Policy Statement

The use of the Presage ST2 Assay to evaluate the prognosis of patients diagnosed with chronic heart failure is considered **investigational**.

The use of the Presage ST2 Assay to guide management (e.g., pharmacologic, device-based, exercise) of patients diagnosed with chronic heart failure is considered **investigational**.

The use of the Presage ST2 Assay in the post cardiac transplantation period is considered **investigational**, including but not limited to:

I. Predicting acute cellular rejection
II. Predicting prognosis

The use of the myTAIHEART assay in the post cardiac transplantation period is considered **investigational**, including but not limited to:

I. Predicting acute cellular rejection
II. Predicting prognosis

The measurement of volatile organic compounds to assist in the detection of moderate grade 2R (formerly grade 3) heart transplant rejection is considered **investigational**.

The use of peripheral blood gene expression profile tests in the management of patients after heart transplantation rejection is considered **investigational**, including but not limited to:

I. Heart transplant graft dysfunction
II. The detection of acute heart transplant

The use of peripheral blood measurement of donor-derived cell-free DNA in the management of patients after renal transplantation is considered **investigational**, including but not limited to:

I. The detection of acute renal transplant rejection
II. Renal transplant graft dysfunction

**NOTE**: Refer to Appendix A to see the policy statement changes (if any) from the previous version.

Policy Guidelines

The U.S. Food and Drug Administration has indicated that the Heartsbreath (Menssana Research) test is only for use as an aid in the diagnosis of grade 3 (now known as grade 2R) heart transplant rejection in patients who have received heart transplants within the preceding year and who have had endomyocardial biopsy within the previous month.

**Coding**

**Effective January 1, 2021**, the following CPT code has been revised:

The following category III CPT code is specific to the Heartsbreath™ test:

- **0085T**: Breath test for heart transplant rejection

The following CPT code is specific to AlloMap®:

- **81595**: Cardiology (heart transplant), mRNA, gene expression profiling by real-time quantitative PCR of 20 genes (11 content and 9 housekeeping), utilizing subfraction of peripheral blood, algorithm reported as a rejection risk score
The following CPT/PLA code is specific to the myTAIHEART™ test:

- **0055U**: Cardiology (heart transplant), cell-free DNA, PCR assay of 96 DNA target sequences (94 single nucleotide polymorphism targets and two control targets), plasma

The following CPT/PLA code is specific to the Molecular Microscope® MMDx-Heart test:

- **0087U**: Cardiology (heart transplant), mRNA gene expression profiling by microarray of 1283 genes, transplant biopsy tissue, allograft rejection and injury algorithm reported as a probability score

The following CPT/PLA code is specific to the Molecular Microscope® MMDx-Kidney test:

- **0088U**: Transplantation medicine (kidney allograft rejection) microarray gene expression profiling of 1494 genes, utilizing transplant biopsy tissue, algorithm reported as a probability score for rejection

The following CPT/PLA code is specific to the Viracor TRACTM dd-cfDNA test:

- **0118U**: Transplantation medicine, quantification of donor-derived cell-free DNA using whole genome next-generation sequencing, plasma, reported as percentage of donor-derived cell-free DNA in the total cell-free DNA

**Description**

Clinical assessment and noninvasive imaging of chronic heart failure can be limited in accurately diagnosing patients with heart failure because symptoms and signs can poorly correlate with objective methods of assessing cardiac dysfunction. For management of heart failure, clinical signs and symptoms (e.g., shortness of breath) are relatively crude markers of decompensation and occur late in the course of an exacerbation. Thus, circulating biomarkers have potential benefit in heart failure diagnosis and management.

In transplant recipients, despite the progress in immunosuppressant therapy, risk of rejection remains. Diagnosis of allograft rejection continues to rely on clinical monitoring and histologic confirmation by tissue biopsy. However, due to limitations of tissue biopsy, including a high degree of interobserver variability in the grading of results and its potential complications, less invasive alternatives have been investigated. Several laboratory-tested biomarkers of transplant rejection have been evaluated and are commercially available for use. The laboratory tests for heart transplant rejection currently evaluated in this policy include the Presage® ST2 Assay kit, which measures the soluble suppression of tumorigenicity-2 protein biomarker; the myTAIHEART assay, which uses cell-free DNA to measure a panel of single nucleotide polymorphisms; the Heartsbreath test, which measures breath markers of oxidative stress; and the AlloMap test, which uses gene expression profiling. Also included in this policy is the AlloSure test, which measures the donor-derived cell-free DNA for renal transplant rejection.

**Related Policies**

- Heart Transplant
- Heart/Lung Transplant

**Benefit Application**

Benefit determinations should be based in all cases on the applicable contract language. To the extent there are any conflicts between these guidelines and the contract language, the contract language will control. 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.
Some state or federal mandates (e.g., Federal Employee Program [FEP]) prohibits plans from denying Food and Drug Administration (FDA)-approved technologies as investigational. In these instances, plans may have to consider the coverage eligibility of FDA-approved technologies on the basis of medical necessity alone.

### Regulatory Status

The U.S. Food and Drug Administration (FDA) has cleared multiple biomarker tests for detection of heart and renal allograft rejection. Table 2 provides a summary of the biomarker tests currently included in this policy that have FDA clearance.

<table>
<thead>
<tr>
<th>Test</th>
<th>Manufacturer</th>
<th>FDAClearance Type, Product Number</th>
<th>FDAClearance Date</th>
<th>Indicated Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heartsbreath™</td>
<td>Menssana Research</td>
<td>Humanitarian device exemption, H030004</td>
<td>2004</td>
<td>To aid in diagnosing grade 3 heart transplant rejection in patients who have received heart transplants within the preceding year. The device is intended as an adjunct to, and not as a substitute for, endomyocardial biopsy and is also limited to patients who have had endomyocardial biopsy within the previous month.</td>
</tr>
<tr>
<td>AlloMap® Molecular</td>
<td>CareDx, formerly XDx</td>
<td>510(k), k073482</td>
<td>2008</td>
<td>The test is to be used in conjunction with clinical assessment, for aiding in the identification of heart transplant recipients with stable allograft function and a low probability of moderate-to-severe transplant rejection. It is intended for patients at least 15 years old who are at least 2 months posttransplant.</td>
</tr>
<tr>
<td>Express Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presage® ST2 Assay kit</td>
<td>Critical Diagnostics</td>
<td>510(k), k093758</td>
<td>2011</td>
<td>For use with clinical evaluation as an aid in assessing the prognosis of patients diagnosed with chronic heart failure</td>
</tr>
</tbody>
</table>

There are also commercially available laboratory-developed biomarker tests for detection of heart and renal allograft rejection. 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. To-date, AlloSure is regulated under the Clinical Laboratory Improvement Amendments standards and all testing is performed at the CareDx reference laboratory.

myTAIHEART is also a laboratory developed test (LDT) developed for clinical diagnostic performance exclusively in the College of American Pathologists (CAP) and Clinical Laboratory Improvement Amendment (CLIA) accredited TAI Diagnostics Clinical Reference Laboratory. This test was developed and its performance characteristics determined by TAI Diagnostics.

Neither test has been cleared or approved by the FDA.
Background

Heart Failure

Heart failure is a major cause of morbidity and mortality worldwide. The term heart failure refers to a complex clinical syndrome that impairs the heart's ability to move blood through the circulatory system. The prevalence of heart failure in the U.S. between 2013 and 2016 was an estimated 6.2 million for Americans ≥20 years old, up from 5.7 million from between 2009 and 2012. Heart failure is the leading cause of hospitalization among people older than age 65 years, with direct and indirect costs estimated at $37 billion annually in the U.S. Although survival has improved with treatment advances, absolute mortality rates of heart failure remain near 50% within 5 years of diagnosis.

Physiology

Heart failure can be caused by disorders of the pericardium, myocardium, endocardium, heart valves or great vessels, or metabolic abnormalities. Individuals with heart failure may present with a wide range of left ventricular (LV) anatomy and function. Some have normal LV size and preserved ejection fraction; others have severe LV dilatation and depressed ejection fraction. However, most patients present with key signs and symptoms secondary to congestion in the lungs from impaired LV myocardial function. They include dyspnea, orthopnea, and paroxysmal dyspnea. Other symptoms include weight gain due to fluid retention, fatigue, weakness, and exercise intolerance secondary to diminished cardiac output.

Diagnosis

Initial evaluation of a patient with suspected heart failure is typically based on clinical history, physical examination, and chest radiograph. Because people with heart failure may present with nonspecific signs and symptoms (e.g., dyspnea), accurate diagnosis can be challenging. Therefore, noninvasive imaging procedures (e.g., echocardiography, radionuclide angiography) are used to quantify pump function of the heart, thus identifying or excluding heart failure in patients with characteristic signs and symptoms. These tests can also be used to assess prognosis by determining the severity of the underlying cardiac dysfunction. However, clinical assessment and noninvasive imaging can be limited in accurately evaluating patients with heart failure because symptoms and signs can poorly correlate with objective methods of assessing cardiac dysfunction. Thus, invasive procedures (e.g., cardiac angiography, catheterization) are used in select patients with presumed heart failure symptoms to determine the etiology (i.e., ischemic vs. nonischemic) and physiologic characteristics of the condition.

Treatment

Patients with heart failure may be treated using a number of interventions. Lifestyle factors such as the restriction of salt and fluid intake, monitoring for increased weight, and structured exercise programs are beneficial components of self-management. A variety of medications are available to treat heart failure. They include diuretics (e.g., furosemide, hydrochlorothiazide, spironolactone), angiotensin-converting enzyme inhibitors (e.g., captopril, enalapril, lisinopril), angiotensin receptor blockers (e.g., losartan, valsartan, candesartan), b-blockers (e.g., carvedilol, metoprolol succinate), and vasodilators (e.g., hydralazine, isosorbide dinitrate). Numerous device-based therapies also are available. Implantable cardioverter defibrillators reduce mortality in patients with an increased risk of sudden cardiac death. Cardiac resynchronization therapy improves symptoms and reduces mortality for patients who have disordered LV conduction evidenced by a wide QRS complex on electrocardiogram. Ventricular assist devices are indicated for patients with end-stage heart failure who have failed all other therapies and are also used as a bridge to cardiac transplantation in select patients.

Heart Failure Biomarkers

Because of limitations inherent in standard clinical assessments of patients with heart failure, a number of objective disease biomarkers have been investigated to diagnose and assess heart failure patient prognosis, with the additional goal of using biomarkers to guide therapy.
include a number of proteins, peptides, or other small molecules whose production and release into circulation reflect the activation of remodeling and neurohormonal pathways that lead to LV impairment. Examples include B-type natriuretic peptide (BNP), its analogue N-terminal pro B-type natriuretic peptide (NT-proBNP), troponin T and I, renin, angiotensin, arginine vasopressin, C-reactive protein, and norepinephrine.\textsuperscript{1,7}

BNP and NT-proBNP are considered the reference standards for biomarkers in assessing heart failure patients. They have had substantial impact on the standard of care for diagnosis of heart failure and are included in the recommendations of all major medical societies, including the American College of Cardiology Foundation and American Heart Association,\textsuperscript{1,7} European Society of Cardiology,\textsuperscript{8} and the Heart Failure Society of America.\textsuperscript{9} Although natriuretic peptide levels are not 100% specific for the clinical diagnosis of heart failure, elevated BNP or NT-proBNP levels in the presence of clinical signs and symptoms reliably identify the presence of structural heart disease due to remodeling and heightened risk for adverse events. Natriuretic peptides also can help in determining prognosis of heart failure patients, with elevated blood levels portending poorer prognosis.

In addition to diagnosing and assessing prognosis of heart failure patients, blood levels of BNP or NT-proBNP have been proposed as an aid for managing patients diagnosed with chronic heart failure.\textsuperscript{1,10,11} Levels of either biomarker rise in response to myocardial damage and LV remodeling, whereas they tend to fall as drug therapy ameliorates symptoms of heart failure. Evidence from a large number of randomized controlled trials (RCTs) that have compared BNP- or NT-proBNP-guided therapy with clinically guided adjustment of pharmacologic treatment of patients who had chronic heart failure has been assessed in recent systematic reviews and meta-analyses. However, these analyses have not consistently reported a benefit for BNP-guided management. Savarese et al (2013) published the largest meta-analysis to date, a patient-level meta-analysis that evaluated 2686 patients from 12 RCTs.\textsuperscript{10} This meta-analysis showed that NT-proBNP-guided management was associated with significant reductions in all-cause mortality and heart failure–related hospitalization compared with clinically guided treatment. Although BNP-guided management in this meta-analysis was not associated with significant reductions in these parameters, differences in patient numbers and characteristics may explain the discrepancy. Troughton et al (2014) conducted a second patient-level meta-analysis that included 11 RCTs with 2000 patients randomized to natriuretic peptide-guided pharmacologic therapy or usual care.\textsuperscript{11} The results showed that, among patients 75 years of age or younger with chronic heart failure, most of whom had impaired left ventricular ejection fraction, natriuretic peptide-guided therapy was associated with significant reductions in all-cause mortality compared with clinically guided therapy. Natriuretic-guided therapy also was associated with significant reductions in hospitalization due to heart failure or cardiovascular disease.

### Suppression of Tumorigenicity-2 Protein Biomarker

A protein biomarker, ST2, has elicited interest as a potential aid to predict prognosis and manage therapy of heart failure.\textsuperscript{12-18} This protein is a member of the interleukin-1 (IL-1) receptor family. It is found as a transmembrane isoform (ST2L) and a soluble isoform (sST2), both of which have circulating IL-33 as their primary ligand. ST2 is a unique biomarker that has pluripotent effects in vivo. Thus, binding between IL-33 and ST2L is believed to have an immunomodulatory function via T-helper type 2 lymphocytes and was initially described in the context of cell proliferation, inflammatory states, and autoimmune diseases.\textsuperscript{13} However, the IL-33/ST2L signaling cascade is also strongly induced through mechanical strain of cardiac fibroblasts or cardiomyocytes. The net result is mitigation of adverse cardiac remodeling and myocardial fibrosis, which are key processes in the development of heart failure.\textsuperscript{20} The soluble isoform of ST2 is produced by lung epithelial cells and cardiomyocytes and is secreted into circulation in response to exogenous stimuli, mechanical stress, and cellular stretch. This form of ST2 binds to circulating IL-33, acting as a “decoy,” thus inhibiting the IL-33-associated antiremodeling effects of the IL-33/ST2L signaling pathway. Thus, on a biologic level, IL-33/ST2L signaling plays a role in modulating the balance of inflammation and neurohormonal activation and is viewed as
pivotal for protection from myocardial remodeling, whereas sST2 is viewed as attenuating this protection. In the clinic, blood concentrations of sST2 appear to correlate closely with adverse cardiac structure and functional changes consistent with remodeling in patients with heart failure, including abnormalities in filling pressures, chamber size, and systolic and diastolic function.7,14,16

An enzyme-linked immunosorbent-based assay is commercially available for determining sST2 blood levels (Presage ST2 Assay).17 The manufacturer claims a limit of detection of 1.8 ng/mL for sST2, and a limit of quantification of 2.4 ng/mL, as determined according to Clinical and Laboratory Standards Institute guideline EP-17-A. Mueller and Dieplinger (2013) reported a limit of detection of 2.0 ng/mL for sST2 in their study.17 In the same study, the assay had a within-run coefficient of variation of 2.5% and a total coefficient of variation less than 4.0%, demonstrated linearity within the dynamic range of the assay calibration curve, and exhibited no relevant interference or cross-reactivity.

The ST2 biomarker is not intended to diagnosis heart failure because it is a relatively nonspecific marker that is increased in many other disparate conditions that may be associated with acute or chronic manifestations of heart failure.16,17 Although the natriuretic peptides (BNP, NT-proBNP) reflect different physiologic aspects of heart failure compared with sST2, they are considered the reference standard biomarkers when used with clinical findings to diagnose, prognosticate, and manage heart failure and as such are the comparator to sST2.

Heart Transplant Rejection
Most cardiac transplant recipients experience at least a single episode of rejection in the first year after transplantation. The International Society for Heart and Lung Transplantation (2005) modified its grading scheme for categorizing cardiac allograft rejection.21 The revised (R) categories are listed in Table 2.

<table>
<thead>
<tr>
<th>New Grade</th>
<th>Definition</th>
<th>Old Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>0R</td>
<td>No rejection</td>
<td></td>
</tr>
<tr>
<td>1R</td>
<td>Mild rejection</td>
<td>1A, 1B, and 2</td>
</tr>
<tr>
<td>2R</td>
<td>Moderate rejection</td>
<td>3A</td>
</tr>
<tr>
<td>3R</td>
<td>Severe rejection</td>
<td>3B and 4</td>
</tr>
</tbody>
</table>

Acute cellular rejection is most likely to occur in the first 6 months after transplantation, with a significant decline in the incidence of rejection after this time. Although immunosuppressants are required on a life-long basis, dosing is adjusted based on graft function and the grade of acute cellular rejection determined by histopathology. Endomyocardial biopsies are typically taken from the right ventricle via the jugular vein periodically during the first 6 to 12 months posttransplant. The interval between biopsies varies among clinical centers. A typical schedule is weekly for the first month, once or twice monthly for the following 6 months, and several times (monthly to quarterly) between 6 months and 1 year posttransplant. Surveillance biopsies may also be performed after the first postoperative year (e.g., on a quarterly or semiannual basis). This practice, although common, has not been demonstrated to improve transplant outcomes. Some centers no longer routinely perform endomyocardial biopsies after 1 year in patients who are clinically stable.

While the endomyocardial biopsy is the criterion standard for assessing heart transplant rejection, it is limited by a high degree of interobserver variability in the grading of results and potential morbidity that can occur with the biopsy procedure. Also, the severity of rejection may not always coincide with the grading of the rejection by biopsy. Finally, a biopsy cannot be used to identify patients at risk of rejection, limiting the ability to initiate therapy to interrupt the development of rejection. For these reasons, an endomyocardial biopsy is considered a flawed criterion standard by many. Therefore, noninvasive methods of detecting cellular rejection have been explored. It is hoped that noninvasive tests will assist in determining appropriate patient management and avoid overuse or underuse of treatment with steroids and other
immunosuppressants that can occur with false-negative and false-positive biopsy reports. Two techniques are commercially available for the detection of heart transplant rejection.

**Noninvasive Heart Transplant Rejection Tests**

**Heartsbreath Test**
The Heartsbreath test, a noninvasive test that measures breath markers of oxidative stress, has been developed to assist in the detection of heart transplant rejection. In heart transplant recipients, oxidative stress appears to accompany allograft rejection, which degrades membrane polyunsaturated fatty acids and evolving alkanes and methylalkanes that are, in turn, excreted as volatile organic compounds in breath. The Heartsbreath test analyzes the breath methylated alkane contour, which is derived from the abundance of C4 to C20 alkanes and monomethylalkanes and has been identified as a marker to detect grade 3 (clinically significant) heart transplant rejection.

**AlloMap**
Another approach has focused on patterns of gene expression of immunomodulatory cells, as detected in the peripheral blood. For example, microarray technology permits the analysis of the expression of thousands of genes, including those with functions known or unknown. Patterns of gene expression can then be correlated with known clinical conditions, permitting a selection of a finite number of genes to compose a custom multigene test panel, which then can be evaluated using polymerase chain reaction techniques. AlloMap is a commercially available molecular expression test that has been developed to detect acute heart transplant rejection or the development of graft dysfunction. The test involves polymerase chain reaction-expression measurement of a panel of genes derived from peripheral blood cells and applies an algorithm to the results. The proprietary algorithm produces a single score that considers the contribution of each gene in the panel. The score ranges from 0 to 40. The AlloMap website states that a lower score indicates a lower risk of graft rejection; the website does not cite a specific cutoff for a positive test. All AlloMap testing is performed at the CareDx reference laboratory in California.

**Presage ST2 Assay**
In addition to its use as a potential aid to predict prognosis and manage therapy of heart failure, elevated serum ST2 levels have also been associated with increased risk of antibody-mediated rejection following heart transplant. For this reason, ST2 has also been proposed as a prognostic marker post heart transplantation and as a test to predict acute cellular rejection (graft-versus-host disease). The Presage ST2 Assay, described above, is a commercially available sST2 test that has been investigated as a biomarker of heart transplant rejection.

**myTAIHEART Biomarker**
Using proprietary myTAIHEART software, the myTAIHEART test uses multiplexed, high-fidelity amplification followed by allele-specific qPCR of a panel of 94 highly informative bi-allelic single nucleotide polymorphisms (SNPs) and two controls to quantitatively genotype cell free DNA in the patient’s plasma after cardiac transplant, and accurately distinguish “donor specific” cell free DNA originating from the engrafted heart from “self-specific” cell free DNA originating from the recipient’s native cells. The ratio of donor specific cell free DNA to total cell free DNA is reported as the donor fraction (%) and categorizes the patient as at low or increased risk of moderate (grade 2R) to severe (grade 3R) acute cellular rejection: low donor fractions indicate less damage to the transplanted heart and a lower risk for rejection, while increased donor fractions indicate more damage to the transplanted heart and an increased risk for rejection. Testing with myTAIHEART does not require a donor specimen. The test is indicated for use in heart transplant recipients who are 2 months of age or older and ≥ 8 days post-transplant, restricted to use in single organ post-heart transplant patients, and is contraindicated in patients who:

- Are Pregnant
- Currently Have Or In The Past Have Had Another Transplanted Organ (Solid Organ Or Allogeneic Bone Marrow)
- Have Post-Transplant Lymphoproliferative Disease
• Have Cancer Or Have Had Cancer Within The Previous 2 Years
• Are On Mechanical Circulatory Support
• Are Closely Related To The Transplant Donor

Other laboratory-tested biomarkers of heart transplant rejection have been evaluated. They include brain natriuretic peptide, troponin, and soluble inflammatory cytokines. Most have had low accuracy in diagnosing rejection. Preliminary studies have evaluated the association between heart transplant rejection and micro-RNAs or high-sensitivity cardiac troponin in cross-sectional analyses but the clinical use has not been evaluated.24,25

Renal Transplant Rejection
Allograft dysfunction is typically asymptomatic and has a broad differential, including graft rejection. Diagnosis and rapid treatment are recommended to preserve graft function and prevent loss of the transplanted organ. For a primary kidney transplant, graft survival at 1 year is 94.7%; at 5 years, graft survival is 78.6%.26

Surveillance of transplant kidney function relies on routine monitoring of serum creatinine, urine protein levels, and urinalysis.27 Allograft dysfunction may also be demonstrated by a drop in urine output or, rarely, as pain over the transplant site. With clinical suspicion of allograft dysfunction, additional noninvasive workup including ultrasonography or radionuclide imaging may be used. A renal biopsy allows a definitive assessment of graft dysfunction and is typically a percutaneous procedure performed with ultrasonography or computed tomography guidance. Biopsy of a transplanted kidney is associated with fewer complications than biopsy of a native kidney because the allograft is typically transplanted more superficially than a native kidney. Renal biopsy is a low-risk invasive procedure that may result in bleeding complications; loss of a renal transplant, as a complication of renal biopsy, is rare.28

Kidney biopsies allow for diagnosis of acute and chronic graft rejection, which may be graded using the Banff Classification.29,30 Pathologic assessment of biopsies demonstrating acute rejection allows clinicians to further distinguish between acute cellular rejection and antibody-mediated rejection, which are treated differently.

Noninvasive Renal Transplant Rejection Tests
Allosure
Cell-free DNA (cfDNA), released by damaged cells, is normally present in healthy individuals.31 In patients who have received transplants, donor-derived cell-free DNA (dd-cfDNA) may also be present. It is proposed that allograft rejection, which is associated with damage to transplanted cells, may result in an increase in dd-cfDNA. AlloSure is a commercially available, next-generation sequencing assay that quantifies the fraction of dd-cfDNA in renal transplant recipients relative to total cfDNA by measuring 266 single nucleotide variants. Separate genotyping of the donor or recipient is not required but patients who receive a kidney transplant from a monozygotic (identical) twin are not eligible for this test. The fraction of dd-cfDNA relative to total cfDNA present in the peripheral blood sample is cited in the report.

Literature Review
Evidence reviews assess whether a medical test is clinically useful. A useful test provides information to make a diagnostic 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.
Use of Soluble Suppression of Tumorigenicity-2 Levels in Chronic Heart Failure Patients

Clinical Context and Test Purpose

The purpose of the Soluble Suppression of Tumorigenicity-2 (sST2) assay is to determine prognosis and/or to guide management in patients with chronic heart failure as an alternative to or an improvement on existing tests and clinical assessment.

The question addressed in this evidence review is: Do sST2 assays determine prognosis and/or guide treatment in patients with chronic heart failure and improve net health outcomes?

The following PICO was used to select literature to inform this review.

Patients
The relevant population of interest is individuals with chronic heart failure.

Interventions
The test being considered is sST2 assay to determine prognosis and/or to guide management. Elevated sST2 levels are purported to predict higher risk of poor outcomes.

Patients with chronic heart failure are actively managed by cardiologists.

Comparators
Comparators of interest include standard prognostic markers, including B-type natriuretic peptide levels and clinical assessment.

Patients with chronic heart failure are actively managed by cardiologists.

Outcomes
The general outcomes of interest are overall survival (OS), quality of life, and hospitalizations. Follow-up of 6-12 months would be appropriate to assess quality of life outcomes.

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 sST2 testing, methodologically credible studies were selected using the following principles:

- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described
- Included a validation cohort separate from the development cohort

Review of Evidence
Subanalyses from Randomized Controlled Trials

A number of clinical studies in which sST2 blood levels were determined using the Presage ST2 Assay have reported that there is an association between ST2 levels and adverse outcomes in patients diagnosed with chronic heart failure. A substantial body of biomarker evidence has been reported retrospectively from subsets of patients enrolled in randomized controlled trials.
(RCTs) of heart failure interventions. These RCTs include the Valsartan Heart Failure Trial (Val-HeFT)\textsuperscript{32}, Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training (HF-ACTION)\textsuperscript{33}, Controlled Rosuvastatin Multinational Trial in Heart Failure (CORONA)\textsuperscript{34}; and ProBNP Outpatient Tailored Chronic Heart Failure study (PROTECT).\textsuperscript{35} Although patients in these RCTs were well-characterized and generally well-matched between study arms, the trials were neither intended nor designed specifically to evaluate biomarkers as risk predictors. At present, no prospectively gathered evidence is available from an RCT in which sST2 levels were compared with levels of a B-type natriuretic peptide (BNP or N-terminal pro B-type natriuretic peptide [NT-proBNP]) to predict risk for adverse outcomes among well-defined cohorts of patients with diagnosed chronic heart failure. Key results of larger individual studies are summarized in Table 3.

**Table 3. Summary of Selected Clinical Studies of sST2 to Predict Outcomes in Chronic Heart Failure Patients**

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Mean Age, y</th>
<th>Study Description and Biomarkers</th>
<th>Primary Endpoints</th>
<th>Mean FU</th>
<th>Synopsis of Findings</th>
</tr>
</thead>
</table>
| Ky et al (2011)\textsuperscript{36} | Ambulatory CHF (N = 1,141, 75% of Penn HF Study population) | 56          | Retrospective analysis of sST2 and NT-proBNP levels and their incremental usefulness over clinical SHFM | Mortality or cardiac transplant | 2.8 y  | • Elevated sST2 levels associated with increased risk (adjusted \(P=0.002\))  
• sST2 in plus NT-proBNP levels showed moderate improvement over SHFM in predicting outcomes (\(P=0.017\)) |
| Bayes-Genis et al (2012)\textsuperscript{37} | Ambulatory decompensated HF (N = 891) | 70          | Retrospective analysis of sST2 and NT-proBNP levels from consecutive series | Mortality                         | 2.8 y  | • Elevated sST2 and NT-proBNP levels provided independent and additive prognostic information for elevated risk of mortality (\(P<0.001\)) |
| Broch et al (2012)\textsuperscript{38} | Ischemic CHF (N = 1,149, 30% of CORONA RCT) | 72          | Retrospective analysis of sST2, NT-proBNP, and CRP levels | CV mortality, nonfatal myocardial infarction or stroke | 2.6 y  | • Elevated sST2 levels independently associated with increased risk for mortality, hospitalization due to HF, or any CV hospitalization (\(P<0.001\))  
• sST2 did not provide additive prognostic information vs. NT-proBNP |
| Felker et al (2013)\textsuperscript{39} | Ambulatory HF (N = 910, 39% of HF-ACTION RCT) | 59          | Retrospective analysis of sST2 and NT-proBNP levels | Mortality, hospitalization, functional capacity | 2.5 y  | • Elevated sST2 levels independently associated with increased risk for mortality, hospitalization due to HF, or any CV hospitalization (\(P<0.001\))  
• sST2 did not provide additive prognostic information vs. NT-proBNP |
<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Mean Age, y</th>
<th>Study Description and Biomarkers</th>
<th>Primary Endpoints</th>
<th>Mean FU</th>
<th>Synopsis of Findings</th>
</tr>
</thead>
</table>
| Gaggin et al (2013) | Recently decompensated CHF (n=151, 100% of PROTECT RCT) | 63 | Retrospective analysis of sST2 and NT-proBNP levels | Composite outcome (worsening HF, hospitalization for HF, clinically significant CV events) | 0.8 y | *Elevated sST2 levels associated with increased risk for adverse CV outcome (P<0.001)*  
* sST2 and NT-proBNP provided independent prognostic information  
* sST2 did not provide additive prognostic information vs. NT-proBNP |
| Anand et al (2014) | CHF (n=1,650, 33% of Val-HeFT RCT) | 63 | Retrospective analysis of sST2, NT-proBNP, and other biomarker levels | All-cause mortality and composite outcome (mortality, SCD with resuscitation, hospitalization for HF, or administration of IV inotropic or vasodilator drug for ≥4 h without hospitalization) |  | *Elevated sST2 levels independently associated with increased risk of poor outcomes (P<0.001)*  
* Baseline sST2 levels did not provide substantial prognostic information when added to a clinical model that included NT-proBNP levels |
<p>| Zhang et al (2015) | De novo HF or decompensated CHF (N=1161) | 58 | Prospective analysis of sST2 in hospitalized sample at 1 center in China | All-cause mortality | 1 y | <em>Elevated sST2 levels independently associated with increased risk of all-cause mortality (P&lt;0.001) after adjustment for clinical risk factors and NT-proBNP levels</em> |
| Dupuy et al (2016) | HF for ≥6 mo (N = 178) | 75 | Prospective analysis of sST2, NT-proBNP, and other biomarker levels in sample from 1 | All-cause mortality and CV mortality | 42 mo&lt;sup&gt;a&lt;/sup&gt; | <em>Elevated sST2 levels independently associated with increased risk for all-cause mortality and CV mortality (P&lt;0.001)</em> |</p>
<table>
<thead>
<tr>
<th>Study Description and Biomarkers</th>
<th>Primary Endpoints</th>
<th>Mean FU</th>
<th>Study Description and Biomarkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>center in France</td>
<td>• In multivariate analysis, sST2 and CRP significantly associated with all-cause mortality and CV mortality</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CHF: chronic heart failure; CRP: C-reactive protein; CV: cardiovascular; FU: follow-up; HF: heart failure; IV: intravenous; NT-proBNP: N-terminal pro B-type natriuretic peptide; RCT: randomized controlled trial; SCD: sudden cardiac death; SHFM: Seattle Heart Failure Model; sST2: soluble suppression of tumorigenicity-2.

**Meta-Analyses**

Aimo et al (2017) pooled findings of studies on the prognostic value of sST2 for chronic heart failure in a meta-analysis. The meta-analysis selected 7 studies, including post hoc analyses of RCTs, and calculated the association between the Presage ST2 Assay and health outcomes. A pooled analysis of 7 studies found that sST2 was a statistically significant predictor of overall mortality (hazard ratio [HR] = 1.75; 95%CI, 1.37-2.22). Moreover, a pooled analysis of 5 studies found that sST2 was a significant predictor of cardiovascular mortality (HR = 1.79; 95% CI, 1.22 to 2.63).

**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 RCTs.

No evidence is available from randomized or nonrandomized controlled studies in which outcomes from groups of well-matched patients managed using serial changes in sST2 blood levels were compared with those managed using the reference standard of BNP or NT-proBNP levels.

**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.

No inferences can be drawn about the clinical utility of sST2 levels for chronic heart failure.

**Section Summary: Use of Soluble Suppression of Tumorigenicity-2 in Chronic Heart Failure Patients**

Several analyses, mostly retrospective, have evaluated whether sST2 levels are associated with disease prognosis, especially mortality outcomes. Studies mainly found that elevated sST2 levels were statistically associated with elevated risk of mortality. A pooled analysis of study results found that sST2 levels significantly predicted overall mortality and cardiovascular mortality. Several studies, however, found that sST2 test results did not provide additional prognostic information compared with BNP or NT-proBNP levels. In general, it appears that elevated sST2 levels predict higher risk of poor outcomes better than lower levels. The available evidence is limited by interstudy inconsistency and differences in patient characteristics, particularly the severity of heart failure, its etiology, duration, and treatment. Furthermore, most of the evidence is limited by interstudy inconsistency and differences in patient characteristics, particularly the severity of heart failure, its etiology, duration, and treatment.
was obtained from retrospective analyses of sST2 levels in subsets of larger patient cohorts within RCTs, potentially biasing the findings. The evidence primarily shows associations between elevated sST2 levels and poor outcomes, but does not go beyond that in demonstrating a clinical connection among biomarker status, treatment received, and clinical outcomes.

**Use of Soluble ST2 Suppression of Tumorigenicity-2 in Post-Heart Transplantation Patients**

**Clinical Context and Test Purpose**

The purpose of sST2 assay is to determine prognosis and/or to predict acute cellular rejection in patients with heart transplantation an alternative to or an improvement on existing tests.

The question addressed in this evidence review is: Does the use of the sST2 assay determine prognosis and/or predict acute cellular rejection in patients undergoing heart transplantation and improve net health outcomes?

The following PICO was used to select literature to inform this review.

**Patients**

The relevant population of interest is individuals with heart transplantation.

Patients undergoing heart transplant are actively managed by cardiologists and transplant specialists.

**Interventions**

The test being considered is sST2 assay to determine prognosis and/or to predict acute cellular rejection.

**Comparators**

Comparators of interest include endomyocardial biopsy for predicting acute cellular rejection. Patients undergoing heart transplant are actively managed by cardiologists and transplant specialists.

**Outcomes**

The general outcomes of interest are OS, quality of life, and hospitalizations.

**Table 4. Significant Outcomes for Post-Heart Transplantation Patients.**

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Details</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morbid events</td>
<td>Short-term and long-term events, such as acute cellular rejection, myocardial infarction, and stroke</td>
<td>30 days, 6 months, 1-5 years</td>
</tr>
<tr>
<td>Hospitalizations</td>
<td>Inpatient hospital admissions</td>
<td>30 days, 6 months, 1-5 years</td>
</tr>
</tbody>
</table>

**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 sST2 testing, methodologically credible studies were selected using the following principles:

- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard
• Patient/sample clinical characteristics were described
• Patient/sample selection criteria were described
• Included a validation cohort separate from the development cohort

Review of Evidence
Serum ST2 levels have been proposed as a prognostic marker post heart transplantation and as a test to predict acute cellular rejection (graft-versus-host disease). There is very little evidence available for these indications. Januzzi et al (2013) retrospectively assessed sST2 levels in 241 patients post–heart transplant. Over a follow-up out to 7 years, sST2 levels were predictive of total mortality (HR = 2.01; 95% CI, 1.15-3.51; P=.01). Soluble ST2 levels were also associated with risk of acute cellular rejection, with a significant difference between the top and bottom quartiles of sST2 levels in the risk of rejection (P=.003).

Pascual-Figal et al (2011), 26 patients with post–cardiac transplantation and an acute rejection episode. Soluble ST2 levels were measured during the acute rejection episode and compared with levels measured when acute rejection was not present. Soluble ST2 levels were higher during the acute rejection episode (130 ng/mL) than during the nonrejection period (50 ng/mL; P=.002). Elevated sST2 levels greater than 68 ng/mL had a positive predictive value of 53% and a negative predictive value of 83% for the presence of acute cellular rejection. The addition of sST2 levels to serum BNP resulted in incremental improvement in identifying rejection episodes.

Table 5. Summary of Key Nonrandomized Clinical Validity Study Characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Type</th>
<th>Country</th>
<th>Dates</th>
<th>Participants</th>
<th>Treatment</th>
<th>Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Januzzi (2013)⁴⁵</td>
<td>Retrospective</td>
<td>United States</td>
<td>NR</td>
<td>Post-cardiac transplantation</td>
<td>sST2 levels assessment (n=241)</td>
<td>Median 7.1 years</td>
</tr>
<tr>
<td>Pascual-Figal (2011)⁴⁶</td>
<td>Retrospective</td>
<td>Spain</td>
<td>2002-2007</td>
<td>Post-cardiac transplantation with acute rejection</td>
<td>sST2 levels assessment (n=26)</td>
<td>Median 3 months</td>
</tr>
</tbody>
</table>

NR: not reported, sST2: soluble suppression of tumorigenicity-2.

Table 6. Summary of Key Nonrandomized Clinical Validity Study Results

<table>
<thead>
<tr>
<th>Study</th>
<th>Total Mortality</th>
<th>ST2 Levels</th>
<th>Positive Predictive Value</th>
<th>Negative Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Januzzi (2013)⁴⁵</td>
<td>2.02 (1.16-3.52)</td>
<td>≥ 30 ng/mL at 7-year follow-up</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>HR (95% CI)</td>
<td>.01</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pascual-Figal (2011)⁴⁶</td>
<td></td>
<td></td>
<td>53%</td>
<td>83%</td>
</tr>
<tr>
<td>Rejection Episode</td>
<td>NR</td>
<td>130 ng/mL (IQR 60-238 ng/mL)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Nonrejection Period</td>
<td>NR</td>
<td>50 ng/mL (IQR 28-80 ng/mL)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>P-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HR: hazard ratio; NA: not applicable; NR: not reported; IQR: interquartile range.

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 RCTs.

No RCTs were identified using sST2 levels that directed patient management in heart transplantation patients and which assessed patient outcomes.

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.

No inferences can be drawn about the clinical utility of sST2 levels for patients with heart transplantation.

Section Summary: Use of Soluble Suppression of Tumorigenicity-2 sST2 in Post-Heart Transplantation Patients
Few studies are available, and they are observational and retrospective. No prospective studies were identified that provide high-quality evidence on the ability of sST2 levels to predict transplant outcomes. One retrospective study (N = 241) found that sST2 levels were associated with acute cellular rejection and mortality; another study (N = 26) found that sST2 levels were higher during an acute rejection episode than before rejection.

Use of myTAIHEART in Post-Heart Transplantation Patients
Clinical Context and Test Purpose
The purpose of myTAIHEART is to determine prognosis and/or to predict acute cellular rejection in patients with heart transplantation as an alternative to or an improvement on existing tests.

The question addressed in this evidence review is: Does the use of the myTAIHEART assay determine prognosis and/or predict acute cellular rejection in patients undergoing heart transplantation and improve net health outcomes?

The following PICO was used to select literature to inform this review.

Patients
The relevant population of interest is individuals with heart transplantation.

Patients undergoing heart transplant are actively managed by cardiologists and transplant specialists.

Interventions
The test being considered is myTAIHEART assay to determine prognosis and/or to predict acute cellular rejection.

Comparators
Comparators of interest include endomyocardial biopsy for predicting acute cellular rejection.

Patients undergoing heart transplant are actively managed by cardiologists and transplant specialists.

Outcomes
The general outcomes of interest are OS, quality of life, and hospitalizations.

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 myTAIHEART testing, methodologically credible studies were selected using the following principles:
• Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
• Included a suitable reference standard
• Patient/sample clinical characteristics were described
• Patient/sample selection criteria were described
• Included a validation cohort separate from the development cohort

Review of Evidence
Serum donor fraction (%), defined as the ratio of donor specific cell free DNA to total cell free DNA, has been proposed as a test to predict acute cellular rejection. In study funded by TAI Diagnostics, Inc., North et al (2020) performed a blinded clinical validation study on 158 matched pairs of endomyocardial biopsy-plasma samples collected from 76 volunteer adult and pediatric heart transplant recipients (ages 2 months or older, and 8 days more more post-transplant) between June of 2010 and Aug 2016 from 2 Milwaukee transplant centers. Based on acute cellular rejection grade as defined by the 2004 International Society for Heart and Lung Transplantation (ISHLT) classification, Receiver Operating Characteristic (ROC) analysis was performed to evaluate diagnostic accuracy across all possible cutoffs. To maximize diagnostic accuracy, Youden’s Index was used to select the optimal cutoff, found to correspond to a donor fraction value of 0.32%. Using this cutoff, clinical performance characteristics of the assay included a negative predictive value (NPV) of 100.00% for grade 2R or higher acute cellular rejection, with 100.00% sensitivity and 75.48% specificity; Area under the Curve (AUC) for this analysis was 0.842, indicative of robust ability of the donor fraction assay to rule out 2R or greater acute cellular rejection for donor fraction values less than 0.32%. There was no statistically significant correlation of donor fraction with age. Donor fraction elevation can also be caused by other forms of injury to the donor heart such as acute cellular rejection 1R, acute antibody-mediated rejection (AMR), and presence of coronary artery vasculopathy (CAV), thereby requiring correlation of myTAIHEART results with other clinical indicators.

In study funded by a grant from the National Institutes of Health and TAI Diagnostics, Inc., Richmond et al (2019) assessed 174 post-cardiac transplant patients from 7 centers (ages 2.4 months-73.4 years) days with myTAIHEART testing ( before transplant; 1, 4, and 7 days following transplant; and at discharge from transplant hospitalization) using blinded analysis of biopsy-paired samples. All the patients were followed for at least 1 year. Donor fraction, defined as the ratio of cell free DNA specific to the transplanted organ to the total amount of cell free DNA present in a blood sample was higher in acute cellular rejection 1R/2R (n = 15) than acute cellular rejection 0R (healthy) (n = 42) (P = 0.02); an optimal donor fraction threshold (0.3%) was determined by the use of Receiver Operating Characteristic (ROC) analysis, revealing an AUC of 0.814 with a sensitivity of 0.65, specificity of 0.93, and an NPV of 81.8% for the absence of any allograft rejection.

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 RCTs.

No studies were identified using myTAIHEART levels that directed patient management in heart transplantation patients and which assessed patient outcomes.

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.

No inferences can be drawn about the clinical utility of myTAIHEART levels for patients with heart transplantation.

Section Summary: Use of myTAIHEART in Post–Heart Transplantation Patients
Few studies are available, and they are observational and of small sample size. No high-quality evidence was found supporting the ability of myTAIHEART levels to predict transplant outcomes. A clinical validation study (n=76) reports donor fraction cutoff value of 0.32% cutoff provides a negative predictive value (NPV) of 100.00% for grade 2R or higher acute cellular rejection, with 100.00% sensitivity and 75.48% specificity; an additional prospective study (N = 174) found that myTAIHEART levels were associated with acute cellular rejection with an optimal donor fraction threshold of 0.3% to rule out the presence of either acute cellular rejection or antibody-mediated rejection.

Measurement of Volatile Organic Compounds for Heart Transplant
Clinical Context and Test Purpose
The purpose of measuring volatile organic compounds in patients with a heart transplant is to assess for heart allograft rejection in a noninvasive manner.

The question addressed in this evidence review is: Does the measurement of volatile organic compounds improve the diagnostic assessment of allograft rejection in heart transplant patients?

The following PICO was used to select literature to inform this review.

Patients
The relevant population of interest are individuals with a heart transplant.

Interventions
The test being considered measures volatile organic compounds to assess for allograft rejection. Patients with a heart transplant are actively managed by cardiologists and transplant specialists; measurement for volatile organic compounds takes place in an outpatient setting.

Comparators
The following test is currently being used to diagnose heart allograft rejection: routine endomyocardial biopsy. Patients with a heart transplant are actively managed by cardiologists and transplant specialists; a routine endomyocardial biopsy is generally performed in an outpatient setting.

Outcomes
The general outcomes of interest are OS, test validity, morbid events, and hospitalizations. Follow-up over months to years is necessary to monitor for signs of allograft rejection.
Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires a 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 the clinical validity of measuring volatile organic compounds, studies that met the following eligibility criteria were considered:
- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard (describe the reference standard)
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described.

Review of Evidence
The U.S. Food and Drug Administration approval of the Heartsbreath test was based on the results of the Heart Allograft Rejection: Detection with Breath Alkanes in Low Levels (HARDBALL) study sponsored by the National Heart, Lung, and Blood Institute. The HARDBALL study was a 3-year, multicenter study of 1061 breath samples in 539 heart transplant patients. Before the scheduled endomyocardial biopsy, patient breath was analyzed by gas chromatography and mass spectroscopy for volatile organic compounds. The amount of C4 to C20 alkanes and monomethylalkanes was used to derive the marker for rejection, known as the breath methylated alkane contour. The breath methylated alkane contour results were compared with subsequent biopsy results, as interpreted by 2 readers using the International Society for Heart and Lung Transplantation biopsy grading system as the criterion standard for rejection.

The authors of the HARDBALL study reported that the abundance of breath markers that measured oxidative stress was significantly greater in grade 0, 1, or 2 rejection than in healthy normal persons. In contrast, in grade 3 rejection, the abundance of breath markers that measure oxidative stress was found to be reduced, most likely due to accelerated catabolism of alkanes and methylalkanes that make up the breath methylated alkane contour. The authors also reported that in identifying grade 3 rejection, the negative predictive value (NPV) of the breath test (97.2%) was similar to endomyocardial biopsy (96.7%) and that the breath test could potentially reduce the total number of biopsies performed to assess for rejection in patients at low-risk for grade 3 rejection. The sensitivity of the breath test was 78.6% vs 42.4% with biopsy. However, the breath test had a lower specificity (62.4%) and a lower positive predictive value (PPV; 5.6%) in assessing grade 3 rejection than a biopsy (specificity, 97%; PPV=45.2%). In addition, the breath test was not evaluated in grade 4 rejection.

Findings from the HARDBALL study were published by Phillips et al (2004). No subsequent studies evaluating the use of the Heartsbreath test to assess for graft rejection were identified in literature updates.

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 (RCTs).

No RCTs assessing the measurement of volatile organic compounds to diagnose cardiac allograft rejection were 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.

Because the clinical validity of measuring volatile organic compounds to assess for cardiac allograft rejection has not been established, a chain of evidence to support clinical utility cannot be constructed.

Section Summary: Measurement of Volatile Organic Compounds for Heart Transplant
A published study found that for identifying grade 3 (now grade 2R) rejection, the NPV of the breath test the study evaluated (97.2%) was similar to endomyocardial biopsy (96.7%), and the sensitivity of the breath test (78.6%) was better than that for biopsy (42.4%). However, the breath test had a lower specificity (62.4%) and a lower PPV (5.6%) in assessing grade 3 rejection than a biopsy (specificity, 97%; PPV=45.2%). The breath test was also not evaluated for grade 4 rejection. At present, no studies evaluating the clinical utility for the measurement of volatile organic compound testing for heart transplant have been identified.

Gene Expression Profiling for Heart Transplant
Clinical Context and Test Purpose
The purpose of the GEP of patients with a heart transplant is to assess for allograft rejection.

The question addressed in this evidence review is: Does the use of GEP improve the diagnostic assessment of allograft rejection in heart transplant patients?

The following PICO was used to select literature to inform this review.

Patients
The relevant population of interest are individuals with heart transplants.

Interventions
The test being considered is GEP to assess for allograft rejection (i.e., AlloMap). Patients with heart transplants are actively managed by cardiologists and transplant specialists; blood samples for GEP are taken in an outpatient setting.

Comparators
The following test is currently being used to diagnose cardiac allograft rejection: routine endomyocardial biopsy. Patients with heart transplants are actively managed by cardiologists and transplant specialists; a routine endomyocardial biopsy is generally performed in an outpatient setting.

Outcomes
The general outcomes of interest are OS, test validity, morbid events, and hospitalizations. Follow-up over months to years to monitor for signs of allograft rejection.

Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires a 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 the clinical validity of GEP testing, studies that met the following eligibility criteria were considered:
- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard (describe the reference standard)
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described.

**Review of Evidence**

**Systematic Reviews**
A Blue Cross Blue Shield Association Technology Evaluation Center (TEC) Assessment (2011) reviewed the evidence on the use of GEP using the AlloMap test. The Assessment concluded that the evidence was insufficient to permit conclusions about the effect of the AlloMap test on health outcomes. Key evidence in the TEC Assessment is described below.

**Nonrandomized Studies**
Patterns of gene expression for the development of the AlloMap test were studied in the Cardiac Allograft Rejection Gene Expression Observation (CARGO) study, which included 8 U.S. cardiac transplant centers enrolling 629 cardiac transplant recipients. The study included the discovery and validation phases. In the discovery phase, patient blood samples were obtained during the endomyocardial biopsy, and the expression levels of more than 7000 genes involved in immune responses were assayed and compared with the biopsy results. A subset of 252 candidate genes was identified, from which a panel of 11 genes was selected for evaluation. A proprietary algorithm was applied to the results, producing a single score that considers the contribution of each gene in the panel.

The validation phase of the CARGO study, published by Deng et al (2006), was prospective, blinded, and enrolled 270 patients. Primary validation was conducted using samples from 63 patients independent from discovery phases of the study and enriched for biopsy-proven evidence of rejection. A prospectively defined test cutoff value of 20 resulted in a sensitivity of 84% of patients with moderate/severe rejection but a specificity of 38%. Of note, in the “training set” used in the study, these rates were 80% and 59%, respectively. The authors evaluated the 11-gene expression profile on 281 samples collected at 1 year or more from 166 patients who were representative of the expected distribution of rejection in the target population (and not involved in discovery or validation phases of the study). When a test cutoff of 30 was used, the NPV (no moderate/severe rejection) was 99.6%; however, only 3.2% of specimens had grade 3 or higher rejection. In this population, grade 1B scores were found to be significantly higher than grade 0, 1A, and 2 scores but were similar to grade 3 scores.

A second prospective multicenter study evaluating the clinical validity of GEP with the AlloMap test (CARGO II) was published by Crespo-Leiro et al (2016). The study enrolled 499 heart transplant recipients undergoing surveillance for allograft rejection. The reference standard for rejection status was histologic grade from an endomyocardial biopsy performed on the same day as blood samples were collected. Blood samples need to be collected 55 days or more posttransplant, more than 30 days after blood transfusion, more than 21 days after administration of prednisone 20 mg/day or more, and more than 60 days after treatment for a prior rejection. Patients had a total of 1579 eligible blood samples for which paired GEP scores and endomyocardial biopsy rejection grades were available.
As in the original CARGO study, the proportion of cases of rejection was small. The prevalence of moderate-to-severe rejection (grade 2R/>3A) reported by local pathologists was 3.2%, which was reduced to 2.0% when confirmation from 1 or more other independent pathologist was required. At a GEP cutoff of 34, for patients who were at least 2 to 6 months posttransplant, the sensitivity of GEP for detecting grade 2R/>3A was 25.0%, and the specificity was 88.7%. The PPV and NPV were 4.0% and 98.4%, respectively. Using the same cutoff of 34, for patients more than 6 months posttransplant, the sensitivity of GEP was 25.0% the specificity was 88.8%, the PPV was 4.3%, and the NPV was 98.3%. The number of true-positives used in the above calculations was 5 (9.1%) of 55 for patients at least 2 to 6 months posttransplant and 6 (10.2%) of 59 for patients more than 6 months posttransplant.

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 RCTs.

Randomized Controlled Trials
Kobashigawa et al (2015) published the results of a pilot RCT evaluating the use of the AlloMap test in patients who were 55 days to 6 months posttransplant. The trial design was similar to that of the Invasive Monitoring Attenuation through Gene Expression (IMAGE) RCT, discussed next. Sixty subjects were randomized to rejection monitoring with AlloMap or with endomyocardial biopsy at prespecified intervals of 55 days and 3, 4, 5, 6, 8, 10, and 12 months posttransplant. The threshold for a positive AlloMap test was set at 30 for patients 2 to 6 months posttransplant and 34 for patients after 6 months posttransplant, based on data from the CARGO study. Endomyocardial biopsy outside of the scheduled visits was obtained in either group if there was clinical or echocardiographic evidence of graft dysfunction and for the AlloMap group if the score was above the specified threshold. The incidence of the primary outcome at 18 months posttransplant (a composite outcome of the first occurrence of any of the following: death or transplant, rejection with hemodynamic compromise, or allograft dysfunction due to other causes) did not differ significantly between the AlloMap and biopsy groups (10% vs 17%; p=0.44). The number of biopsy-proven rejection episodes (International Society for Heart and Lung Transplantation grading system ≥2R) within the first 18 months did not differ significantly between groups (3 in the AlloMap group vs 1 in the biopsy group; p=0.31). Of the rejections in the AlloMap group, 1 was detected after an elevated routine AlloMap test, while 2 were detected after patients presenting with hemodynamic compromise. As in the IMAGE study, a high proportion of rejection episodes were detected by clinical signs or symptoms (however, this study had only 3 rejection episodes in the AlloMap group).

In 2010, the results of the IMAGE study were published. This was an industry-sponsored, nonblinded, noninferiority RCT that compared outcomes in 602 patients managed with the AlloMap test (n=297) or with routine endomyocardial biopsies (n=305). The trial included adults from 13 centers who underwent cardiac transplantation between 1 and 5 years prior to participating, were clinically stable and had a left ventricular ejection fraction of at least 45%. To increase enrollment, the trial protocol was later amended to include patients who had undergone transplantation between 6 months and 1 year prior to participating; this subgroup ultimately comprised only 15% of the final sample (n=87). Each transplant center used its own protocol for determining the intervals for routine testing. At all sites, patients in both groups underwent clinical and echocardiographic assessments in addition to the assigned surveillance strategy. According to the study protocol, patients underwent biopsy if they had signs or symptoms of rejection or allograft dysfunction at clinic visits (or between visits) or if the
Laboratory Tests Post Transplant and for Heart Failure

Echocardiogram showed a left ventricular ejection fraction decrease of at least 25% compared with the initial visit. Additionally, patients in the AlloMap group underwent biopsy if their test score was above a specified threshold; however, if they had 2 elevated scores with no evidence of rejection found on 2 previous biopsies, no additional biopsies were required. The AlloMap test score varied from 0 to 40, with higher scores indicating a higher risk of transplant rejection. The investigators initially used 30 as the cutoff for a positive score; the protocol was amended to use a cutoff of 34 to minimize the number of biopsies needed. Fifteen patients in the AlloMap group and 26 in the biopsy group did not complete the trial.

The primary outcome was a composite variable: (1) the first occurrence of rejection with hemodynamic compromise; (2) graft dysfunction due to other causes; (3) death; or (4) retransplantation. Use of the AlloMap test was considered noninferior to the biopsy strategy if the 1-sided upper boundary of the 95% confidence interval (CI) for the hazard ratio comparing the 2 strategies was less than the prespecified margin of 2.054. The margin was derived using the estimate of a 5% event rate per year in the biopsy group, taken from published observational studies, and allowing for an event rate of up to 10% per year in the AlloMap group.

According to Kaplan-Meier analysis, the 2-year event rate was 14.5% in the AlloMap group and 15.3% in the biopsy group. The corresponding hazard ratio was 1.04 (95% CI, 0.67 to 1.68). The upper boundary of the CI of the hazard ratio (1.68) fell within the prespecified noninferiority margin (2.054); thus, GEP was considered noninferior to endomyocardial biopsy. Death from all causes, a secondary outcome, did not differ significantly between groups. There were 13 (6.3%) deaths in the AlloMap group and 12 (5.5%) in the biopsy group (P=0.82). During follow-up, there were 34 treated episodes of graft rejection in the AlloMap group. Only 6 (18%) of the 34 patients with graft rejection presented solely with elevated AlloMap scores. Twenty (59%) patients presented with clinical signs/symptoms and/or graft dysfunction on echocardiogram and 7 patients had an elevated AlloMap score plus clinical signs/symptoms with or without graft dysfunction on echocardiogram. In the biopsy group, 22 patients were detected solely due to an abnormal biopsy.

A total of 409 biopsies were performed in the AlloMap group and 1249 in the biopsy group. Most biopsies in the AlloMap group (67%) were performed because of elevated gene profiling scores. Another 17% were performed due to clinical or echocardiographic manifestations of graft dysfunction, and 13% were performed as part of routine follow-up after treatment for rejection. There was 1 (0.3%) adverse event associated with biopsy in the AlloMap group and 4 (1.4%) in the biopsy group. In terms of quality of life, the physical health and mental health summary scores of the 12-Item Short-Form Health Survey were similar in the 2 groups at baseline and did not differ significantly between groups at 2 years.

A limitation of the trial was that the threshold for a positive AlloMap test was changed partway through the study; thus, the optimal test cutoff remains unclear. Moreover, the trial was not blinded, which could have affected treatment decisions based on clinical findings, such as whether to recommend a biopsy. In addition, the study did not include a group that only received clinical and echocardiographic assessment, so the value of AlloMap testing beyond that of clinical management alone cannot be determined. The uncertain incremental benefit of the AlloMap test is highlighted by the finding that only 6 of the 34 treated episodes of graft rejection detected during follow-up in the AlloMap group were initially identified solely due to an elevated GEP score. Since 22 episodes of asymptomatic rejection were detected in the biopsy group, the AlloMap test does not appear to be a sensitive test, possibly missing more than half of the episodes of asymptomatic rejection. Because clinical outcomes were similar in the 2 groups, there are at least 2 possible explanations: the clinical outcome of the study may not be sensitive to missed episodes of rejection, or it is not necessary to treat asymptomatic rejection. In addition, the trial was only statistically powered to rule out more than a doubling of the rate of the clinical outcome, which some may believe is an insufficient margin of noninferiority. Finally, only 15% of the final study sample had undergone transplantation less than 1 year before study.
participation; therefore, findings might not be generalizable to the population of patients 6 to 12 months posttransplant.

**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.

Because the clinical validity of GEP testing to assess for cardiac allograft rejection has not been established, a chain of evidence to support clinical utility cannot be constructed.

**Section Summary: Gene Expression Profiling for Heart Transplant**
The 2 studies (CARGO, CARGO II) examining the diagnostic performance of GEP using the AlloMap test for detecting moderate or severe rejection were flawed by lack of a consistent threshold (i.e., 20, 30, or 34) for determining positivity and by a small number of positive cases. In the available studies, although the NPVs were relatively high (i.e., at least 88%), the performance characteristics were calculated based on detection of 10 or fewer cases of rejection each. Moreover, the PPV in the CARGO II study was only 4.0% for patients who were at least 2 to 6 months posttransplant and 4.3% for patients more than 6 months posttransplant.

The most direct evidence on the clinical utility of GEP using the AlloMap test comes from a large RCT comparing a GEP-directed strategy with an endomyocardial biopsy-directed strategy for detecting rejection; it found that the GEP-directed strategy was noninferior. However, given the high proportion of rejection episodes in the GEP-directed strategy group detected by clinical signs/symptoms, the evidence is insufficient to determine that health outcomes are improved because of the uncertain incremental benefit of GEP. In addition, a minority of subjects assessed were in the first year posttransplant. Results from a pilot RCT would suggest that GEP may have a role in evaluating for heart transplant rejection beginning at 55 days posttransplant, but the trial was insufficiently powered to permit firm conclusions about the noninferiority of early GEP use.

**Donor-Derived Cell-Free DNA Testing for Renal Transplant**

**Clinical Context and Test Purpose**
The purpose of dd-cfDNA testing in patients with renal transplant and clinical suspicion of allograft rejection is to detect allograft rejection.

The question addressed in this evidence review is: Does testing for dd-cfDNA improve outcomes in renal transplant patients with clinical suspicion of allograft rejection?

The following PICO was used to select literature to inform this review.

**Patients**
The relevant population of interest are individuals with renal transplants and clinical suspicion of allograft rejection.

**Interventions**
The test being considered is dd-cfDNA testing to assess for renal allograft rejection (i.e., AlloSure). Patients with a renal transplant are actively managed by nephrologists and transplant specialists; dd-cfDNA testing is performed in an outpatient setting.

**Comparators**
The following test is currently being used to confirm a clinical suspicion of allograft rejection: renal biopsy. Patients with a renal transplant are actively managed by nephrologists and transplant specialists; a renal biopsy is performed in an outpatient setting.

**Outcomes**
The general outcomes of interest are OS, test validity, morbid events, and hospitalizations. Follow-up over months to years is needed to monitor for signs of allograft rejection.
Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires a 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 the clinical validity dd-cfDNA testing, studies that met the following eligibility criteria were considered:
- Reported on the accuracy of the marketed version of the technology (including any algorithms used to calculate scores)
- Included a suitable reference standard (describe the reference standard)
- Patient/sample clinical characteristics were described
- Patient/sample selection criteria were described.

Review of Evidence
Development of the AlloSure test was conducted in the multicenter prospective study by Bloom et al (2017), which both recruited patients who were less than 3 months after renal transplant (n=245) and recruited renal transplant patients requiring a biopsy for suspicion of graft rejection (n=139). For the primary analysis, an active rejection was defined as the combined categories of T cell-mediated rejection, acute/active antibody-mediated rejection, and chronic/active antibody-mediated rejection as defined by the Banff working groups. Only patients undergoing biopsy were considered; further exclusion of biopsies that were not for cause had an inadequate or incomplete collection of biopsies or corresponding blood samples or had prior allograft in situ. These exclusions resulted in the main study cohort of 102 patients (107 biopsies).

Within this population, acute rejection was noted in 27 patients (27 biopsies). After statistical analysis accounting for multiple biopsies from the same patient, the threshold dd-cfDNA fraction corresponding to acute rejection was set to 1.0% or higher. In the main study group, this resulted in a sensitivity of 59% (95% CI, 44% to 74%) and specificity of 85% (95% CI, 79% to 81%) for detecting active rejection vs no rejection. Using the original data set including all biopsies performed for clinical suspicion of rejection, 58 cases of acute rejection were diagnosed in 204 biopsies (170 patients). This PPV was 61% and the NPV 84%. Biopsies performed for surveillance (n=34 biopsies) were excluded from analysis in this study, as only 1 biopsy for surveillance demonstrated acute rejection. Study limitations included the absence of a validation data set.

Huang et al (2019) conducted a smaller single center that recruited 63 renal transplant patients with suspicion of rejection that had AlloSure assessment of dd-cfDNA within 30 days of an allograft biopsy. Median years from transplant to dd-dfDNA measurement was 2.0 (interquartile range, 0.3 to 6.5). Within this population, biopsy found acute rejection in 34 (54%) of patients; 10 (15.9%) were cell-mediated only, 22 (25.4%) were antibody-mediated only, and 2 (3.2%) were mixed cell-mediated and antibody-mediated. In contrast to the study by Bloom et al (2017), the optimal threshold for a positive dd-cfDNA result was identified as ≥0.74%. For the outcome of any rejection (i.e., cell-mediated, antibody-mediated, or mixed), use of this threshold was associated with an overall sensitivity of 79.4%, specificity of 72.4%, PPV of 77.1%, and NPV of 75.0%. Discrimination of rejection differed by biopsy findings, however. For the subgroup of patients with antibody-mediated rejection, the sensitivity was 100%, specificity was 71.8%, PPV was 68.6%, and NPV was 100%. The dd-cfDNA test did not discriminate rejection in patients with cell-mediated rejection, as evidenced by an AUC of 0.43 (95% CI, 0.17 to 0.66). Major limitations of this study is its small sample size and single-center setting.
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 RCTs.

No RCTs assessing the clinical utility of the dd-cfDNA (AlloSure) testing to diagnose renal allograft rejection were 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.

Because the clinical validity of dd-cfDNA (AlloSure) testing to assess for renal allograft rejection has not been established, a chain of evidence support clinical utility cannot be constructed.

Section Summary: Donor-Derived Cell-Free DNA Testing for Renal Transplant
A discovery phase prospective study using the AlloSure test has been performed in a multicenter setting. A subsequent smaller single-center study that explored variation in clinical validity based on different rejection mechanisms found the strongest performance characteristics for AlloSure with antibody-mediated rejection. Larger studies validating the dd-cfDNA threshold for active rejection are needed to develop conclusions. At present, no studies evaluating the clinical utility for the dd-cfDNA (AlloSure) testing were identified.

Summary of Evidence
For individuals who have chronic heart failure who receive the sST2 assay to determine prognosis and/or to guide management, the evidence includes correlational studies and 2 meta-analyses. Relevant outcomes are overall survival, quality of life, and hospitalization. Most of the evidence is from reanalysis of existing randomized controlled trials and not from studies specifically designed to evaluate the predictive accuracy of sST2, and prospective and retrospective cross-sectional studies made up a large part of 1 meta-analysis. Studies have mainly found that elevated sST2 levels are statistically associated with elevated risk of mortality. A pooled analysis of study results found that sST2 significantly predicted overall mortality and cardiovascular mortality. Several studies, however, found that sST2 test results did not provide additional prognostic information compared with N-terminal pro B-type natriuretic peptide levels. Moreover, no comparative studies were identified on the use of the sST2 assay to guide management of patients diagnosed with chronic heart failure. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have heart transplantation who receive sST2 assay to determine prognosis and/or to predict acute cellular rejection, the evidence includes a small number of retrospective observational studies on the Presage ST2 Assay. Relevant outcomes are overall survival, morbid events, and hospitalization. No prospective studies were identified that provide high-quality evidence on the ability of sST2 to predict transplant outcomes. One retrospective study \((n = 241)\) found that sST2 levels were associated with acute cellular rejection and mortality; another study \((n = 26)\) found that sST2 levels were higher during an acute rejection episode than before rejection. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have heart transplantation who receive myTAIHEART assay to determine acute cellular rejection, the evidence includes observational studies. A validation study using 158
Laboratory Tests Post Transplant and for Heart Failure

Page 26 of 35

matched endomyocardial biopsy-plasma pairs from 76 pediatric and adult heart transplant recipients (ages 2 months or older, and 8 days more post-transplant) found a donor-specific fraction cutoff (0.32%) that produced a 100% negative predictive value for Grade 2 or higher acute cellular rejection. A prospective observational blinded study (n=174; pediatric=101, adult=73) using biopsy-paired samples found that myTAIHEART level was associated with acute cellular and antibody-mediated rejection in both adult and pediatric heart transplant populations, and that an optimal donor fraction threshold (0.3%) ruled out the presence of either acute cellular rejection or antibody-mediated rejection. Both studies received industry funding. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have a heart transplant who receive a measurement of volatile organic compounds to assess cardiac allograft rejection, the evidence includes a diagnostic accuracy study. Relevant outcomes are overall survival, test validity, morbid events, and hospitalizations. The published study found that, for identifying grade 3 (now grade 2R) rejection, the negative predictive value of the breath test the study evaluated (97.2%) was similar to endomyocardial biopsy (96.7%) and the sensitivity of the breath test (78.6%) was better than that for biopsy (42.4%). However, the breath test had a lower specificity (62.4%) and a lower positive predictive value (5.6%) in assessing grade 3 rejection than a biopsy (specificity, 97%; positive predictive value, 45.2%). The breath test was also not evaluated for grade 4 rejection. This single study is not sufficient to determine the clinical validity of the test measuring volatile organic compounds and no studies on clinical utility were identified. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals who have a heart transplant who receive gene expression profiling (GEP) to assess cardiac allograft rejection, the evidence includes 2 diagnostic accuracy studies and several randomized controlled trials evaluating clinical utility. Relevant outcomes are overall survival, test validity, morbid events, and hospitalizations. The 2 studies, Cardiac Allograft Rejection Gene Expression Observation (CARGO, CARGO II) examining the diagnostic performance of GEP for detecting moderate-to-severe rejection lacked a consistent threshold for defining a positive GEP test (i.e., 20, 30, or 34) and reported a low number of positive cases. In the available studies, although the negative predictive values were relatively high (i.e., at least 88%), the performance characteristics were only calculated based on 10 or fewer cases of rejection; therefore, performance data may be imprecise. Moreover, the positive predictive value in CARGO II was only 4.0% for patients who were at least 2 to 6 months posttransplant and 4.3% for patients more than 6 months posttransplant. The threshold indicating a positive test that seems to be currently accepted (a score of 34) was not prespecified; rather it evolved partway through the data collection period in the Invasive Monitoring Attenuation through Gene Expression (IMAGE) study. In addition, the IMAGE study had several methodologic limitations (e.g., lack of blinding); further, the IMAGE study failed to provide evidence that GEP offers incremental benefit over biopsy performed on the basis of clinical exam or echocardiography. Patients at the highest risk of transplant rejection are patients within 1 year of the transplant, and, for that subset, there remains insufficient data on which to evaluate the clinical utility of GEP. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals with a renal transplant and clinical suspicion of allograft rejection who receive testing of dd-cfDNA to assess renal allograft rejection, the evidence includes small diagnostic accuracy studies. Relevant outcomes are OS, test validity, morbid events, and hospitalizations. One study examined the diagnostic performance of dd-cfDNA for detecting moderate-to-severe rejection; the NPV was moderately high (84%), and performance characteristics were calculated on 27 cases of active transplant rejection. The threshold indicating a positive test was not prespecified. A subsequent smaller single-center study that explored variation in clinical validity based on different rejection mechanisms found the strongest performance characteristics for AlloSure with antibody-mediated rejection. The evidence is insufficient to determine the effects of the technology on health outcomes.
Supplemental Information

Clinical Input From Physician Specialty Societies and Academic Medical Centers

While the various physician specialty societies and academic medical centers may collaborate with and make recommendations during this process, through the provision of appropriate reviewers, input received does not represent an endorsement or position statement by the physician specialty societies or academic medical centers, unless otherwise noted.

2012 Input

In response to requests from Blue Cross Blue Shield Association, input was received from 7 academic medical centers and 1 specialty society in 2012. Input was mixed on whether AlloMap should be investigational. Four reviewers agreed with the investigational status, 1 disagreed, and 3 indicated it was a split decision/other. Reviewers generally agreed that the sensitivity and specificity have not yet been adequately defined for AlloMap and that the negative predictive value was not sufficiently high to preclude the need for biopsy. There was mixed input about the need for surveillance cardiac biopsies to be performed in the absence of clinical signs and/or symptoms of rejection.

2008 Input

In response to requests from Blue Cross Blue Shield Association, input was received from 2 academic medical centers and 2 physician specialty societies in 2008. Three reviewers agreed that these approaches for monitoring heart transplant rejection are considered investigational. The American College of Cardiology disagreed with the policy, stating that the College considers the available laboratory tests to have good potential to diagnose heart transplant rejection and reduce the frequency of invasive biopsies performed on heart transplant patients, although questions remained as to their role in clinical practice.

Practice Guidelines and Position Statements

American College of Cardiology et al

In 2017, the American College of Cardiology Foundation, American Heart Association, and Heart Failure Society published a focused update of their 2013 guideline on the management of heart failure. Part of the focus of the update was on biomarkers. The guidelines stated that soluble suppression of tumorigenicity-2 (ST2) is a biomarker for myocardial fibrosis that may predict hospitalization and death in patients with heart failure and provides additive prognostic information to natriuretic peptide levels. The guidelines were based on a class IIb recommendation (weak; benefit ≥ risk) with level B-NR evidence (moderate-quality, nonrandomized) for the use of ST2 as an option to provide additive prognostic information to established clinical evaluation and biomarkers. The guidelines did not address other uses of ST2 or myTAI HEART.

International Society of Heart and Lung Transplantation

In 2010, the International Society of Heart and Lung Transplantation issued guidelines for the care of heart transplant recipients. The guidelines included the following recommendations (see Table 7).

Table 7. Guidelines for Postoperative Care of Heart Transplant Recipients

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>COR</th>
<th>LOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>“The standard of care for adult HT recipients is to perform periodic EMB during the first 6 to 12 postoperative months for surveillance of HT rejection.”</td>
<td>IIa</td>
<td>C</td>
</tr>
<tr>
<td>“After the first post-operative year, EMB surveillance for an extended period of time (e.g., every 4-6 months) is recommended in HT patients at higher risk for late acute rejection...”</td>
<td>IIa</td>
<td>C</td>
</tr>
<tr>
<td>“Gene Expression Profiling (AlloMap) can be used to rule out the presence of ACR of grade 2R or greater in appropriate low-risk patients, between 6 months and 5 years after HT.”</td>
<td>IIa</td>
<td>B</td>
</tr>
</tbody>
</table>

ACR: acute heart rejection; COR: class of recommendation; EMB: endomyocardial biopsy; HT: heart transplant; LOE: level of evidence.
Kidney Disease Improving Global Outcomes

The Kidney Disease Improving Global Outcomes (2009) issued guidelines for the care of kidney transplant recipients. The guidelines included the following recommendations (see Table 8).

Table 8. Guidelines for Biopsy in Renal Transplant Recipients

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>SOR</th>
<th>LOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>“We recommend kidney allograft biopsy when there is a persistent, unexplained increase in serum creatinine.”</td>
<td>Level 1</td>
<td>C</td>
</tr>
<tr>
<td>“We suggest kidney allograft biopsy when serum creatinine has not returned to baseline after treatment of acute rejection.”</td>
<td>Level 2</td>
<td>D</td>
</tr>
<tr>
<td>“We suggest kidney allograft biopsy every 7-10 days during delayed function.”</td>
<td>Level 2</td>
<td>C</td>
</tr>
<tr>
<td>“We suggest kidney allograft biopsy if expected kidney function is not achieved within the first 1-2 months after transplantation.”</td>
<td>Level 2</td>
<td>D</td>
</tr>
<tr>
<td>“We suggest kidney allograft biopsy when there is new onset of proteinuria.”</td>
<td>Level 2</td>
<td>C</td>
</tr>
<tr>
<td>“We suggest kidney allograft biopsy when there is unexplained proteinuria ≥3.0 g/g creatinine or ≥3.0 g per 24 hours.”</td>
<td>Level 2</td>
<td>C</td>
</tr>
</tbody>
</table>

LOE: level of evidence; SOR: strength of recommendation.

U.S. Preventive Services Task Force Recommendations

Not applicable.

Medicare National Coverage

The Centers for Medicare & Medicaid Services (2008) issued a noncoverage decision for the Heartsbreath test. The Centers determined that the evidence did not adequately define the technical characteristics of the test; nor did it demonstrate that Heartsbreath testing could predict heart transplant rejection, and therefore the test would not improve health outcomes in Medicare beneficiaries.

For AlloMap, Allosure, myTAIHEART, and the Presage ST2 Assay there are no national coverage determinations. In the absence of a national coverage determination, coverage decisions are left to the discretion of local Medicare carriers. Palmetto (2012) conducted a technical assessment and determined that AlloMap met Medicare’s reasonable and necessary criteria. Palmetto GBA and Noridian have local coverage determinations on AlloSure.

Ongoing and Unpublished Clinical Trials

Some currently ongoing and unpublished trials that might influence this review are listed in Table 9.

Table 9. Summary of Key Active 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></td>
<td>AlloMap</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ongoing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC101833195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC01833195</td>
<td>Outcomes AlloMap Registry: the Long-term Management and</td>
<td>2444</td>
<td>Feb 2020 (active, not recruiting)</td>
</tr>
<tr>
<td></td>
<td>Outcomes of Heart Transplant Recipients With AlloMap Testing (OAR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC02178943</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC02178943</td>
<td>Utility of Donor-Derived Cell-free DNA in Association With Gene-</td>
<td>100</td>
<td>Feb 2020 (recruiting)</td>
</tr>
<tr>
<td></td>
<td>Expression Profiling (AlloMap) in Heart Transplant Recipients (D-OAR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AlloSure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ongoing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC03326076</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC03326076</td>
<td>Evaluation of Patient Outcomes From the Kidney Allograft</td>
<td>4000</td>
<td>Dec 2025</td>
</tr>
<tr>
<td></td>
<td>Outcomes AlloSure Registry</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NCT: national clinical trial.

a Denotes industry-sponsored or cosponsored trial.
References


8. McMurray JJ, Adamopoulos S, Anker SD, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail. Aug 2012; 14(8): 803-69. PMID 22828712


49. Blue Cross Blue Shield Technology Evaluation Center (TEC). Gene expression profiling as a noninvasive method to monitor for cardiac allograft rejection. TEC Assessment Program. 2011;26(8).


**Documentation for Clinical Review**

- No records required

**Coding**

This Policy relates only to the services or supplies described herein. Benefits may vary according to product design; therefore, contract language should be reviewed before applying the terms of the Policy. Inclusion or exclusion of codes does not constitute or imply member coverage or provider reimbursement.

<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT®</td>
<td>0055U</td>
<td>Cardiology (heart transplant), cell-free DNA, PCR assay of 96 DNA target sequences (94 single nucleotide polymorphism targets and two control targets), plasma</td>
</tr>
<tr>
<td></td>
<td>0085T</td>
<td>Breath test for heart transplant rejection <em>(Code revision effective 1/1/2021)</em></td>
</tr>
<tr>
<td></td>
<td>0087U</td>
<td>Cardiology (heart transplant), mRNA gene expression profiling by microarray of 1283 genes, transplant biopsy tissue, allograft rejection and injury algorithm reported as a probability score</td>
</tr>
<tr>
<td></td>
<td>0088U</td>
<td>Transplantation medicine (kidney allograft rejection) microarray gene expression profiling of 1494 genes, utilizing transplant biopsy tissue, algorithm reported as a probability score for rejection</td>
</tr>
<tr>
<td></td>
<td>0118U</td>
<td>Transplantation medicine, quantification of donor-derived cell-free DNA using whole genome next-generation sequencing, plasma, reported as percentage of donor-derived cell-free DNA in the total cell-free DNA</td>
</tr>
<tr>
<td></td>
<td>81595</td>
<td>Cardiology (heart transplant), mRNA, gene expression profiling by real-time quantitative PCR of 20 genes (11 content and 9 housekeeping), utilizing subfraction of peripheral blood, algorithm reported as a rejection risk score</td>
</tr>
<tr>
<td></td>
<td>83006</td>
<td>Growth stimulation expressed gene 2 (ST2, Interleukin 1 receptor-like-1)</td>
</tr>
<tr>
<td>HCPCS</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>

**Policy History**

This section provides a chronological history of the activities, updates and changes that have occurred with this Medical Policy.

<table>
<thead>
<tr>
<th>Effective Date</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/05/2007</td>
<td>BCBSA Medical Policy adoption</td>
</tr>
<tr>
<td>03/11/2008</td>
<td>Update CPT code</td>
</tr>
<tr>
<td>10/01/2010</td>
<td>Policy Revision with title change from Heart Transplant Rejection Breath Test</td>
</tr>
<tr>
<td>07/01/2011</td>
<td>Policy revision with position change</td>
</tr>
</tbody>
</table>
### Effective Date | Action
--- | ---
12/15/2014 | Policy title change from Heart Transplant Rejections Laboratory Tests
02/15/2015 | Policy revision with position change
08/31/2015 | Policy revision without position change
02/01/2016 | Administrative Update
07/01/2016 | Policy revision without position change
08/01/2017 | Policy revision without position change
12/01/2017 | Policy revision without position change
07/01/2018 | Coding update
12/01/2018 | Policy title change from Laboratory Tests for Heart Transplant Rejection
07/01/2019 | Coding update
10/01/2019 | Coding update
12/01/2019 | Policy revision without position change
01/01/2021 | Coding update

### Definitions of Decision Determinations

**Medically Necessary:** Services that are Medically Necessary include only those which have been established as safe and effective, are furnished under generally accepted professional standards to treat illness, injury or medical condition, and which, as determined by Blue Shield, are: (a) consistent with Blue Shield medical policy; (b) consistent with the symptoms or diagnosis; (c) not furnished primarily for the convenience of the patient, the attending Physician or other provider; (d) furnished at the most appropriate level which can be provided safely and effectively to the patient; and (e) not more costly than an alternative service or sequence of services at least as likely to produce equivalent therapeutic or diagnostic results as to the diagnosis or treatment of the Member’s illness, injury, or disease.

**Investigational/Experimental:** A treatment, procedure, or drug is investigational when it has not been recognized as safe and effective for use in treating the particular condition in accordance with generally accepted professional medical standards. This includes services where approval by the federal or state governmental is required prior to use, but has not yet been granted.

**Split Evaluation:** Blue Shield of California/Blue Shield of California Life & Health Insurance Company (Blue Shield) policy review can result in a split evaluation, where a treatment, procedure, or drug will be considered to be investigational for certain indications or conditions, but will be deemed safe and effective for other indications or conditions, and therefore potentially medically necessary in those instances.

### Prior Authorization Requirements (as applicable to your plan)

Within five days before the actual date of service, the provider must confirm with Blue Shield that the member's health plan coverage is still in effect. Blue Shield reserves the right to revoke an authorization prior to services being rendered based on cancellation of the member's eligibility. Final determination of benefits will be made after review of the claim for limitations or exclusions.

Questions regarding the applicability of this policy should be directed to the Prior Authorization Department at (800) 541-6652, or the Transplant Case Management Department at (800) 637-2066 ext. 3507708 or visit the provider portal at www.blueshieldca.com/provider.
### POLICY STATEMENT

<table>
<thead>
<tr>
<th>BEFORE</th>
<th>AFTER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Policy Statement:</strong></td>
<td><strong>Policy Statement:</strong></td>
</tr>
<tr>
<td>The measurement of volatile organic compounds to assist in the detection of moderate grade 2R (formerly grade 3) heart transplant rejection is considered <em>investigational</em>.</td>
<td>The use of the Presage ST2 Assay to evaluate the prognosis of patients diagnosed with chronic heart failure is considered <em>investigational</em>.</td>
</tr>
</tbody>
</table>
| The use of peripheral blood gene expression profile tests in the management of patients after heart transplantation is considered *investigational*, including but not limited to:  
  - Heart transplant graft dysfunction  
  - The detection of acute heart transplant rejection | The use of the Presage ST2 Assay to guide management (e.g., pharmacologic, device-based, exercise) of patients diagnosed with chronic heart failure is considered *investigational*. |
| The use of peripheral blood measurement of donor-derived cell-free DNA in the management of patients after renal transplantation is considered *investigational*, including but not limited to:  
  - The detection of acute renal transplant rejection  
  - Renal transplant graft dysfunction | The use of the Presage ST2 Assay in the post cardiac transplantation period is considered *investigational*, including but not limited to:  
  I. Predicting acute cellular rejection  
  II. Predicting prognosis |
| The measurement of volatile organic compounds to assist in the detection of moderate grade 2R (formerly grade 3) heart transplant rejection is considered *investigational*. | The use of peripheral blood gene expression profile tests in the management of patients after heart transplantation rejection is considered *investigational*, including but not limited to:  
  I. Heart transplant graft dysfunction  
  II. The detection of acute heart transplant rejection |
| The use of peripheral blood measurement of donor-derived cell-free DNA in the management of patients after renal transplantation is considered *investigational*, including but not limited to:  
  I. The detection of acute renal transplant rejection  
  II. Renal transplant graft dysfunction | The use of the myTAIHEART assay in the post cardiac transplantation period is considered *investigational*, including but not limited to:  
  I. Predicting acute cellular rejection  
  II. Predicting prognosis |