaging decisions would depend on SLNB results. A negative DecisionDx (class 1) test result would indicate that a patient could avoid an SLNB. Therefore, the potential beneficial outcomes of a true-negative result are avoidance of an SLNB and related adverse effects and burdens. The potential harmful outcomes of a false-negative result are reduced time to recurrence due to not identifying node-positive patients that would be eligible for beneficial adjuvant treatment and potentially reduced survival.

The risk of recurrence decreases over time but does not reach zero. In a study of 1568 patients with stage I melanoma, Dicker et al (1999) found that 80% of the recurrences occurred within the first 3 years. A prospective study by Garbe et al (2003) reported that, for stage I and II patients, the risk of recurrence was low after 4.4 years. Among 4731 patients treated for more than 10 years at 1 institution, Faries et al (2013) found the majority of recurrences occurred in the first 5 years. However, 7% of patients experienced recurrence after 10 years (median, 16 years). Even among stage I/II patients, recurrence after 10 years occurred in 2% of patients. Five-year recurrence-free survival (RFS) is the outcome and time-point of interest.

Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid
Study Selection Criteria
For the evaluation of clinical validity of the DecisionDx test, studies that meet the following eligibility criteria were considered:
- Reported on a validation cohort that was independent of the development cohort;
- Reported on the accuracy of the marketed version of the technology;
- Included a suitable reference standard (five-year RFS or five-year metastasis-free survival [MFS]);
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described.
Several papers were excluded from the evaluation of clinical validity. Hsueh et al (2017) and Podlipnik et al (2019) were excluded from the evaluation of the clinical validity of the DecisionDx test because they did not report 5-year outcomes (median follow-up, 1.5 years and 2 years respectively).\textsuperscript{50,51} Samples used in Gerami et al (2015)\textsuperscript{48} and Ferris et al (2017)\textsuperscript{24} appear to overlap with the samples from Gerami et al (2015)\textsuperscript{42}, and each other and will not be considered independent validation studies for inclusion in the tables. They are described briefly following the clinical validity tables. Samples used in both papers by Gastman et al (2018) are stated to overlap previous validation studies.\textsuperscript{52,53} Vetto et al (2019) included a retrospective cohort that was used to develop the model and is thus not eligible for inclusion, as well a prospective cohort with some overlapping samples and without report of 5-year outcomes.\textsuperscript{54}

Three independent clinical validity studies meeting eligibility criteria have been conducted. Characteristics and results are summarized in Tables 9 and 10 and briefly in the paragraphs that follow.

Table 9. Clinical Validity Study Characteristics of the DecisionDx Test for Diagnosing Melanoma

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Population</th>
<th>Design</th>
<th>Reference Standard / Outcome Measure</th>
<th>Threshold Score for Positive DecisionDx Test</th>
<th>Timing of Reference and DecisionDx Tests</th>
<th>Blinding of Assessors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerami et al (2015)\textsuperscript{42}, Validation subset</td>
<td>Adults</td>
<td>Retrospective</td>
<td>5-y RFS</td>
<td>Class 2 is high-risk</td>
<td>Patient diagnosed between 1998 and 2009</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not consecutive or randomly selected</td>
<td>Risk threshold not provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stage I-IV cutaneous melanoma (87% stage I/II)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>At least 5 y of FU (median, 7.0 y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zager et al (2018)\textsuperscript{43}.</td>
<td>Stage I-III cutaneous melanoma (68% stage I/II)</td>
<td>Retrospective</td>
<td>5-y RFS</td>
<td>Class 2 is high-risk</td>
<td>Patients diagnosed between 2000 and 2014</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not consecutive or randomly selected</td>
<td>Class 1: probability score 0 to 0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Class 2: probability score 0.5 to 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>At least 5 y of FU (median, 7.5 y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhaw et al (2018)\textsuperscript{55}.</td>
<td>Patients who were treated for primary invasive CM of any Breslow depth within the last 5 years and had had GEP testing (86% stage I, 14% stage II) Mean follow-up of 23 months; only 20 patients had 5-year follow-up</td>
<td>Retrospective Consecutive</td>
<td>5-y MFS</td>
<td></td>
<td>Institution offered DecisionDx testing to newly diagnosed and those treated within the previous five years</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Commercial test cutoffs used</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| FU: follow-up; RFS: recurrence-free survival; MFS: metastasis-free survival; GEP: gene expression profiling; CM: cutaneous melanoma.
Table 10. Clinical Validity Study Results of the DecisionDx Test for Diagnosing Melanoma

<table>
<thead>
<tr>
<th>Study</th>
<th>Initial / Final N</th>
<th>Excluded Samples</th>
<th>Events and Kaplan-Meier 5-Year RFS or MFS&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sensitivity&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Specificity&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PPV&lt;sup&gt;b&lt;/sup&gt;</th>
<th>NPV&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerami et al (2015); Validation subset</td>
<td>Unclear/104</td>
<td>Samples excluded if melanoma dx not confirmed, dissectible area not acceptable</td>
<td>· 4 events RFS=79 (NR) · 31 events RFS=97 (NR) p&lt;0.01 vs class 1</td>
<td>89 (73 to 97)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>83 (72 to 91)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>72 (56 to 85)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>93 (84 to 98)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Overall</td>
<td>Unclear/78</td>
<td></td>
<td>· 3 events RFS=98 (NR) · 18 events RFS=37 (NR) p&lt;0.01 vs class 1</td>
<td>86 (64 to 97)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>84 (72 to 93)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>67 (46 to 83)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>94 (84 to 99)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Zager et al (2018)</td>
<td>Did not meet analytic quality control thresholds</td>
<td></td>
<td>· 42 events RFS=88 (85 to 92) · 100 events RFS=52 (46 to 60)</td>
<td>70 (62 to 78)</td>
<td>71 (67 to 76)</td>
<td>48 (41 to 55)</td>
<td>87 (82 to 90)</td>
</tr>
<tr>
<td>Overall</td>
<td>601 / 523</td>
<td></td>
<td>· 11 events RFS=96 (94 to 99) · 30 events RFS=85 (74 to 97)</td>
<td>35 (14 to 62)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>87 (82 to 91)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>15 (6 to 31)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>95 (91 to 98)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>AJCC stage I</td>
<td>Unclear/264</td>
<td></td>
<td>· 9 events RFS=74 (60 to 91) · 30 events RFS=55 (44 to 69)</td>
<td>77 (61 to 89)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>43 (29 to 57)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>49 (36 to 62)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>72 (53 to 86)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>AJCC stage II</td>
<td>Unclear/93</td>
<td></td>
<td>· 3 events MFS=93 (82 to 100) · 8 events MFS=69 (52 to 90)</td>
<td>77 (46 to 94)</td>
<td>87 (82 to 91)</td>
<td>24 (13 to 40)</td>
<td>99 (96 to 100)</td>
</tr>
<tr>
<td>Greenhaw et al (2018)</td>
<td>256 / 256</td>
<td>None excluded but only 20 had 5-year follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AJCC: American Joint Committee on Cancer; Dx: diagnosis; NPV: negative predictive value; NR: not reported; PPV: positive predictive value; RFS: recurrence-free survival; MFS: metastasis-free survival

<sup>a</sup> Values are percentages with 95% confidence interval.

<sup>b</sup> Confidence intervals not provided in the report; calculated from data provided.

The validation cohort in Gerami et al (2015) included patients with stage 0, I, II, III, or IV disease from 6 U.S. centers (n=104).42 A complete disposition of samples received from the institutions and those included in the analysis was not provided. For 78 patients in the validation cohort with AJCC stage I or II cutaneous melanoma who had either a metastatic event or had more than 5 years of follow-up without metastasis, 5-year disease-free survival was 98% (CIs not reported) for DecisionDx class I patients and 37% for DecisionDx class II patients. The PPV and NPV were 67% and 94%, respectively. CIs for performance characteristics were calculated in Table 10 based on data provided. Reclassification of patients in AJCC stages to DecisionDx classes is shown in Table 11.
Table 11. Reclassification of Patients Based on AJCC Stages to DecisionDx Classes in the Gerami Validation Cohort

<table>
<thead>
<tr>
<th>AJCC Stage</th>
<th>DecisionDx Class</th>
<th>Class 1 (Low-Risk), N (row %)</th>
<th>Class 2 (High-Risk), N (row %)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total stage I</td>
<td></td>
<td>50 (89%)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6 (11%)</td>
<td>56</td>
</tr>
<tr>
<td>IA</td>
<td></td>
<td>37</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td></td>
<td>10</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total stage II</td>
<td></td>
<td>10 (29%)</td>
<td>24 (71%)</td>
<td>34</td>
</tr>
<tr>
<td>IIA</td>
<td></td>
<td>5</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>IIB</td>
<td></td>
<td>5</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>IIC</td>
<td></td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total stage III</td>
<td></td>
<td>1 (8%)</td>
<td>11 (92%)</td>
<td>12</td>
</tr>
<tr>
<td>Total stage IV</td>
<td></td>
<td>0 (0%)</td>
<td>2 (100%)</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>61</td>
<td>43</td>
<td>104</td>
</tr>
</tbody>
</table>

Adapted from Gerami et al (2015).<sup>42</sup>
AJCC: American Joint Committee on Cancer.
<sup>a</sup> The subclass for n=3 class 1 samples are not reported.

Zager et al (2018) reported results of a second clinical validity study including AJCC stage I, II, or III primary melanoma tumors from 16 U.S. sites.<sup>43</sup> The samples were independent of the other validation studies. Of the 601 cases submitted from the institutions, 523 were included in the analysis (357 stage I/II). The excluded samples did not meet pre- and post-analytic quality control thresholds. SLNB status was untested in 36% of the patients, negative in 34%, and positive in 30%. The report did not describe any adjuvant therapy that the patients received. Overall, 42 (13%) recurrence events occurred in DecisionDx class 1 patients and 100 (48%) recurrence events occurred in DecisionDx class 2 patients. The 5-year RFS estimated by Kaplan-Meier was 88% (95% CI, 85% to 92%) in class 1 and 52% (95% CI, 46% to 60%) in class 2. The reported sensitivity and specificity were 70% (95% CI, 62% to 78%) and 71% (95% CI, 67% to 76%), respectively, with a PPV of 48% (95% CI, 41% to 55%) and a NPV of 87% (95% CI, 82% to 90%). For comparison, the performance characteristics for 5-year RFS for sentinel lymph node status among those with SLNB were: sensitivity, 66% (95% CI, 57% to 74%); specificity, 65% (95% CI, 58% to 71%); PPV, 52% (95% CI, 44% to 60%); and NPV, 76% (95% CI, 69% to 82%). Estimates stratified by AJCC stage I or II are shown in Table 10. The reclassification of patients based on SLNB status using DecisionDx classes is shown in Table 12. If DecisionDx were used as a triage test such that only class 2 received SLNB, then 159 class 1 patients would not have undergone SLNB. Of the 159 patients in class 1, 56 were SLNB-positive and were therefore eligible for adjuvant therapy. It is not clear if the SLNB-positive patients in this study received adjuvant therapy. Of the 56 patients who were DecisionDx class 1 and SLNB-positive, 22 recurrence events occurred by 5 years.

Relevance, design, and conduct gaps are summarized in Tables 13 and 14.

Table 12. Reclassification of Patients Based on SLNB Status to DecisionDx Classes

<table>
<thead>
<tr>
<th>SLNB</th>
<th>DecisionDx Class 1 (Low-Risk)</th>
<th>DecisionDx Class 1 (High-Risk)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>n (%)</td>
<td>Events 5-Year RFS (95% CI), %</td>
<td>n (%) Events 5-Year RFS (95% CI), %</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>103 (65) 15 87 (81 to 94)</td>
<td>77 (43) 28 67 (57 to 79)</td>
<td>180</td>
</tr>
<tr>
<td>Positive</td>
<td>56 (35) 22 61 (49 to 76)</td>
<td>101 (57) 60 37 (28 to 49)</td>
<td>157</td>
</tr>
<tr>
<td>Total</td>
<td>159</td>
<td>178</td>
<td>337&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Adapted from Zager et al (2017).<sup>41</sup>
CI: confidence interval; RFS: recurrence-free survival; SLNB: sentinel lymph node biopsy.
<sup>a</sup> 337 patients had DecisionDx results and SLNB results.
Greenhaw et al (2018) reported results of an independent study of the DecisionDx test using their institution’s melanoma registry and including patients who had been treated for cutaneous melanoma within the last 5 years and undergone DecisionDx testing. Study characteristics and results were reported in the preceeding Tables 9 and 10. Two-hundred fifty-six patients were tested; 84% were categorized as DecisionDx class 1 (low-risk) and 16% were DecisionDx class 2 (high-risk). 219 (86%) of tumors were AJCC stage I and 37 (14%) were AJCC stage II. None of the 18 stage I/class 2 tumors metastasized but 1 (0.5%) of 201 stage I/class 1 tumors metastasized. Ten (42%) of the stage II/class 2 tumors metastasized and 2 (15%) of the 13 stage II/class 1 tumors metastasized.

### Table 13. Clinical Validity Study Relevance Gaps of the DecisionDx Test

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Outcomes</th>
<th>Duration of Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerami et al (2015); Validation subset</td>
<td>4. Study population includes AJCC stage III/IV lesions (13%), although analysis for only stage I/II was provided</td>
<td>1. Risk threshold for classification into class 1 or 2 not provided.</td>
<td>3. Not compared to other prediction tools</td>
<td>2. Evidence-based treatment or surveillance pathway using the test is not described</td>
<td></td>
</tr>
<tr>
<td>Zager et al (2018)</td>
<td>4. Study population includes AJCC stage III lesions (32%), although analysis for only stage I/II was provided</td>
<td>3. Not compared to other prediction tools</td>
<td></td>
<td>2. Evidence-based treatment or surveillance pathway using the test is not described</td>
<td></td>
</tr>
<tr>
<td>Greenhaw et al (2018)</td>
<td>3. Not compared to other prediction tools</td>
<td></td>
<td></td>
<td>2. Evidence-based treatment or surveillance pathway using the test is not described</td>
<td>1. Only 20 patients had 5-year follow-up</td>
</tr>
</tbody>
</table>

The evidence gaps stated in this table are those notable in the current review; this is not a comprehensive gaps assessment. AJCC: American Joint Committee on Cancer.

- **Population** key: 1. Intended use population unclear; 2. Clinical context is unclear; 3. Study population is unclear; 4. Study population not representative of intended use.
- **Intervention** key: 1. Classification thresholds not defined; 2. Version used unclear; 3. Not intervention of interest.
- **Comparator** key: 1. Classification thresholds not defined; 2. Not compared to credible reference standard; 3. Not compared to other tests in use for same purpose.
- **Outcomes** key: 1. Study does not directly assess a key health outcome; 2. Evidence chain or decision model not explicated; 3. Key clinical validity outcomes not reported (sensitivity, specificity and predictive values); 4. Reclassification of diagnostic or risk categories not reported; 5. Adverse events of the test not described (excluding minor discomforts and inconvenience of venipuncture or noninvasive tests).
- **Follow-Up** key: 1. Follow-up duration not sufficient with respect to natural history of disease (true-positives, true-negatives, false-positives, false-negatives cannot be determined).

### Table 14. Clinical Validity Study Design and Conduct Gaps of the DecisionDx Test

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection</th>
<th>Blinding</th>
<th>Delivery of Test</th>
<th>Selective Reporting</th>
<th>Completeness of Follow-Up</th>
<th>Statistical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerami et al (2015); Validation subset</td>
<td>2. Not consecutive or random</td>
<td>1. Time between collection of biopsy and DecisionDx not described</td>
<td>1. No registration reported</td>
<td>1. No description of number of samples (if any) that failed to produce results or were indeterminate</td>
<td>1. CIs not reported but were calculated based on data provided</td>
<td></td>
</tr>
</tbody>
</table>
In a subsequent analysis of patients with melanoma who had undergone SLNB, Gerami et al (2015) compared prognostic classification by DecisionDx-Melanoma with biopsy results. A total of 217 patients comprised a convenience sample from a database of 406 patients previously tested with DecisionDx-Melanoma. Patients who had undergone SLNB appear to overlap with patients in Gerami et al (2015) discussed previously. Most (73%) patients had a negative SLNB, and 27% had a positive SLNB. DecisionDx-Melanoma classified 76 (35%) tumors as low-risk (class I) and 141 (65%) tumors as high-risk (class II). Within the group of SLNB-negative patients, the 5-year OS rate was 91% in class I patients and 55% in class II patients. Within the group of SLNB-positive patients, the 5-year OS rate was 77% in class I patients and 57% in class II patients.

Ferris et al (2017) compared the accuracy of DecisionDx-Melanoma with the web-based AJCC Individualized Melanoma Patient Outcome Prediction Tool. The study included 205 patients who appear to overlap with the patients in the second Gerami et al (2015) study described above. AJCC-predicted 5-year survival for each patient was categorized into low and high-risk based on both a 68% predicted 5-year survival and a 79% predicted 5-year survival. The 68% and 79% cutpoints were reported to correspond to 5-year survival in patients with stage IIA and IIB, respectively, although it is unclear whether those cutpoints were prespecified, whether they were based on internal or external estimates of risk, or whether they are commonly used in practice. The prognostic sensitivity and specificity for death (median follow-up, 7 years) of the Decision-Dx Melanoma were 78% and 69%, respectively (CIs not reported). The sensitivity and specificity for the AJCC calculator with the 79% cutpoint were 60% and 74%, respectively. The combination of the DecisionDx-Melanoma and AJCC tools had a sensitivity of 82% and specificity of 62%. The cross-classification for the DecisionDx-Melanoma and AJCC tools for 5-year OS is shown in Table 15.
### Table 15. Cross-Classification for the DecisionDx-Melanoma and AJCC Tool (79% Cutpoint) for 5-Year Overall Survival

<table>
<thead>
<tr>
<th>Risk Classification (DecisionDx-Melanoma vs AJCC)</th>
<th>N</th>
<th>No. of Events</th>
<th>5-Year Overall Survival, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low/low</td>
<td>105</td>
<td>9</td>
<td>96</td>
</tr>
<tr>
<td>Low/high</td>
<td>13</td>
<td>2</td>
<td>83</td>
</tr>
<tr>
<td>High/low</td>
<td>30</td>
<td>11</td>
<td>71</td>
</tr>
<tr>
<td>High/high</td>
<td>57</td>
<td>28</td>
<td>44</td>
</tr>
</tbody>
</table>

Adapted from Ferris et al (2017).24
AJCC: American Joint Committee on Cancer.

**Section Summary: Clinically Valid**

To use prognostic information for decision-making, performance characteristics should be consistent and precise. Two independent studies, using archived tumor specimens, have reported 5-year RFS in AJCC stage I or II patients.

If the test is to be used to select stage I and II patients for adjuvant therapy or enhanced surveillance then it should identify a group with high-risk of recurrence. Gerami et al (2015) reported RFS rates of 37% for DecisionDx class 2 (high-risk) in patients in AJCC stage I and II patients. However, Zager et al (2018) reported RFS rates of 85% (95% CI, 74% to 97%) for DecisionDx class 2 patients in AJCC stage I and 55% (95% CI, 44% to 69%) for DecisionDx class 2 in AJCC stage II disease. In addition, to ‘rule-in’ patients for additional treatment or surveillance, the test should have specificity and PPV. Zager et al (2018) and Greenhaw et al (2018) the specificities were 71% and 87% respectively while the PPV were only 48% and 24%, respectively. The low PPV suggests that the majority of patients identified as high-risk by the the DecisionDx test would not develop metastasis and would be unnecessarily subjected to additional treatment or surveillance.

If the test is to be used to select stage I and II patients who can avoid SLNB, then it should identify a group who are eligible for SLNB but have low-risk of recurrence. Gerami et al (2015) reported RFS rates of 98% in DecisionDx class 1 (low-risk) without CIs in AJCC stage I or II patients. Zager et al (2018) reported RFS rates of 96% (95% CI, 94% to 99%) for DecisionDx class 1 in patients with AJCC stage I disease and RFS rates of 74% (95% CI, 60% to 91%) for DecisionDx class 1 in patients with AJCC stage II disease.

Although CIs were not available for the first study, RFS does not appear to be well-characterized in either DecisionDx risk group as evidenced by the variation in estimates across studies.

Zager et al (2017) also reported that 56 of 159 (35%) patients who were DecisionDx class 1 (low-risk) were SLNB-positive and in those patients 22 recurrences (39%) occurred over 5 years.43 If the DecisionDx test were used as a triage for SLNB, these patients would not undergo SLNB and would likely not receive adjuvant therapy, which has shown to be effective at prolonging time to recurrence in node-positive patients.

Greenhaw et al (2018) also reported that in 219 AJCC stage I patients, 201 had DecisionDx class 1 (low-risk) scores and 18 had DecisionDx class 2 (high-risk) scores. The only metastasis in stage I patients occurred in a patient with a DecisionDx class 1 score. Therefore, with respect to the proposed uses of identifying higher risk patients that should receive adjuvant therapy or enhanced surveillance, none of their stage 1 patients benefited from DecisionDx testing but 18
(8%) were incorrectly identified as high-risk for metastasis and could have received unnecessary treatment or surveillance.

**Clinical Useful**
A test is clinically useful if the results inform management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

**Direct Evidence**
Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

No direct evidence of clinical utility was identified.

**Chain of Evidence**
Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

Four decision-impact studies have been published reporting on the impact of DecisionDx on physicians’ management decisions. Given the lack of established clinical validity and no reported long-term outcomes of the test used to select patients for active surveillance, it is not known whether any management changes were clinically appropriate.

For the proposed use of the test as a triage for SLNB (identify patients who can avoid SLNB), performance characteristics are not well-characterized.

For the proposed use of the test as a replacement for SLNB (identify patients who are AJCC stage I/II who should receive adjuvant therapy), performance characteristics are also not well-characterized. In addition, an evidence-based management pathway would be needed to support the chain of evidence. The existing RCTs demonstrating that adjuvant therapy reduces recurrence included node positive patients.

For the proposed use of the test to identify patients who are AJCC stage I/II who should receive enhanced surveillance, there is also a lack of evidence that imaging surveillance or increased frequency of surveillance improves outcomes in stage I/I patients. The National Comprehensive Cancer Network guidelines state that imaging surveillance is not recommended for stage I-IIA and can be 'considered' for IIB-IV but that there is an absence of meaningful data on the association of rigorous routine surveillance imaging with improved long-term outcome for stage IIB-IIC and the recommendations regarding consideration of imaging surveillance remain controversial. While earlier detection of recurrence is thought to be beneficial because lower tumor burden and younger age are associated with improved treatment response and survival, this has not been proven and RCTs are needed to assess whether enhanced surveillance improves survival. The optimal frequency and duration of follow-up surveillance are not standardized and how the surveillance would be altered for DecisionDx class 2 patients has not be defined.
No evidence was identified that demonstrated that adjuvant therapy or increased surveillance improves net health outcomes in AJCC stage I or II patients who are DecisionDx class 2.

Section Summary: Clinically Useful
There is no direct evidence of clinical utility. A chain of evidence for clinical utility cannot be created due to lack of robust evidence of clinical validity and lack of evidence-based management pathway.

SUMMARY OF EVIDENCE
For individuals with suspicious pigmented lesions (based on ABCDE and/or ugly duckling criteria) being considered for biopsy who receive GEP with the DermTech PLA to determine which lesions should proceed to biopsy, the evidence includes observational studies. The relevant outcomes are OS, disease-specific survival, validity, and resource utilization. The PLA has one clinical validity study with many methodologic and reporting limitations. Therefore, performance characteristics are not well-characterized. Also, the test has not been compared with dermoscopy, another tool frequently used to make biopsy decisions. No direct evidence of clinical utility was identified. Given that the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility through a chain of evidence. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have melanocytic lesions with indeterminate histopathologic features who receive GEP with the myPath Melanoma test added to histopathology to aid in the diagnosis of melanoma, the evidence includes observational studies. The relevant outcomes are OS, disease-specific survival, test validity, change in disease status, treatment-related morbidity. The myPath test has one clinical validity study, which includes long-term follow-up for metastasis as the reference standard. However, it is not clear if the study population included lesions that were indeterminate following histopathology and the study had other methodologic and reporting limitations. Therefore, performance characteristics are not well-characterized. No direct evidence of clinical utility was identified. Given that the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility through a chain of evidence. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals with AJCC stage I or II cutaneous melanoma who receive GEP with the DecisionDx-Melanoma test to inform management decisions regarding enhanced surveillance, the evidence includes retrospective observational studies. The relevant outcomes are OS, disease-specific survival, test validity, change in disease status, resource utilization and treatment-related morbidity. The DecisionDx-Melanoma test has three independent clinical validity studies that have reported five-year RFS in AJCC stage I or II patients. Gerami et al (2015) reported RFS rates of 37% for DecisionDx class 2 (high-risk) in patients in AJCC stage I and II patients combined. Zager et al (2018) reported RFS rates of 85% (95% CI, 74% to 97%) for DecisionDx class 2 patients in AJCC stage 1 and 55% (95% CI, 44% to 69%) for DecisionDx class 2 in AJCC stage II disease. RFS does not appear to be well-characterized as evidenced by the variation in estimates across studies. This indication is to 'rule-in' patients for enhanced surveillance; therefore, specificity and PPV are key performance characteristics. Zager et al (2018) and Greenhaw et al (2018) the specificities were 71% and 87% respectively while the PPV were 48% and 24%, respectively. The PPV suggests that the majority of patients identified as high-risk by the the DecisionDx test would not develop metastasis and would be unnecessarily subjected to additional surveillance. Greenhaw et al (2018) also reported that in 219 AJCC stage I patients, 201 had...
DecisionDx class 1 (low-risk) scores and 18 had DecisionDx class 2 (high-risk) scores. The only metastasis in stage I patients occurred in a patient with a DecisionDx class 1 score. Therefore, none of their stage 1 patients benefited from DecisionDx testing but 18 (8%) were incorrectly identified as high-risk for metastasis, and could have received unnecessary surveillance. There is no evidence that changes to the frequency and methods for surveillance improve outcomes. Given that the evidence is insufficient to demonstrate test performance and there is no evidence that changes in surveillance improve outcomes, no inferences can be made about clinical utility through a chain of evidence. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals with AJCC stage I or II cutaneous melanoma who receive GEP with the DecisionDx-Melanoma test to inform management decisions regarding adjuvant therapy, the evidence includes retrospective observational studies. The relevant outcomes are OS, disease-specific survival, test validity, change in disease status, resource utilization and treatment-related morbidity. The DecisionDx-Melanoma test has three independent clinical validity studies that have reported five-year RFS in AJCC stage I or II patients. Gerami et al (2015) reported RFS rates of 37% for DecisionDx class 2 (high-risk) in patients in AJCC stage I and II patients combined. Zager et al (2018) reported RFS rates of 85% (95% CI, 74% to 97%) for DecisionDx class 2 patients in AJCC stage 1 and 55% (95% CI, 44% to 69%) for DecisionDx class 2 in AJCC stage II disease. RFS does not appear to be well-characterized as evidenced by the variation in estimates across studies. This indication is to 'rule-in' patients for adjuvant therapy; therefore, specificity and PPV are key performance characteristics. Zager et al (2018) and Greenhaw et al (2018) the specificities were 71% and 87% respectively while the PPV were 48% and 24%, respectively. The PPV suggests that the majority of patients identified as high-risk by the DecisionDx test would not develop metastasis and would be unnecessarily subjected to additional treatment. Greenhaw et al (2018) also reported that in 219 AJCC stage I patients, 201 had DecisionDx class 1 (low-risk) scores and 18 had DecisionDx class 2 (high-risk) scores. The only metastasis in stage I patients occurred in a patient with a DecisionDx class 1 score. Therefore, none of their stage 1 patients benefited from DecisionDx testing but 18 (8%) were incorrectly identified as high-risk for metastasis and could have received unnecessary treatment. There is no evidence that adjuvant therapy improves outcomes in these patients. Given that the evidence is insufficient to demonstrate test performance and there is no evidence that adjuvant therapy improves outcomes, no inferences can be made about clinical utility through a chain of evidence. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals with cutaneous melanoma with clinically negative sentinel node basins who are being considered for SLNB who receive GEP with the DecisionDx-Melanoma test to determine whether to perform SLNB, the evidence includes retrospective observational studies. The relevant outcomes are OS, disease-specific survival, test validity, change in disease status, resource utilization and treatment-related morbidity. The DecisionDx-Melanoma test has three independent clinical validity studies that have reported five-year RFS in AJCC stage I or II patients. Gerami et al (2015) reported RFS rates of 98% in DecisionDx class 1 (low-risk) without CIs, in AJCC stage I or II patients. Zager et al (2017) reported RFS rates of 96% (95% CI, 94% to 99%) for DecisionDx class 1 in patients with AJCC stage I disease; they also reported RFS rates of 74% (95% CI, 60% to 91%) for DecisionDx class 1 in patients with AJCC stage II disease. Although CIs were not available for the first study, RFS does not appear to be well-characterized as evidenced by the variation in estimates across studies. Zager et al (2017) also reported that in 56 patients who were DecisionDx class 1 (low-risk) but SLNB-positive, 22 recurrences (39%) occurred over 5 years. If the DecisionDx test were used as a triage for SLNB,
these patients would not undergo SLNB and would likely not receive adjuvant therapy, which has shown to be effective at prolonging time to recurrence in node-positive patients. No direct evidence of clinical utility was identified. Given that the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility through a chain of evidence. The evidence is insufficient to determine the effects of the technology on health outcomes.

**PRACTICE GUIDELINES AND POSITION STATEMENTS**

**National Comprehensive Cancer Network**

The National Comprehensive Cancer Network guidelines (v.2.2019) for melanoma made the following statements on use of gene expression profiling. "While there is interest in newer prognostic molecular techniques such as gene expression profiling to differentiate melanomas at low versus high risk for metastasis, routine (baseline) prognostic genetic testing of primary cutaneous melanoma (before or following sentinel lymph node biopsy) is not recommended outside of a clinical study (trial)."

The guidelines state the following regarding diagnostic testing for indeterminate melanocytic neoplasms following histopathology: "They may be used on a case-by-case basis in ambiguous melanocytic tumors; however, their utility is still under evaluation, and more data are needed before they can be routinely recommended." Specifically regarding the GEP test, the guidelines state that "...long-term follow-up is required to validate the prognostic significance of this test." The guidelines state the following regarding prognostic testing: "Commercially available GEP tests are marketed as being able to classify cutaneous melanoma into separate categories based on metastasis. However, it remains unclear whether these tests provide clinically actionable prognostic information when used in addition to or in comparison with known clinicopathologic factors or multivariable nomograms that incorporate patient sex, age, tumor location and thickness, ulceration, mitotic rate, lymphovascular invasion, microsatellites, and SLNB status. Furthermore, the impact of these tests on treatment outcomes or follow-up schedules has not been established."

**American Academy of Dermatology**

The American Academy of Dermatology (2019) published guidelines of care for the management of primary cutaneous melanoma. The guidelines state the following regarding GEP tests:

- Regarding diagnostic GEP tests:
  - "Diagnostic molecular techniques are still largely investigative and may be appropriate as ancillary tests in equivocal melanocytic neoplasms, but they are not recommended for routine diagnostic use in CM. These include comparative genomic hybridization, fluorescence in situ hybridization, gene expression profiling (GEP), and (potentially) next-generation sequencing."
  - "Ancillary diagnostic molecular techniques (eg, CGH, FISH, GEP) may be used for equivocal melanocytic neoplasms."

- Regarding prognostic GEP tests:
  - "...there is also insufficient evidence of benefit to recommend routine use of currently available prognostic molecular tests, including GEP, to provide more accurate prognosis beyond currently known clinicopathologic factors" (Strength of evidence: C, Level of evidence II/III)
  - "Going forward, GEP assays should be tested against all known histopathologic prognostic factors and contemporary eighth edition of AJCC CM staging to assess their additive value in prognostication."
"Routine molecular testing, including GEP, for prognostication is discouraged until better use criteria are defined. The application of molecular information for clinical management (e.g., sentinel lymph node eligibility, follow-up, and/or therapeutic choice) is not recommended outside of a clinical study or trial."

**U.S. PREVENTIVE SERVICES TASK FORCE RECOMMENDATIONS**

Not applicable.

**ONGOING AND UNPUBLISHED CLINICAL TRIALS**

Some currently unpublished trials that might influence this review are listed in Table 16.

**Table 16. Summary of Key Trials**

<table>
<thead>
<tr>
<th>NCT No.</th>
<th>Trial Name</th>
<th>Planned Enrollment</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ongoing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCT02355574a</td>
<td>An Ongoing, 5-year Post Market Study to Track Clinical Application of DecisionDx-Melanoma Gene Expression Profile (GEP) Assay Results and the Impact on Patient Outcomes and Health Economics</td>
<td>1672</td>
<td>Jun 2024</td>
</tr>
<tr>
<td>NCT02355587a</td>
<td>An Open, 5-year Registry Study to Track Clinical Application of DecisionDx-Melanoma Gene Expression Profile Assay Results and Associated Patient Outcomes</td>
<td>5000</td>
<td>Feb 2024</td>
</tr>
</tbody>
</table>

NCT: national clinical trial.
a Denotes industry-sponsored or cosponsored trial.

**CODING**

The following codes for treatment and procedures applicable to this policy are included below for informational purposes. Inclusion or exclusion of a procedure, diagnosis or device code(s) does not constitute or imply member coverage or provider reimbursement. Please refer to the member’s contract benefits in effect at the time of service to determine coverage or non-coverage of these services as it applies to an individual member.

**CPT/HCPCS**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>81479</td>
<td>Unlisted molecular pathology procedure</td>
</tr>
<tr>
<td>81599</td>
<td>Unlisted multianalyte assay with algorithmic analysis</td>
</tr>
<tr>
<td>84999</td>
<td>Unlisted chemistry procedure</td>
</tr>
<tr>
<td>0089U</td>
<td>Oncology (melanoma), gene expression profiling by RTqPCR, PRAME and LINCO0518, superficial collection using adhesive patch(es)</td>
</tr>
<tr>
<td>0090U</td>
<td>Oncology (cutaneous melanoma) mRNA gene expression profiling by RT-PCR of 23 genes (14 content and 9 housekeeping), utilizing formalin-fixed paraffin embedded tissue, algorithm reported as a categorical result (i.e., benign, indeterminate, or malignant)</td>
</tr>
</tbody>
</table>

**ICD-10 Diagnoses**

Experimental / Investigational for all diagnoses related to this medical policy.
REFERENCES


52. Gastman, BB, Gerami, PP, Kurley, SS, Cook, RR, Leachman, SS, Vetto, JJ. Identification of patients at risk of metastasis using a prognostic 31-gene expression profile in subpopulations of melanoma


Other References
1. Blue Cross and Blue Shield of Kansas Oncology Liaison Committee, May 2019.