Policy

Blue Cross and Blue Shield of Kansas City (Blue KC) will not provide coverage for Molecular Testing in the Management of Pulmonary Nodules. This is considered investigational.

When Policy Topic is covered
n/a

When Policy Topic is not covered
Plasma-based proteomic screening, including but not limited to BDX-XL2, in patients with undiagnosed pulmonary nodules detected by computed tomography is considered investigational.

Gene expression profiling on bronchial brushings using Percepta® Bronchial Genomic Classifier or similar assays in patients with indeterminate bronchoscopy results is considered investigational.

The EarlyCDT Lung for the early detection of lung cancer is considered investigational.

Description of Procedure or Service

<table>
<thead>
<tr>
<th>Populations</th>
<th>Interventions</th>
<th>Comparators</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals: With undiagnosed pulmonary nodules detected by computed tomography</td>
<td>Interventions of interest are: Plasma-based proteomic screening</td>
<td>Comparators of interest are: Standard diagnostic workup of pulmonary nodules</td>
<td>Relevant outcomes include: Overall survival, Disease-specific survival, Test accuracy, Test validity, Morbid events, Hospitalizations, Resource utilization</td>
</tr>
<tr>
<td>Individuals: With undiagnosed pulmonary nodules following</td>
<td>Interventions of interest are: Gene expression profiling of</td>
<td>Comparators of interest are: Standard diagnostic workup of pulmonary nodules</td>
<td>Relevant outcomes include: Overall survival, Disease-specific survival</td>
</tr>
</tbody>
</table>
Plasma-based proteomic screening and gene expression profiling of bronchial brushing are molecular tests available in the diagnostic workup of pulmonary nodules. To rule out malignancy, invasive diagnostic procedures such as computed tomography-guided biopsies, bronchoscopies, or video-assisted thoracoscopic are often required, but each carry procedure-related complications ranging from postprocedure pain to pneumothorax. Molecular diagnostic tests have been proposed to aid in risk-stratifying patients to eliminate or necessitate the need for subsequent invasive diagnostic procedures.

For individuals with undiagnosed pulmonary nodules detected by computed tomography who receive plasma-based proteomic screening, the evidence includes prospective cohorts and prospective-retrospective studies. The relevant outcomes are overall survival, disease-specific survival, test accuracy and validity, morbid events, hospitalizations, and resource utilization. Clinical validation studies were identified for two versions of a proteomic classifier. This classifier has undergone substantial evolution, from a 13 protein assay to a 2 protein assay integrated with clinical factors. Because of this evolution, the most relevant studies are with the most recent version two. One validation study on the version two has been identified. The classifier has been designed to have high specificity for malignant pulmonary nodules, and the validation study showed a specificity of 97% for patients with low to moderate pretest probability (< 50%) of a malignant pulmonary nodule. The primary limitation of this study is that a high number of patients were excluded from the study due to incomplete clinical data or because they were subsequently determined to be outside of the intended use population. It is unclear if the intended use population was determined a priori. Validation in an independent sample in the intended use population is needed. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals with undiagnosed pulmonary nodules following indeterminate bronchoscopy results for suspected lung cancer who receive gene expression profiling of bronchial brushings, the evidence includes multicenter prospective studies. The relevant outcomes are overall survival, disease-specific survival, test accuracy and validity, morbid events, hospitalizations, and resource utilization. Reported receiver operating characteristic curve values ranged from 0.74 to 0.81, with a negative predictive value of 91%. Among patients with a low and intermediate pretest probability of cancer with an inconclusive bronchoscopy, 77 (85%) patients underwent invasive diagnostic procedures. However, there was a relatively high number of missed cancers. No validation of the test in other populations was identified. Also, where the test would fall in the clinical pathway (ie, other than indeterminate bronchoscopy) is uncertain. The evidence is insufficient to determine the effects of the technology on health outcomes.
**Background**  
**Pulmonary Nodules**  
Pulmonary nodules are a common clinical problem that may be found incidentally on a chest x-ray or computed tomography (CT) scan or during lung cancer screening studies of smokers. The primary question after the detection of a pulmonary nodule is the probability of malignancy, with subsequent management of the nodule based on various factors such as the radiographic characteristics of the nodules (eg, size, shape, density) and patient factors (eg, age, smoking history, previous cancer history, family history, environmental/occupational exposures). The key challenge in the diagnostic workup for pulmonary nodules is appropriately ruling in patients for invasive diagnostic procedures and ruling out patients who should forgo invasive diagnostic procedures. However, due to the low positive predictive value of pulmonary nodules detected radiographically, many unnecessary invasive diagnostic procedures and/or surgeries are performed to confirm or eliminate the diagnosis of lung cancer.

**Proteomics**  
Proteomics is the study of the structure and function of proteins. The study of the concentration, structure, and other characteristics of proteins in various bodily tissues, fluids, and other materials has been proposed as a method to identify and manage various diseases, including cancer. In proteomics, multiple test methods are used to study proteins. Immunoassays use antibodies to detect the concentration and/or structure of proteins. Mass spectrometry is an analytic technique that ionizes proteins into smaller fragments and determines mass and composition to identify and characterize them.

**Plasma-Based Proteomic Screening for Pulmonary Nodules**  
Plasma-based proteomic screening has been investigated to risk-stratify pulmonary nodules as likely benign to increase the number of patients who undergo serial CT scans of their nodules (active surveillance), instead of invasive procedures such as CT-guided biopsy or surgery. Additionally, proteomic testing may also determine a likely malignancy in clinically low-risk or intermediate-risk pulmonary nodules, thereby permitting earlier detection in a subset of patients.

Xpresys Lung and BDX-XL2 are plasma-based proteomic screening tests that measures the relative abundance of proteins from multiple disease pathways associated with lung cancer using an analytic technique called multiple reaction monitoring mass spectroscopy. The role of the tests is to aid physicians in differentiating likely benign from likely malignant nodules. If the test yields a likely benign result, patients may choose active surveillance via serial CT scans to monitor the pulmonary nodule. If the test yields a likely malignant result, invasive diagnostic procedures would be indicated. The test is therefore only used in the management of pulmonary nodules to rule in or out invasive diagnostic procedures and does not diagnose lung cancer.
Gene expression profiling
GEP is the measurement of the activity of genes within cells. Messenger RNA serves as the bridge between DNA and functional proteins. Multiple molecular techniques such as Northern blots, ribonuclease protection assay, in situ hybridization, spotted complementary DNA arrays, oligonucleotide arrays, reverse transcriptase polymerase chain reaction, and transcriptome sequencing are used in GEP. An important role of GEP in molecular diagnostics is to detect cancer-associated gene expression of clinical samples to assess for the risk for malignancy.

Gene Expression Profiling for an Indeterminate Bronchoscopy Result
The Percepta Bronchial Genomic Classifier is a 23-gene, GEP test that analyzes genomic changes in the airways of current or former smokers to assess a patient's risk of having lung cancer, without the direct testing of a pulmonary nodule. The test is indicated for current and former smokers following an indeterminate bronchoscopy result to determine the subsequent management of pulmonary nodules (eg, active surveillance or invasive diagnostic procedures), and does not diagnose lung cancer.

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. Xpresys® Lung 2 (BDX-XL2 [Integrated Diagnostics [Indi], purchased by Bidesix) and Percepta® Bronchial Genomic Classifier (Veracyte) 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.

EarlyCDT
EarlyCDT(cancer detection test)–Lung measures a panel of six autoantibodies (p53, NY-ESO-1, CAGE, GBU4-5, Annexin 1, and SOX2) each selected for their involvement in the development of lung cancer. The test detects the body’s immune response in the form of antibodies to antigens, which are produced by solid tumor cancer cells. In healthy individuals, the immune system does not produce antibodies against normal tissue antigens. When cancer develops, cancer cells produce antigens that are abnormal or in some way modified from those found in normal tissue. The immune system detects these antigens and reacts by producing an abundance of autoantibodies that circulate in the bloodstream. As these autoantibodies form and circulate in the body during stages I and II, the earliest stages of lung cancer, their measurement signifies the presence of cancerous cells, indicating that a tumor exists.

Risk assessment and early detection of lung cancer in asymptomatic high-risk patient populations are the primary benefits of EarlyCDT-Lung. It is also used as a follow-up test for high-risk patients who have indeterminate lung nodules identified by CT.
Rationale
This evidence review was created in May 2017 and has been updated regularly with searches of the MEDLINE database. The most recent literature update was performed through March 26, 2019.

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.

Plasma-Based Proteomic Screening of Pulmonary Nodules

Clinical Context and Test Purpose
The purpose of plasma-based proteomic screening in individuals with undiagnosed pulmonary nodule(s) is to stratify clinical risk for malignancy and eliminate or necessitate the need for invasive diagnostic procedures.

The relevant question addressed in this evidence review is: Does plasma-based proteomic screening appropriately eliminate or necessitate the need for invasive diagnostic procedures and lead to improved health outcomes?

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

Patients
The relevant population of interest are individuals with undiagnosed pulmonary nodules. In particular, as outlined in the evidence-based American College of Chest Physicians guidelines (2013) on the diagnosis and management of lung cancer, decision-making about a single indeterminate lung nodule 8 to 30 mm in diameter on a computed tomography (CT) scan is complicated, requiring input about the patient's pretest probability of lung cancer, the characteristics of the lung nodule on CT, and shared decision-making between the patient and physician about follow-up.1 Therefore, additional information in the segment of patients with an indeterminate lung nodule, 8 to 30 mm in diameter would be particularly useful.

Interventions
The test being considered is plasma-based proteomic screening. Of particular focus are the Xpresys Lung test 2 (BDX-XL2). BDX-XL2 measures the abundance of 2 plasma proteins (LG3BP and C163A) and combines the results with 5 clinical risk
factors (age, smoking status, nodule diameter, edge characteristics, and location) to provide a posttest probability of a lung nodule being benign.

**Comparators**
The following practice is currently being used: standard clinical management using clinical and radiographic risk factors.

**Outcomes**
The potential beneficial outcomes of primary interest are avoiding an unneeded invasive biopsy of a nodule that would be negative for lung cancer or initiating a biopsy for a nodule that would otherwise have been followed with serial CTs.

Potential harmful outcomes are those resulting from false-positive or false-negative test results. False-positive test results can lead to unnecessary invasive diagnostic procedures and procedure-related complications. False-negative test results can lead to lack of pulmonary nodule surveillance or lack of appropriate invasive diagnostic procedures to diagnose malignancy.

**Timing**
The time frame for evaluating test performance varies from the initial CT scan to an invasive diagnostic procedure up to two years later, which would be the typical follow-up needed for some lung nodules.

**Setting**
The primary setting would be in outpatient pulmonology or primary care offices.

**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).

Several studies were identified that reported on the development and validation of plasma-based classifier test to predict malignancy (Xpresys Lung, and Xpresys Lung 2 [BDX-XL2]). Initial development was based on a 13-protein plasma classifier as reported by Li et al (2013).\(^2\)

Vachani et al (2015) reported the validation of Xpresys Lung, which was an 11-protein plasma classifier designed to identify likely benign lung nodules (see Tables 1 and 2).\(^3\). This retrospective, blinded analysis evaluated existing samples (n=141) associated with indeterminate pulmonary nodules 8 to 30 mm in diameter. The performance of the classifier in identifying benign nodules was tested at predefined reference values. For example, using a population-based non-
small-cell lung cancer prevalence estimate of 23% for indeterminate pulmonary nodules 8 to 30 mm in diameter, the classifier identified likely benign lung nodules with a 90% negative predictive value (NPV) and a 26% positive predictive value, at 92% sensitivity and 20% specificity, with the lower bound of the classifier's performance at 70% sensitivity and at 48% specificity. Additional sample diagnostic characteristics, selected to keep the study's target NPV of 90%, are shown in Table 2. Classifier scores for the overall cohort were statistically independent of patient age, tobacco use, nodule size, and chronic obstructive pulmonary disease diagnosis. The classifier also demonstrated incremental diagnostic performance in combination with a 4-parameter clinical model.

Vachani et al (2015) reported on a multicenter prospective-retrospective study of patients with indeterminate pulmonary nodules. A plasma protein classifier was used on 475 patients with nodules 8 to 30 mm in diameter who had an invasive procedure to confirm the diagnosis. Using the classifier, 32.0% (95% confidence interval [CI], 19.5% to 46.7%) of surgeries and 31.8% (95% CI, 20.9% to 44.4%) of invasive procedures (biopsy and/or surgery) on benign nodules could have been avoided, while 24.0% (95% CI, 19.2% to 29.4%) of patients with malignancy would have been triaged to CT surveillance. By comparison, 24.5% (95% CI, 16.2% to 34.4%) of patients with malignancy were routed to CT surveillance using clinical parameters alone.

Kearney et al (2017) conducted an exploratory study that combined the 11-protein plasma classifier (Xpresys Lung) with clinical risk factors using 222 samples associated with a lung nodule of 8 to 20 mm in diameter from the reclassification study by Vachani et al (2015) described above. The study determined that the ratio of LG3BP to a normalizer protein C163A was the diagnostic and normalizer protein pair with the highest area under the curve (60%). At sensitivity of 90% and specificity of 33%, the ratio of the proteomic marker was more accurate than clinical risk factors, and the combination of the clinical risk factors with the proteomic markers was more accurate than either alone. This study led to the development of the Xpresys Lung version 2, which includes LG3BP, C163A, and clinical risk factors.

Silvestri et al (2018) reported the validation of the Xpresys Lung version 2 (BDX-XL2) in a prospective multicenter observational study (Pulmonary Nodule Plasma Proteomic Classifier [PANOPTIC]) that enrolled 685 patients with lung nodules of 8 to 30 mm and a low pretest probability of malignancy ≤ 50%. After exclusions for missing clinical data or a pretest probability of > 50%, 178 patients remained in the intended use population. Of these, 66 were classified as likely benign, 65 of which had a benign nodule, while 1 of 29 malignant nodules (3%) was misclassified as likely benign. Of the 149 benign nodules in the study, 44% were correctly classified as likely benign. For the 71 patients who had invasive procedures, 42 had benign nodules. Use of the integrated proteomic classifier would have reduced the number of patients undergoing an invasive procedure to 27, a 36% relative risk reduction, with 1 malignant nodule misclassified as benign.
Table 1. Study Characteristics of Clinical Validity

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Population</th>
<th>Design</th>
<th>Reference Standard</th>
<th>Threshold for Positive Index Test</th>
<th>Timing of Reference and Index Tests</th>
<th>Blinding of Assessors</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vachani et al (2015)</td>
<td>141 samples associated with indeterminate pulmonary nodules</td>
<td>Retrospective analysis with existing samples</td>
<td>Selected to keep NPV of 90%</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Xpresys Lung</td>
</tr>
<tr>
<td>Silvestri et al (2018) &amp; PANOPTIC</td>
<td>178 patients with 8 to 30 mm lung nodules and low pretest probability</td>
<td>Prospective multicenter observational</td>
<td>Definitive diagnosis, nodule resolution, or 1 year of radiographic stability</td>
<td>Not reported</td>
<td>Retrospective evaluation of performance</td>
<td>Yes</td>
<td>Xpresys Lung version 2</td>
</tr>
</tbody>
</table>

NPV: negative predictive value.

Table 2. Summary of Diagnostic Performance Studies for Proteomic Tests to Predict Malignancy

<table>
<thead>
<tr>
<th>Study</th>
<th>Prevalence, %</th>
<th>Cutoff Value</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, %</th>
<th>NPV, %</th>
<th>PPV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vachani et al (2015)</td>
<td>23.1</td>
<td>0.35</td>
<td>93.2</td>
<td>18.5</td>
<td>90.1</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>23.1</td>
<td>0.34</td>
<td>93.7</td>
<td>18.5</td>
<td>90.1</td>
<td>25.6</td>
</tr>
<tr>
<td></td>
<td>23.1</td>
<td>0.33</td>
<td>94.7</td>
<td>17.6</td>
<td>90.3</td>
<td>25.5</td>
</tr>
<tr>
<td>Silvestri et al (2018)</td>
<td>NR</td>
<td></td>
<td>97 (82 to 100)</td>
<td>44 (36 to 52)</td>
<td>98 (92 to 100)</td>
<td></td>
</tr>
</tbody>
</table>

CI: confidence interval; NPV: negative predictive value; NR: not reported; PPV: positive predictive value.

Limitations of the two validation studies are described in Tables 3 and 4. The primary limitation of the study by Vachani et al (2015) is that the technology is very different from the current marketed version. The primary limitation of the study by Silvestri et al (2018) is that a high number of patients were excluded from the study due to incomplete clinical data or because they were subsequently determined to be outside of the intended use population. It is unclear if the intended use population was determined a priori.

Table 3. Relevance Limitations

<table>
<thead>
<tr>
<th>Study</th>
<th>Population a</th>
<th>Intervention b</th>
<th>Comparator c</th>
<th>Outcomes d</th>
<th>Duration of Follow-Up e</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silvestri et al (2018)</td>
<td>4. The enrolled patients included those who were outside of</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.

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>
<thead>
<tr>
<th>Study</th>
<th>Selectiona</th>
<th>Blindingb</th>
<th>Delivery of Testc</th>
<th>Selective Reportingd</th>
<th>Data Completenesse</th>
<th>Statisticalf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vachani et al (2015)a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silverstri et al (2018)b</td>
<td></td>
<td></td>
<td></td>
<td>2. Data was collected but not reported for the 214 patients with pretest probability &gt;50%</td>
<td>2. A high number of patients (n=234) were excluded</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.

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

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

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.


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

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

**Section Summary: Clinically Valid**

Clinical validation studies were identified for two versions of a proteomic classifier. This classifier has undergone substantial evolution, from a 13 protein assay to a 2 protein assay integrated with clinical factors. Because of this evolution, the most relevant studies are with the most recent version two. One validation study on the version two has been identified. The classifier has been designed to have high specificity for malignant pulmonary nodules, and the validation study showed a specificity of 97% for patients with low to moderate pretest probability (≤ 50%) of
a malignant pulmonary nodule. The primary limitation of this study is that a high number of patients were excluded from the study due to incomplete clinical data or because they were subsequently determined to be outside of the intended use population. It is unclear if the intended use population was determined a priori. Validation in an independent sample in the intended use population is needed.

**Clinically Useful**
A test is clinically useful if the use of the results informs 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.

No evidence directly demonstrating improved outcomes in patients managed with the Xpresys Lung 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.

A chain of evidence was developed, which addresses two key questions: (1) Does the use of a proteomic classifier with high NPV in patients with undiagnosed pulmonary nodules detected by CT change clinical management (in this case, reduction of invasive procedures)? (2) Do those management changes improve outcomes relative to a clinical classifier?

**Changes in Management**
The clinical setting in which a proteomic classifier with high NPV is used, are individuals with undiagnosed pulmonary nodules detected by CT.

Indirect evidence suggests that 36% of invasive procedures (biopsy and/or surgery) on benign nodules could have been avoided, if the test is used in patients with a low to moderate (≤ 50%) pretest probability of malignancy. Three percent of malignant lesions may be missed, although these patients would be followed by CT to verify lack of progression.

**Improved Outcomes**
Indirect evidence suggests that use of a proteomic classifier with high NPV has the potential to reduce the number of unnecessary invasive procedures to definitively diagnose benign disease vs malignancy. Compared with the standard care plan, some patients without cancer will have avoided an unnecessary invasive procedure, which is weighed against the increase in missed cancers in patients
who had lung cancer but tested as negative on the proteomic classifier with high NPV test.

Whether the tradeoff between avoiding unneeded surgeries and the potential for missed cancer is worthwhile depends, in part, on patient and physician preferences. Missed malignancies would likely be continued to be followed by active surveillance using low-dose CT imaging. In the context of lung cancers, overall survival (OS) depends on detection of lung cancer at early, more treatable stages.

Avoiding invasive procedures in situations where patients are at very low likelihood of having lung cancer is likely beneficial, given the known complications of invasive procedures (eg pneumothorax). However, reductions in unnecessary invasive procedures must be weighed against outcomes and harms associated with a missed diagnosis of lung cancer at earlier, more treatable stages.

**Section Summary: Clinically Useful**
Indirect evidence suggests that a proteomic classifier with high NPV has the potential to reduce the number of unnecessary invasive procedures to definitively diagnose benign disease vs malignancy. However, stronger clinical validity data would be needed to rely on indirect evidence for clinical utility.

**Gene Expression Profiling of Indeterminate Bronchoscopy Results**

**Clinical Context and Test Purpose**
The purpose of GEP of bronchial brushings in individuals who undergo bronchoscopy for the diagnosis of suspected lung cancer but who have an indeterminate cytology result is to stratify the clinical risk for malignancy and eliminate the need for invasive diagnostic procedures.

The relevant question addressed in this evidence review is: Does GEP of bronchial brushings reduce the need for invasive diagnostic procedures and lead to improved health outcomes?

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

**Patients**
The relevant population of interest, according to the manufacturer, are individuals with physician-assessed low or intermediate pretest risk of malignancy who are current or former smokers with inconclusive bronchoscopy results for suspected lung cancer.

**Interventions**
The test being considered is GEP of bronchial brushings.

**Comparators**
The following practice is currently being used: standard clinical management without GEP. The management of patients with suspected lung cancer who have
an indeterminate bronchoscopy result is not entirely standardized. However, it is likely that in standard practice many patients would have a surgical biopsy, transthoracic needle aspiration, or another test, depending on the location of the nodule. According to guidelines from the American College of Chest Physicians (2013), in patients with suspected lung cancer with a central lesion, bronchoscopy is recommended to confirm the diagnosis. If bronchoscopy results are nondiagnostic and a suspicion of lung cancer remains, additional testing is recommended (grade 1B recommendation).

Outcomes
The potential beneficial outcomes of primary interest are avoiding an unneeded invasive biopsy of a nodule that would be negative for lung cancer.

Potential harmful outcomes are those resulting from false-positive or false-negative test results. False-positive test results can lead to unnecessary invasive diagnostic procedures and procedure-related complications. False-negative test results can lead to lack of pulmonary nodule surveillance or lack of appropriate invasive diagnostic procedures to diagnose malignancy.

Timing
The time frame for outcomes measures varies from the short-term development of invasive diagnostic procedure-related complications to long-term procedure-related complications, development of malignancy, or OS.

Setting
The primary setting would be in outpatient pulmonology offices.

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).

Whitney et al (2015) reported on the development and initial validation of an RNA-based gene expression classifier from airway epithelial cells designed to be predictive of cancer in current and former smokers undergoing bronchoscopy for suspected lung cancer. Samples were from patients in the Airway Epithelium Gene Expression In the Diagnosis of Lung Cancer (AEGIS) trials, which were two prospective, observational, cohort studies (AEGIS-1, AEGIS-2), for current or former smokers undergoing bronchoscopy for suspected lung cancer. Cohort details are described in Silvestri et al (2015), below. A total of 299 samples from AEGIS-1 (223 cancer-positive and 76 cancer-free subjects) were used to derive the classifier. Data from 123 patients in a prior study with a nondiagnostic
bronchoscopy were used as an independent test set. In the final model, the classifier included 17 genes, patient age, and gene expression correlates and was reported as a dichotomous score (≥0.65 as cancer-positive, <0.65 as cancer-negative). The performance characteristics of the classifier in the training and test set are shown in Table 2.

Silvestri et al (2015) reported on the diagnostic performance of the gene expression classifier developed in Whitney et al (2015), in a sample of 639 patients enrolled in 2 multicenter prospective studies (AEGIS-1, n=298 patients; AEGIS-2, n=341 patients). The study enrolled patients who were undergoing clinically indicated bronchoscopy for a diagnosis of possible lung cancer and had a history of smoking. Before the bronchoscopy, the treating physician assessed each patient's probability of having cancer with a 5-level scale (<10%, 10-39%, 40-60%, 61-85%, >85%). Patients were followed until a diagnosis was established (either at the time of bronchoscopy or subsequently by another biopsy means) or until 12 months after bronchoscopy.

A total of 855 patients in AEGIS-1 and 502 patients in AEGIS-2 met enrollment criteria. After exclusions due to sample quality issues, loss to follow-up, lack of final diagnosis, or nonprimary lung cancer, 341 subjects were available in the validation set for AEGIS-2. For AEGIS-1, patients were randomized to the development (described above) or validation (n=298) sets. Of the 639 patients in the validation study who underwent bronchoscopy, 272 (43%; 95% CI 39 to 46%) had a nondiagnostic examination. The prevalence of lung cancer was 74% and 78% in AEGIS-1 and AEGIS-2, respectively. The overall test characteristics in AEGIS-1 and AEGIS-2 are summarized in Table 2. The classifier improved prediction of cancer compared with bronchoscopy alone but comparisons with a clinical predictor were not reported. For the subset of 272 patients with a nondiagnostic bronchoscopy, the classifier performance was presented by the pretest physician-predicted risk of cancer. For most subpopulations, there was a very high NPV. However, there were 13 false-negatives, 10 of which were considered at high-risk (>60%) of cancer pre-bronchoscopy.

### Table 5. Study Characteristics of Clinical Validity

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Population</th>
<th>Design</th>
<th>Reference Standard</th>
<th>Threshold for Positive Index Test</th>
<th>Timing of Reference and Index Tests</th>
<th>Blinding of Assessors</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silvestri et al (2015)</td>
<td>639 Current or former smokers undergoing bronchoscopy for suspected lung cancer</td>
<td>Prospective, observational, cohort studies</td>
<td>Diagnosis or until 12 months after bronchoscopy</td>
<td>NR</td>
<td>Following diagnosis or 12 months</td>
<td>Yes</td>
<td>272 patients had a nondiagnostic bronchoscopy and were included in the analysis</td>
</tr>
</tbody>
</table>

NR: not reported.
Table 6. Summary of Clinical Validity Studies for GEC to Predict Malignancy in Bronchial Samples

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>AUC (95% CI)</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>PPV, % (95% CI)</th>
<th>NPV, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whitney et al (2015)</td>
<td>Training set, entire population (n=299)</td>
<td>0.78 (0.73 to 0.82)</td>
<td>93</td>
<td>57</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training set, subset with nondiagnostic bronchoscopy (n=134)</td>
<td>0.78 (0.71 to 0.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test set with nondiagnostic bronchoscopy (n=123)</td>
<td>0.81 (0.73 to 0.88)</td>
<td>92 (78 to 98)</td>
<td>53 (42 to 63)</td>
<td>47 (36 to 58)</td>
<td>94 (83 to 99)</td>
</tr>
<tr>
<td>Silvestri et al (2015)</td>
<td>AEGIS-1 (n=298)</td>
<td>0.78 (0.73 to 0.83)</td>
<td>88 (83 to 95)</td>
<td>47 (37 to 58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AEGIS-2 (n=341)</td>
<td>0.74 (0.68 to 0.80)</td>
<td>89 (84 to 92)</td>
<td>47 (36 to 59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subset of all patients with nondiagnostic bronchoscopy, by pretest cancer probability risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Risk &lt;10% (n=61)</td>
<td>7 (1 to 24)</td>
<td>100 (89 to 100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Risk 10%-60% (n=84)</td>
<td>40 (27 to 55)</td>
<td>91 (75 to 98)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Risk &gt;60% (n=108)</td>
<td>84 (75 to 81)</td>
<td>38 (15 to 65)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Risk unknown (n=19)</td>
<td>47 (21 to 73)</td>
<td>100 (40 to 100)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUC: area under the curve; CI: confidence interval; GEC: gene expression classifier; NPV: negative predictive value; PPV: positive predictive value.

Table 7. Relevance Limitations

<table>
<thead>
<tr>
<th>Study</th>
<th>Populationa</th>
<th>Interventionb</th>
<th>Comparatore</th>
<th>Outcomesd</th>
<th>Duration of Follow-Upa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silvestri et al (2015)</td>
<td>4. Only included patients with a history of smoking</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.

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.
&d 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 8. Study Design and Conduct Limitations

<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>Data Completeness&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Statistical&lt;sup&gt;f&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silvestri et al (2015)&lt;sup&gt;9&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2. High number of excluded samples</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> 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> Data Completeness 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.

Vachani et al (2016) reported on rates of invasive procedures from AEGIS-1 and -2. Of 222 patients, 188 (85%) had an inconclusive bronchoscopy and follow-up procedure data available for analysis. Seventy-seven (41%) patients underwent an additional 99 invasive procedures, which included surgical lung biopsy in 40 (52%) patients. Benign and malignant diseases were ultimately diagnosed in 62 (81%) and 15 (19%) patients, respectively. Among those undergoing surgical biopsy, 20 (50%) were performed in patients with benign disease. If the classifier had been used to guide decision making, procedures could have been avoided in 21 (50%) of 42 patients who had additional invasive testing. Further, among 35 patients with an inconclusive index bronchoscopy who were diagnosed with lung cancer, the sensitivity of the classifier was 89%, with 4 (11%) patients having a false-negative classifier result. Invasive procedures after an inconclusive bronchoscopy occur frequently, and most are performed in patients ultimately diagnosed with benign disease.

Section Summary: Clinically Valid

Two multicenter prospective studies have provided evidence of the clinical validity of a bronchial genomic classifier in current or former cigarette smokers undergoing bronchoscopy for suspicion of lung cancer. For patients with intermediate-risk of lung cancer with a nondiagnostic, bronchoscopic examination, the NPV was 91%. However, there has been limited replication outside of a single trial group.
**Clinically Useful**
A test is clinically useful if the use of the results informs 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.

No evidence directly demonstrating improved outcomes in patients managed with the Percepta Bronchial Genomic Classifier (BGC) 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.

A chain of evidence was developed, which addresses two key questions: (1) Does use of the Percepta BGC in individuals with indeterminate bronchoscopy results for suspected lung cancer change clinical management (in this case, reduction of invasive procedures)? (2) Do those management changes improve outcomes?

**Changes in Management**
The clinical setting in which Percepta BGC is meant to be used is not well-defined—individuals who are suspected to have cancer but who have a nondiagnostic bronchoscopy.

One decision impact study reporting on clinical management changes, but not on outcomes after decisions for invasive procedures were made, has suggested that, in at least some cases, decisions for invasive procedures may be changed. Ferguson et al (2016) reported on the impact of the Percepta BGC on physician decision making for recommending invasive procedures among patients with an inconclusive bronchoscopy. The results revealed that a negative (low-risk) result might reduce invasive procedure recommendations in patients diagnosed with benign disease.

**Improved Outcomes**
Indirect evidence suggests that use of the Percepta BGC has the potential to reduce the number of unnecessary invasive procedures to definitively diagnose benign disease vs malignancy. Compared with the standard care plan, some patients without cancer will have avoided an unnecessary invasive procedure, which is weighed against the small increase in missed cancers in patients who had cancer but tested as negative (low-risk) on the Percepta BGC.
Whether the tradeoff between avoiding unneeded surgeries and the potential for missed cancer is worthwhile depends, in part, on patient and physician preferences. Missed malignancies would likely be continued to be followed by active surveillance by low-dose CT imaging. In the context of lung cancers, OS depends on detection of lung cancer at early, more treatable stages.

Avoiding invasive procedures in situations where patients are at very low likelihood of having lung cancer is likely beneficial, given the known complications of invasive procedures (eg, pneumothorax). However, reductions in unnecessary invasive procedures must be weighed against outcomes and harms associated with a missed diagnosis of lung cancer at earlier, more treatable stages.

Section Summary: Clinically Useful
Direct evidence of the clinical utility for GEP of bronchial brushings is lacking. Indirect evidence suggests that Percepta BGC has the potential to reduce the number of unnecessary invasive procedures to definitively diagnose benign disease vs malignancy. However, long-term follow-up data would be required to determine the survival outcomes in patients with a missed diagnosis of lung cancer at earlier, more treatable stages.

Summary of Evidence
For individuals with undiagnosed pulmonary nodules detected by CT who receive plasma-based proteomic screening, the evidence includes prospective cohorts and prospective-retrospective studies. The relevant outcomes are OS, disease-specific survival, test accuracy and validity, morbid events, hospitalizations, and resource utilization. Clinical validation studies were identified for two versions of a proteomic classifier. This classifier has undergone substantial evolution, from a 13 protein assay to a 2 protein assay integrated with clinical factors. Because of this evolution, the most relevant studies are with the most recent version two. One validation study on the version two has been identified. The classifier has been designed to have high specificity for malignant pulmonary nodules, and the validation study showed a specificity of 97% for patients with low to moderate pretest probability (< 50%) of a malignant pulmonary nodule. The primary limitation of this study is that a high number of patients were excluded from the study due to incomplete clinical data or because they were subsequently determined to be outside of the intended use population. It is unclear if the intended use population was determined a priori. Validation in an independent sample in the intended use population is needed. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals with undiagnosed pulmonary nodules following indeterminate bronchoscopy results for suspected lung cancer who receive GEP of bronchial brushings, the evidence includes multicenter prospective studies. The relevant outcomes are OS, disease-specific survival, test accuracy and validity, morbid events, hospitalizations, and resource utilization. Reported receiver operating characteristic curve values ranged from 0.74 to 0.81, with an NPV of 91%. Among patients with a low and intermediate pretest probability of cancer with an inconclusive bronchoscopy, 77 (85%) patients underwent invasive diagnostic
procedures. However, there was a relatively high number of missed cancers. No validation of the test in other populations was identified. Also, where the test would fall in the clinical pathway (ie, other than indeterminate bronchoscopy) is uncertain. The evidence is insufficient to determine the effects of the technology on health outcomes.

**SUPPLEMENTAL INFORMATION**

**Practice Guidelines and Position Statements**

**American College of Chest Physicians**
The American College of Chest Physicians (2013) has published evidence-based clinical practice guidelines on the diagnosis and management of lung cancer, including pulmonary nodules, which is discussed in the patient population parameters in the Plasma-Based Proteomic Screening Of Pulmonary Nodules section.12,

**American Thoracic Society**
The American Thoracic Society (2017) published a position statement on the evaluation of molecular biomarkers for the early detection of lung cancer.5, The Society states that "a clinically useful molecular biomarker applied to the evaluation of lung nodules may lead to expedited therapy for early lung cancer and/or fewer aggressive interventions in patients with benign lung nodules." To be considered clinically useful, a molecular diagnosis "must lead to earlier diagnosis of malignant nodules without substantially increasing the number of procedures performed on patients with benign nodules" or "fewer procedures for patients with benign nodules without substantially delaying the diagnosis of cancer in patients with malignant nodules."

**U.S. Preventive Services Task Force Recommendations**
Not applicable.

**Medicare National Coverage**
MolDX will provide limited coverage for the BDX-XL2 test (Biodesix) for the management of a lung nodule between 8 and 30mm in diameter, in patients at least 40 years of age and with a pre-test cancer risk of 50% or less, as assessed by the Mayo Clinic Model for Solitary Pulmonary Nodules.

**Ongoing and Unpublished Clinical Trials**
A search of ClinicalTrials.gov in April 2018 did not identify any ongoing or unpublished trials that would likely influence this review.

**REFERENCES**


Billing Coding/Physician Documentation Information

**0080U** Oncology (Lung), Mass Spectrometric Analysis of Galectin-3-Binding Protein And Scavenger Receptor Cysteine-Rich Type 1 Protein M130, With Five Clinical Risk Factors (Age, Smoking Status, Nodule Diameter, Nodule-Spiculation Status and Nodule Location), Utilizing Plasma, Algorithm Reported as a Categorical Probability of Malignancy (effective 1/1/19)

**0092U** Oncology (lung), three protein biomarkers, immunoassay using magnetic nanosensor technology, plasma, algorithm reported as risk score for likelihood of malignancy (effective 7/1/19)

**83520** Immunoassay for analyte other than infectious agent antibody or
infectious agent antigen; quantitative, not otherwise specified

84999  Unlisted chemistry procedure

**ICD-10 Codes**

**R91.1**  Solitary pulmonary nodule

---

**Additional Policy Key Words**

EarlyCDT

**Policy Implementation/Update Information**

- **7/1/17**  New Policy. Considered Investigational.
- **1/1/18**  No policy statement changes.
- **7/1/18**  No policy statement changes.
- **1/1/19**  No policy statement changes.
- **7/1/19**  Name of proteomic plasma assay changed to BDX-XL2.

---

State and Federal mandates and health plan contract language, including specific provisions/exclusions, take precedence over Medical Policy and must be considered first in determining eligibility for coverage. The medical policies contained herein are for informational purposes. The medical policies do not constitute medical advice or medical care. Treating health care providers are independent contractors and are neither employees nor agents Blue KC and are solely responsible for diagnosis, treatment and medical advice. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, photocopying, or otherwise, without permission from Blue KC.