Laboratory Tests for Heart and Kidney Transplant Rejection

DISCLAIMER

Our medical policies are designed for informational purposes only and are not an authorization, explanation of benefits or a contract. Receipt of benefits is subject to satisfaction of all terms and conditions of the coverage. Medical technology is constantly changing, and we reserve the right to review and update our policies periodically.

POLICY

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, including but not limited to the detection of acute heart transplant rejection or heart transplant graft dysfunction, is considered investigational.

The use of peripheral blood measurement of donor-derived cell-free DNA in the management of patients after renal transplantation, including but not limited to the detection of acute renal transplant rejection or renal transplant graft dysfunction, is considered investigational.

POLICY GUIDELINES

The U.S. Food and Drug Administration (FDA) 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

See the Codes table for details.

BENEFIT APPLICATION

BLUECARD/NATIONAL ACCOUNT ISSUES
State or federal mandates (eg, Federal Employee Program) may dictate that certain U.S. Food and Drug Administration–approved devices, drugs, or biologics may not be considered investigational, and thus these devices may be assessed only by their medical necessity.

Monitoring of heart transplantation rejection is a specialized procedure that may require an out-of-network referral. Specifically, the AlloMap test is only performed in the company's laboratory.
BACKGROUND

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. The revised (R) categories are listed in Table 1.

Table 1. Revised Grading Schema for Cardiac Allograft Rejection

<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>

Surveillance
Acute cellular rejection is most likely to occur in the first 6 months, 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 endomyocardial biopsy is the criterion standard for assessing heart transplant rejection, it is limited by a high degree of interobserver variability in 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, 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
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.
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.

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.

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%.

Surveillance
Surveillance of transplant kidney function relies on routine monitoring of serum creatinine, urine protein levels, and urinalysis. 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.

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

Donor-Derived Cell-Free DNA
Cell-free DNA (cfDNA), released by damaged cells, is normally present in healthy individuals. In patients who have received transplants, donor-derived cfDNA (dd-cfDNA) may be also 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. All AlloSure testing is performed at the CareDx reference laboratory.

REGULATORY STATUS
In 2004, the Heartsbreath™ test (Menssana Research) was cleared for marketing by the U.S. Food and Drug Administration through a humanitarian device exemption for use as an 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.

In 2008, AlloMap® Molecular Expression Testing (CareDx, formerly XDx) was cleared for marketing by the Food and Drug Administration through the 510(k) process. The Food and Drug Administration determined that this device was substantially equivalent to existing devices, 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.

RATIONALE
This evidence review was created in November 2004 and has been updated regularly with searches of the MEDLINE database. The most recent literature update was performed through August 22, 2018.

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

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

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.

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 PICOTS were used to select literature to inform this review.

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

Interventions
The test being considered measures volatile organic compounds to assess for allograft rejection.
Comparators
The following test is currently being used to diagnose heart allograft rejection: routine endomyocardial biopsy.

Outcomes
The general outcomes of interest are overall survival, test validity, morbid events, and hospitalizations.

Timing
Follow-up over months to years is to monitor for signs of allograft rejection.

Setting
Patients with a heart transplant are actively managed by cardiologists and transplant specialists.

Study Selection Criteria
For the evaluation of 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.

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).

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 were found to be 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 were 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 finding 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: Clinically Useful**
The 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.

**GENE EXPRESSION PROFILING FOR HEART TRANSPLANT**

**Clinical Context and Test Purpose**
The purpose of the gene expression profiling (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 PICOTS were used to select literature to inform this review.

**Patients**
The relevant population of interest is individuals with heart transplants.
Interventions
The test being considered is GEP to assess for allograft rejection (ie, AlloMap).

Comparators
The following test is currently being used to diagnose cardiac allograft rejection: routine endomyocardial biopsy.

Outcomes
The general outcomes of interest are overall survival, test validity, morbid events, and hospitalizations.

Timing
Follow-up over months to years to monitor for signs of allograft rejection.

Setting
Patients with heart transplant are actively managed by cardiologists and transplant specialists.

Study Selection Criteria
For the evaluation of 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.

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).

Systematic Reviews
A 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 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 one 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.

**Section Summary: Clinically Valid**
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 (ie, 20, 30, or 34) for determining positivity and a small number of positive cases. In the available studies, although the NPVs were relatively high (ie, 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.

**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: 60 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 (composite outcome of the first occurrence of any of the following: death or retransplant, 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, one was detected after an elevated routine AlloMap test, while two 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, 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 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 impacted treatment decisions such as whether to recommend a biopsy, based on clinical findings. In addition, the study did not include a group that only received clinical and echocardiographic assessment, and therefore, 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 GPS 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: Clinically Useful**
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
Laboratory Tests for Heart and Kidney Transplant Rejection

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 donor-derived cell-free DNA (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 PICOTS were used to select literature to inform this review.

**Patients**
The relevant population of interest is 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 (ie, AlloSure).

**Comparators**
The following test is currently being used to confirm clinical suspicion of allograft rejection: renal biopsy.

**Outcomes**
The general outcomes of interest are overall survival, test validity, morbid events, and hospitalizations.

**Timing**
Follow-up over months to years is to monitor for signs of allograft rejection.

**Setting**
Patients with a renal transplant are actively managed by nephrologists and transplant specialists.

**Study Selection Criteria**
For the evaluation of clinical validity of 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.

**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)

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, 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 which were not for cause, had an inadequate or incomplete collection of biopsies or corresponding blood samples, or had prior allograft in situ resulted in the main study cohort (N=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 one biopsy for surveillance demonstrated acute rejection. Study limitations included the absence of a validation data set.

Section Summary: Clinically Valid
A discovery phase prospective study using the AlloSure test has been performed in a multicenter setting. Larger studies validating the dd-cfDNA threshold for active rejection are needed to develop conclusions.

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: Clinically Useful
At present, no studies evaluating the clinical utility for the dd-cfDNA (AlloSure) testing were identified.
SUMMARY OF EVIDENCE
For individuals who have a heart transplant who receive 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 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 (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 (ie, 20, 30, or 34) and reported a low number of positive cases. In the available studies, although the negative predictive values were relatively high (ie, 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 IMAGE study. In addition, the IMAGE study had several methodologic limitations (eg, 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 a diagnostic accuracy study. Relevant outcomes are overall survival, test validity, morbid events, and hospitalizations. The study examined the diagnostic performance of dd-cfDNA for detecting moderate-to-severe rejection; the negative predictive value 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. 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.

Original Policy Date: November 2004
2012 Input
In response to requests, input was received from 7 academic medical centers and 1 specialty society while this policy was under review in 2012. Input was mixed on whether AlloMap should be investigational. Four reviewers agreed with the investigational status, one disagreed, and three 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, input was received from 2 academic medical centers and 2 physician specialty societies while this policy was under review 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

International Society of Heart and Lung Transplantation
The International Society of Heart and Lung Transplantation (2010) issued guidelines for the care of heart transplant recipients. The guidelines included the following recommendations (see Table 2).

Table 2. 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 post-operative months for surveillance of HT rejection.”</td>
<td>Ila</td>
<td>C</td>
</tr>
<tr>
<td>“After the first post-operative year, EMB surveillance for an extended period of time (eg, every 4–6 months) is recommended in HT patients at higher risk for late acute rejection.”</td>
<td>Ila</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>Ila</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 3).

Table 3. 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>
</tbody>
</table>
“We suggest kidney allograft biopsy when there is new onset of proteinuria.” Level 2 C
“We suggest kidney allograft biopsy when there is unexplained proteinuria ≥3.0 g/g creatinine or ≥3.0 g per 24 hours.” Level 2 C

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.\(^\text{21}\) 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, there is no national coverage determination. 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.

For AlloSure, there is no national coverage determination.\(^\text{22}\) In the absence of a national coverage determination, coverage decisions are left to the discretion of local Medicare carriers. Palmetto GBA and Noridian have local coverage determinations on AlloSure.

ONGOING AND UNPUBLISHED CLINICAL TRIALS
Some currently unpublished trials that might influence this review are listed in Table 4.

Table 4. 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>NCT01833195(^a)</td>
<td>Outcomes AlloMap Registry: the Long-term Management and Outcomes of Heart Transplant Recipients With AlloMap Testing (OAR)</td>
<td>2000</td>
<td>Dec 2018</td>
</tr>
<tr>
<td>NCT02178943(^a)</td>
<td>Utility of Donor-Derived Cell-free DNA in Association With Gene-Expression Profiling (AlloMap®) in Heart Transplant Recipients (D-OAR)</td>
<td>100</td>
<td>July 2019</td>
</tr>
<tr>
<td>NCT03326076(^a)</td>
<td>Evaluation of Patient Outcomes From the Kidney Allograft Outcomes AlloSure Registry (KOAR)</td>
<td>1000</td>
<td>Dec 2022</td>
</tr>
</tbody>
</table>

NCT: national clinical trial.
\(^a\) Denotes industry-sponsored or cosponsored trial.

REFERENCES

**CODES**

<table>
<thead>
<tr>
<th>Codes</th>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT</td>
<td>0085T</td>
<td>Breath test for heart transplant rejection</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>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>HCPCS</td>
<td>No codes</td>
<td></td>
</tr>
<tr>
<td>ICD-10-CM</td>
<td>T86.20-T86.298</td>
<td>Complications of heart transplant code range,</td>
</tr>
<tr>
<td></td>
<td>Z48.21</td>
<td>Encounter for aftercare following heart transplant</td>
</tr>
<tr>
<td></td>
<td>Z94.1</td>
<td>Heart transplant status</td>
</tr>
<tr>
<td>ICD-10-PCS</td>
<td>Not applicable. ICD-10-PCS codes are only used for inpatient services. There are no ICD procedure codes for laboratory tests.</td>
<td></td>
</tr>
</tbody>
</table>

**POLICY HISTORY**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/10/14</td>
<td>Replace policy</td>
<td>Policy updated with literature review through March 4, 2014; reference 9 added; policy statements edited for clarity.</td>
</tr>
<tr>
<td>04/23/15</td>
<td>Replace policy</td>
<td>Policy updated with literature review through March 5, 2015. References 2-3 and 12 added. Policy statements unchanged.</td>
</tr>
<tr>
<td>05/19/16</td>
<td>Replace policy</td>
<td>Policy updated with literature review through March 21, 2016; references 1 and 8 added. In first policy statement, “grade 3” changed to “grade 2R/grade 3” due to updated ISHLT rejection grades and brand name of test removed; intent of statements unchanged.</td>
</tr>
<tr>
<td>06/01/17</td>
<td>Replace policy</td>
<td>Blue Cross of Idaho annual review; no change to policy.</td>
</tr>
<tr>
<td>10/30/17</td>
<td>Replace policy</td>
<td>Blue Cross of Idaho adopted changes to policy as noted. Policy updated with literature review through August 28, 2017; no references added; reference 2 updated. Policy statements unchanged.</td>
</tr>
<tr>
<td>10/18/18</td>
<td>Replace policy</td>
<td>Blue Cross of Idaho adopted changes as noted, effective 01/25/2019. Policy updated with literature review through August 22, 2018; references 5-9, 18, 20, and 22 added. Policy updated with literature review through August 22, 2018; references 5-9, 18, 20, and 22 added.</td>
</tr>
</tbody>
</table>
statement added that “The use of peripheral blood measurement of donor-derived cell-free DNA in the management of patients after renal transplantation, including but not limited to the detection of acute renal transplant rejection or renal transplant graft dysfunction, is considered investigational.” Title expanded to include kidney transplant rejection.