Laboratory Tests for Heart Transplant Rejection

Description

Several commercially available laboratory tests assess heart transplant rejection including the Heartsbreath test, which measures breath markers of oxidative stress, and the AlloMap™ test, which conducts gene expression profiling (GEP). These tests are proposed as an alternative to, or adjunct to, endomyocardial biopsy, which is invasive, and its interpretation may have high interobserver variability.

Background

Heart Transplant Rejection

Most cardiac transplant recipients experience at least one episode of rejection in the first year after transplantation. 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 post-transplant. 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 post-transplant. Surveillance biopsies may also be performed after the first postoperative year e.g., on a quarterly or semi-annual 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 the 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, 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 have become commercially available for the detection of heart transplant rejection.
NonInvasive Heart Transplant Rejection Tests
The Heartsbreath test (Menssana Research, Inc.), 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 (BMAC), which is derived from the abundance of C4-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 gene expression of thousands of genes, including those with functions that are 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 (PCR) techniques. AlloMap™ (XDx, Inc.) 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 PCR-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 XDx 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. (1) All AlloMap testing is performed at the XDx reference laboratory in Brisbane, CA.

Other laboratory-tested biomarkers of heart transplant rejection have been evaluated. These include brain natriuretic peptide, troponin, and soluble inflammatory cytokines. Most of these have had low diagnostic 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. (2,3)

Regulatory Status
In February 2004, the Heartsbreath test (Menssana Research, Inc., Newark, NJ) received approval from the U.S. Food and Drug Administration (FDA) through a Humanitarian Device Exemption. The Heartsbreath test is indicated for use as an aid in the diagnosis of grade 3 heart transplant rejection in patients who have received a heart transplant within the preceding year. The device is intended to be used as an adjunct to, and not as a substitute for, endomyocardial biopsy and is limited to patients who have had endomyocardial biopsy within the previous month.

In August 2008, AlloMap Molecular Expression Testing (XDx, Inc., Brisbane, CA) was cleared for marketing by the FDA through the 510(k) process. The FDA 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 that have a low probability of moderate/severe transplant rejection. It is intended for patients at least 15 years-old who are at least 2 months post-transplant.
Related Policies

7.03.08    Heart/Lung Transplant
7.03.09    Heart Transplant

Policy

*This policy statement applies to clinical review performed for pre-service (Prior Approval, Precertification, Advanced Benefit Determination, etc.) and/or post-service claims.

The measurement of volatile organic compounds with the Heartsbreath test to assist in the detection of grade 3 heart transplant rejection is considered **not medically necessary**.

The use of peripheral blood genetic profiling 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 **not medically necessary**.

Policy Guideline

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

Rationale

Heartsbreath Test

Approval of the Heartsbreath test by the U.S. Food and Drug Administration (FDA) 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. (4) The HARDBALL study was a 3-year, multicenter study of 1,061 breath samples in 539 heart transplant patients. Before scheduled endomyocardial biopsy, patient breath was analyzed by gas chromatography and mass spectroscopy for volatile organic compounds. The amount of C4-C20 alkanes and monomethylalkanes was used to derive the marker for rejection, known as the breath methylated alkane contour (BMAC). The BMAC results were compared with subsequent biopsy results, as interpreted by two readers using the International Society for Heart and Lung Transplantation (ISHLT) biopsy grading system as the criterion standard for rejection.

The authors of the HARDBALL study reported that the abundance of breath markers of oxidative stress were significantly greater in grade 0, 1, or 2 rejection than in healthy normal persons. Whereas in grade 3 rejection, the abundance of breath markers of oxidative stress was reduced, most likely due to accelerated catabolism of alkanes and methylalkanes that make up the BMAC. The authors also reported finding that in identifying grade 3 rejection, the negative predictive value (NPP) 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% versus 42.4% with biopsy. However, the breath test had lower specificity (62.4%) and a lower positive predictive value (PPV) (5.6%) in assessing grade 3 rejection than 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 in 2004. No subsequent studies that evaluated use of the Heartsbreath test to assess for graft rejection were identified in literature reviews for policy updates.

**AlloMap™ Test**

A 2011 TEC Assessment reviewed the evidence on the use of AlloMap testing. (5) The Assessment concluded that the evidence is insufficient to permit conclusions about the effect of the AlloMap™ test on health outcomes. Key evidence is described below:

**Analytic/Clinical Validity**

Patterns of gene expression for 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 650 cardiac transplant recipients. (6) The study included discovery and validation phases. In the discovery phase, patient blood samples were obtained at the time of endomyocardial biopsy, and the expression levels of more than 7,000 genes known to be involved in immune responses were assayed and compared to the biopsy results. A subset of 200 candidate genes were identified that showed promise as markers that could distinguish transplant rejection from quiescence, and from there, a panel of 11 genes was selected that could be evaluated using polymerase chain reaction (PCR) assays. A proprietary algorithm is applied to the results of the analysis, producing a single score that considers the contribution of each gene in the panel.

The validation phase of the CARGO study, published in 2006, was prospective, blinded, and enrolled 270 patients. (6) 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 NPP (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 similar to grade 3 scores. The sensitivity and specificity for determining quiescent versus early stages of rejection was not addressed.

Post-CARGO clinical observations have also been published. (7) The multicenter work group identified a number of factors that can affect AlloMap scores, including the time post-transplant, corticosteroid dosing, and transplant vasculopathy. (7, 8) Scores of 34 and greater were considered positive. Analysis of data from a number of centers collected post-CARGO showed that, at 1 year or more post-transplantation, an AlloMap threshold of 34 had a PPV of 7.8% for scores of 3A/2R or greater on biopsy and a NPP of 100% for AlloMap scores below 34. There is insufficient information in this study to
determine whether there are potential study biases in this report. These findings were limited due to a very low number of rejection events; only 5 biopsy samples (2.4%) were found to have a grade of 2R or greater. At 1 year, 28% of the sample showed an elevated AlloMap score (>34) even though there was absence of evidence of rejection on biopsy. The significance of chronically elevated AlloMap scores in the absence of clinical manifestation of graft dysfunction and the actual impact on the number of biopsies performed is currently unknown.

Section Summary
In sum, the studies examining the diagnostic performance of AlloMap testing for detecting moderate/severe rejection are flawed by lack of a consistent threshold for determining positivity and very small sample sizes. The studies that examined cutoff scores of 30 or 34, calculated sensitivities of 80-100%, based on detecting 10 or fewer cases of rejection in each of 3 studies. (5)

Clinical Utility
In 2010, results of the Invasive Monitoring Attenuation through Gene Expression (IMAGE) study were published. (9, 10) This was an industry-sponsored noninferiority randomized controlled trial (RCT) that compared outcomes in 602 patients managed with the AlloMap test (n=297) or routine endomyocardial biopsies (n=305). This study was not blinded. The study included adult patients from 13 centers who underwent cardiac transplantation between 1 and 5 years previously, were clinically stable, and had a left ventricular ejection fraction (LVEF) of at least 45%. To increase enrollment, the study protocol was later amended to include patients who had undergone transplantation between 6 months and 1 year earlier; this sub-group 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 LVEF decrease of at least 25% compared to 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 varies 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 later 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 study.

The primary outcome was a composite variable; the first occurrence of 1) rejection with hemodynamic compromise, 2) graft dysfunction due to other causes, 3) death, or 4) retransplantation. The trial was designed to test the noninferiority of GEP with the AlloMap test compared to endomyocardial biopsies with respect to the primary outcome. Use of the AlloMap test was considered noninferior to the biopsy strategy if the one-sided upper boundary of the 95% confidence interval (CI) for the hazard ratio (HR) 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 in the biopsy group, taken from published observational studies, and allowing for an event rate of up to 10% in the AlloMap group. Secondary outcomes included death, the number of biopsies performed, biopsy-related complications, and quality-of-life using the 12-item short-form (SF-12).
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 HR was 1.04 (95% CI: 0.67 to 1.68). The upper boundary of the CI of the HR, 1.68, fell within the prespecified noninferiority margin (2.054); thus, GEP was considered noninferior to endomyocardial biopsy. Median follow-up was 19 months. The number of patients remaining in the Kaplan-Meier analysis after 300 days was 221 in the biopsy group and 207 in the AlloMap group; the number remaining after 600 days was 137 and 133, respectively. The secondary outcome, death from all-causes at any time during the study, did not differ significantly between groups. There were a total of 13 (6.3%) deaths in the AlloMap group and 12 (5.5%) in the biopsy group (p=0.82). During the follow-up period, there were 34 treated episodes of graft rejection in the AlloMap group. Only 6 of the 34 (18%) patients with rejection presented solely with an elevated AlloMap score. Twenty patients (59%) 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 1,249 in the biopsy group. Most of the 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 SF-12 were similar in the 2 groups at baseline and did not differ significantly between groups at 2 years.

A limitation of the study was that the threshold for a positive AlloMap test was changed partway through the study; thus, the optimal test cutoff remains unclear. Moreover, the study was not blinded which could have impacted treatment decisions such as whether or not to recommend 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 due solely to an elevated gene-profiling score. Since 22 episodes of asymptomatic rejection were detected in the biopsy group, it is likely that the AlloMap test is not 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 study 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 may not be generalizable to the population of patients 6-12 months post-transplant.

In a follow-up analysis of data from the IMAGE RCT, Deng et al evaluated whether variability in gene expression profiling results were predictive of clinical outcomes. (11) For this analysis, the authors included a subset of 369 patients who had at least 2 AlloMap tests done before an event or the study end, and at least 1 endomyocardial biopsy and 1 echocardiogram. Patients were included from both
arms of the IMAGE RCT. AlloMap test results were expressed in 3 ways, as an ordinal score from 0 to 39, a threshold score of 1 or 0 depending on whether the score was 34 or more or not, and as a variability score, the standard deviation of all of the ordinal scores within a patient. The AlloMap results were entered into a multivariable regression model to predict the composite end point, defined as a patient's first occurrence of: rejection with hemodynamic compromise, graft dysfunction due to other causes, death, or retransplantation. AlloMap ordinal score and AlloMap threshold score were not predictive of the composite outcome. AlloMap score variability was significantly associated with the composite outcome, with a hazard ratio for a 1-unit increase in variability of 1.76 (95% CI, 1.4 to 2.3). While this study implies that variability in AlloMap score may be a prognostic factor, clinical application of this finding is uncertain.

In 2015 (preprint), Kobashigawa et al published results of a pilot RCT evaluating the use of the AlloMap test in patients who were 55 days to 6 months posttransplant. (12) The study design was similar to that of the IMAGE RCT: 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 first occurrence of 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 (ISHLT ≥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 presented with hemodynamic compromise. In the AlloMap group, 29 of 42 biopsies were performed due to elevated AlloMap scores; 4 were performed due to signs, symptoms, or echocardiographic manifestations of graft dysfunction; 5 were performed as part of follow-up assessment for treatment for rejection; and 4 were performed outside the study protocol. In the biopsy group, 253 biopsies were performed, 4 of which were performed based on clinical need.

**Section Summary**

The most direct evidence on the clinical utility of the AlloMap test comes from 1 large RCT comparing an AlloMap-directed strategy with an endomyocardial biopsy-directed strategy for detecting rejection, which found that the AlloMap-directed strategy was noninferior. However, given the high proportion of rejection episodes in the AlloMap-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 the AlloMap test. In addition, a minority of included subjects were in the first year post-transplant. Results from a pilot RCT suggests that AlloMap may have a role in evaluating for heart transplant rejection beginning at 55 days post-transplant, but the study was insufficiently powered to allow firm conclusions about the noninferiority of early AlloMap use.
Ongoing and Unpublished Clinical Trials

Some currently unpublished trials that might influence this policy are listed in Table 1.

<table>
<thead>
<tr>
<th>NCT No.</th>
<th>Trial Name</th>
<th>Planned Enrollment</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ongoing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCT01833195</td>
<td>Outcomes AlloMap Registry: the Long-term Management and Outcomes of Heart Transplant Recipients With AlloMap Testing (OAR)</td>
<td>2000</td>
<td>December 2018</td>
</tr>
<tr>
<td>NCT02178943</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>December 2018</td>
</tr>
</tbody>
</table>

NCT: national clinical trial

*Denotes industry sponsored or cosponsored trial

Practice Guidelines and Position Statements

In 2010, the International Society of Heart and Lung Transplantation issued guidelines for the care of heart transplant recipients. (13) The guidelines included the following recommendations:

- The standard of care for adult heart transplant recipients is to perform periodic endomyocardial biopsy during the first 6-12 months after transplant for rejection surveillance.
- After the first year post-transplant, EMB surveillance every 4-6 months is recommended for patients at higher risk of late acute rejection.
- Gene Expression Profiling using the AlloMap test can be used to rule out acute heart rejection (grade 2 or greater) in appropriate low-risk patients between 6 months and 5 years post-transplant.

U.S. Preventive Services Task Force Recommendations

Not applicable.

Summary

There is insufficient evidence on the diagnostic accuracy of the Heartsbreath test, especially for grades 3 and 4 rejection, and no published studies have evaluated the clinical utility of this test. Therefore, use of the Heartsbreath test to assist in the detection of heart transplant rejection is considered **not medically necessary**.

There is some evidence on the diagnostic accuracy of the AlloMap test from the CARGO trial and post-CARGO publications. However, the evidence is not sufficiently rigorous to determine the true sensitivity
and specificity of the test with certainty. The clinical utility of the AlloMap test compared with routine endomyocardial biopsies has been evaluated in 2 subsequent randomized controlled trials (RCTs), the IMAGE study in patients more than 6 months post-transplant, and a small pilot RCT in patients starting at 55 days posttransplant. The threshold indicating a positive test that seems to be currently accepted, a score of 34, evolved partway through the data collection period the IMAGE study. In addition, the IMAGE study had several methodologic limitations, e.g., lack of blinding, and it was not able to determine whether AlloMap offers incremental benefit compared with biopsy performed on the basis of clinical exam and echocardiography. In patients at less than 1 year after transplant, the group that is at highest risk of transplant rejection, there are insufficient data on the clinical utility of AlloMap. Results from a pilot RCT suggest that AlloMap may be noninferior to a strategy of routine biopsies for heart-detecting transplant rejection beginning at 55 days post-transplant, but the study was insufficiently powered to allow firm conclusions. Thus, use of the AlloMap test to assist in the detection of heart transplant rejection is considered not medically necessary.

Medicare National Policy

In December 2008, the Centers for Medicare and Medicaid Services (CMS) issued a noncoverage decision for the Heartsbreath Test. (14) CMS has determined that the evidence does not adequately define the technical characteristics of the test nor demonstrate that Heartsbreath testing to predict heart transplant rejection improves health outcomes in Medicare beneficiaries.

For AlloMap, there is no National Coverage Determination (NCD). In the absence of an NCD, coverage decisions are left to the discretion of local Medicare carriers. Palmetto GBA determines coverage and reimbursement for laboratories that perform molecular diagnostic testing and submit claims to Medicare in Medicare Jurisdiction E (California, Nevada, and Hawaii). Palmetto GBA’s decisions apply for all molecular diagnostic tests for Medicare. In 2012, Palmetto conducted a technical assessment and determined that AlloMap meets Medicare’s reasonable and necessary criteria.

References

5. Blue Cross Blue Shield Technology Evaluation Center (TEC). Gene expression profiling as a noninvasive method to monitor for cardiac allograft rejection. TEC Assessments 2011; Volume 26, Tab 8.


<table>
<thead>
<tr>
<th>Policy History</th>
<th>Date</th>
<th>Action</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 2012</td>
<td>New Policy</td>
<td></td>
<td>Policy updated with literature review, References removed and renumbered. Policy statements unchanged.</td>
</tr>
</tbody>
</table>
Keywords

AlloMap™
Breath Test, Heart Transplant Rejection
Heartsbreath
Transplantation

This policy was approved by the FEP® Pharmacy and Medical Policy Committee on June 19, 2015 and is effective July 15, 2015.

Signature on File

Deborah M. Smith, MD, MPH