Noninvasive laboratory tests have been explored as an alternative to biopsy to detect cellular rejection following heart transplantation.

**MEDICAL POLICY CRITERIA**

I. The measurement of volatile organic compounds to assist in the detection of grade 3 heart transplant rejection is considered **investigational**.

II. 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 **investigational**.

**NOTE:** A summary of the supporting rationale for the policy criteria is at the end of the policy.

**POLICY GUIDELINES**

The HeartsBreath™ test measures breathe markers of oxidative stress (Criteria I), and the AlloMap® test provides gene expression profiling of RNA obtained from peripheral blood samples (Criteria II).
BACKGROUND

After heart transplantation, patients are monitored for cellular rejection by endomyocardial biopsies that are typically obtained from the right ventricle. The interval between biopsies varies among clinical centers. A typical schedule is weekly for the first month, once or twice monthly for the following six months, and several times (monthly to quarterly) between six months and one-year post transplant. Surveillance biopsies may also be performed after the first postoperative year; e.g., on a quarterly or semi-annual basis. Due to the low rate of rejection after one year, some centers no longer routinely perform endomyocardial biopsies after a year in patients who are clinically stable.

Endomyocardial biopsy is invasive and carries significant risk of adverse effects. Additionally, while endomyocardial biopsy is considered the gold standard for assessing heart transplant rejection, biopsy may be limited by a high degree of interobserver variability in grading of results and the significant morbidity and even mortality that can occur with the biopsy procedure. Also, the severity of rejection may not always coincide with the grading of the rejection by biopsy, and 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 gold standard.

Therefore, noninvasive methods of detecting cellular rejection have been explored. It is hypothesized 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 non-invasive techniques are commercially available for the detection of heart transplant rejection. The HeartsBreath™ test measures breathe markers of oxidative stress, and the AlloMap® test provides gene expression profiling of RNA obtained from peripheral blood samples.

NONINVASIVE HEART TRANSPLANT REJECTION TESTS

HeartsBreath™ Test

The Heartsbreath™ test (Menssana Research, Inc) measures breathe markers of oxidative stress non-invasively and is based on the understanding that in heart transplant recipients, oxidative stress appears to accompany allograft rejection. This rejection degrades membrane polyunsaturated fatty acids and evolving alkanes and methylalkanes, which are excreted as volatile organic compounds (VOC) in breath. The Heartsbreath™ test analyzes the breath methylated alkane contour (BMAC), which is derived from the abundance of C4 to C20 alkanes and monomethylalkanes.

AlloMap® Test

Another approach, the AlloMap® test (CareDx, formerly Xdx, Inc.), focuses 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 multi-gene test panel, which can then be evaluated using polymerase chain reaction (PCR) techniques. The test applies an algorithm to the results, which produces a single score that considers the contribution of each gene in the panel. The manufacturer website states that a lower score indicates a lower risk of graft rejection; the website does not cite a specific cut-off for a positive test.\[^1\]

**Additional Tests**

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**

Both the Heartsbreath™ and AlloMap® tests have received approval from the US Food and Drug Administration (FDA):

- In 2004, the Heartsbreath™ test received approval from the FDA through a humanitarian device exemption. The Heartsbreath™ test is indicated for use as an aid in the diagnosis of grade 3 (significant) heart transplant rejection in patients who have received heart transplants within the preceding year. The test is intended to be used as an adjunct to, and not as a substitute for, endomyocardial biopsy. It is also limited to patients who have had endomyocardial biopsy within the previous month.

- AlloMap® received 510k clearance from the FDA for use in conjunction with clinical assessment to identify heart transplant recipients with stable allograft function. The test is intended for patients at least 15 years-old who are at least two months post-transplant and who have a low probability of moderate/severe transplant rejection.

**EVIDENCE SUMMARY**

The principal outcomes associated with detection of acute heart transplant rejection or graft dysfunction include hemodynamic compromise, graft dysfunction, and/or death. Outcomes relating to use of laboratory tests (such as Heartsbreath™ or AlloMap®) proposed for adjunctive use in heart transplant rejection are best understood by comparing outcomes of patients receiving endomyocardial biopsy alone to those receiving biopsy with the laboratory test. Data from adequately powered, blinded, randomized controlled trials (RCTs) are required to control for baseline differences between groups and determine whether additional testing provides a significant advantage over the standard of care in the proposed uses of these laboratory tests.

**HEARTSBREATH™ TEST**

A single non-randomized study was published in 2004 on the use of the Heartsbreath™ test. No subsequent studies that evaluate use of the Heartsbreath™ test to assess for graft rejection have been identified.

The FDA approval of the Heartsbreath™ test was based on the results of the National Heart Lung and Blood Institute-sponsored Heart Allograft Rejection: Detection with Breath Alkanes in
Low Levels (HARDBALL) study.\textsuperscript{[4]} The HARDBALL study was a three-year multicenter study of 1,061 breath samples in 539 heart transplantation patients. Prior to scheduled endomyocardial biopsy, patient breath was analyzed by gas chromatography and mass spectroscopy for VOCs. The amount of C4 to C20 alkanes and monomethylalkanes was used to derive the BMAC. The BMAC results were compared with subsequent biopsy results as interpreted by two readers using the International Society for Heart and Lung Transplantation biopsy grading system as the "gold standard" for rejection.

The authors of the HARDBALL study reported that the abundance of breath markers of oxidative stress were significantly greater in grades 0, 1 or 2 rejection than in healthy normal subjects. However, 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 comprised the BMAC. The authors also reported that in identifying grade 3 rejection, the negative predictive value 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 (5.6\%) in assessing grade 3 rejection than biopsy (specificity 97\%, positive predictive value 45.2\%). Additionally, the breath test was not evaluated in grade 4 rejection.

**ALLOMAP® TEST**

The validation of the clinical use of any genetic test, including the AlloMap® test, focuses on three main principles:

1. The analytic validity of the test, which refers to the technical accuracy of the test in detecting a mutation that is present or in excluding a mutation that is absent;
2. The clinical validity of the test, which refers to the diagnostic performance of the test (sensitivity, specificity, positive and negative predictive values) in detecting clinical disease; and
3. The clinical utility of the test, i.e., how the results of the diagnostic test will be used to change management of the patient and whether these changes in management lead to clinically important improvements in health outcomes.

A 2011 TEC Assessment reviewed the evidence on the use of AlloMap® testing.\textsuperscript{[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 next.

**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 eight U.S. cardiac transplant centers enrolling 650 cardiac transplant recipients.\textsuperscript{[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 7000 genes known to be involved in immune responses were assayed and compared with 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% for 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 one 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 similar to grade 3 scores. The sensitivity and specificity for determining quiescent versus early stages of rejection was not addressed in this study; however, it was addressed in a 2016 study.\[7\]

Crespo-Leiro (2016) published a reanalysis of the CARGO II data to clinically validate the GEP test performance.\[7\] Blood samples for AlloMap® were collected during post-transplant surveillance and were obtained at least 55 days post-transplantation; >30 days after transfusion of blood products; >21 days after administration of ≥20 mg/day of prednisone; and >60 days after treating a prior rejection. Four hundred and ninety-nine patients had 1579 visits with paired endomyocardial biopsy histopathology rejection grades and GEP scores that met inclusion criteria for the study analyses. The reference standard for rejection status was based on histopathology grading of tissue from endomyocardial biopsy. Results indicated that a GEP test score of ≥34 (patients who are >6 months post-transplantation) corresponded to histology-based grade ≥3A (2R) rejection with a PPV of 4.0% at months ≥2–6 post-transplantation, and 4.3% at >6 months post-transplantation. The NPVs were 98.4% at months ≥2–6 post-transplantation and 98.3% at >6 months post-transplantation. In both time windows, the NPVs increased from 98.3 to >99.0% for decreasing threshold values below 34. The corresponding PPVs decreased from 4.3 to 2.1. Post-CARGO clinical observations have also been published.\[8\] The multicenter work group identified a number of factors that can affect AlloMap® scores, including the time post-transplant, corticosteroid dosing, and transplant vasculopathy.\[8,9\] Scores of 34 or higher were considered positive. Analysis of data from a number of centers collected post-CARGO showed that at one year or more post-transplantation, an AlloMap® threshold of 34 had a PPV of 7.8% for scores of 3A/2R or more on biopsy and a NPV 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 five biopsy samples (2.4%) were found to have a grade of 2R or greater. At one year, 28% of the samples 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.

A similar analysis by Fujita (2017) evaluated the longer-term predictive value of AlloMap® in a group of 46 patients from the CARGO II trial who survived at least one year after transplant.\[10\] Mean AlloMap® scores at 6, 9, 12, and 18 months posttransplant were not significantly different from one another, and there was no significant difference in mortality between those with
scores about the median and those below at any time point. The authors also analyzed changes in AlloMap® scores between different time points, and found that only those with an increase in score between six and nine months posttransplant had higher mortality. Changes at all other times were not significantly associated with mortality. The authors concluded that a nine-month score that is less than 1.02-fold of than the six-month score had a NPV of 100%, but that isolated scores at any of the time points were not correlated with survival.

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% to 100%, based on detecting 10 or fewer cases of rejection in each of three studies.[5]

Clinical Utility

Kobashigawa (2015) published results of a pilot RCT evaluating the use of the AlloMap® test in patients who were 55 days to six months posttransplant.[11] The study design was similar to that of the IMAGE RCT described below: 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 two to six months posttransplant and 34 for patients after six 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 (three in the AlloMap® group vs one 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 presented with hemodynamic compromise. In the AlloMap® group, 29 of 42 biopsies were performed due to elevated AlloMap® scores; four were performed due to signs, symptoms, or echocardiographic manifestations of graft dysfunction; five were performed as part of follow-up assessment for treatment for rejection; and four were performed outside the study protocol. In the biopsy group, 253 biopsies were performed, four of which were performed based on clinical need.

In 2010, results of the Invasive Monitoring Attenuation through Gene Expression (IMAGE) study were published.[12,13] This was an industry-sponsored noninferiority RCT that compared outcomes in 602 patients managed with the AlloMap® test (n=297) or routine endomyocardial biopsies (n=305). The study was not blinded. The study included adult patients from 13 centers who underwent cardiac transplantation between one and five 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 six months and one year earlier; 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 LVEF 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 two elevated scores with no evidence of rejection found on two 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 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 gene expression profiling (GEP) with the AlloMap® test compared with 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 two 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 Health Survey (SF-12).

According to Kaplan-Meier analysis, the two 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 six 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 seven 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 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 one (0.3%) adverse event associated with biopsy in the AlloMap® group and four (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 two groups at baseline and did not differ significantly between groups at two 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 two groups, there are at least two 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 one year before study participation; therefore, findings may not be generalizable to the population of patients 6 to 12 months post-transplant.

In a follow-up analysis of data from the IMAGE RCT, Deng (2014) evaluated whether variability in gene expression profiling results were predictive of clinical outcomes.[14] For this analysis, the authors included a subset of 369 patients who had at least two AlloMap® tests done before an event or the study end, and at least one endomyocardial biopsy and one echocardiogram. Patients were included from both arms of the IMAGE RCT. AlloMap® test results were expressed in three 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 one 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.

Section Summary

The most direct evidence on the clinical utility of the AlloMap® test comes from one 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 posttransplant, but the study was insufficiently powered to allow firm conclusions about the noninferiority of early AlloMap® use.
PRACTICE GUIDELINE SUMMARY

In 2010, the International Society of Heart and Lung Transplantation issued consensus-based guidelines for the care of heart transplant recipients.\textsuperscript{[15]} 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 to 12 months after transplant for rejection surveillance.
- After the first year post-transplant, endomyocardial biopsy surveillance every 4 to 6 months is recommended for patients at higher risk of late acute rejection.
- Gene Expression Profiling (AlloMap\textsuperscript{®}) can be used to rule out the presence of acute heart rejection of grade 2R or greater in appropriate low-risk patients, between 6 months and 5 years after heart transplant.

SUMMARY

There is not enough research to show that the Heartsbreath™ test improves health outcomes for patients that have had a heart transplant. Therefore, the measurement of volatile organic compounds to assist in the detection of heart transplant rejection, including use of the Heartsbreath™ test, is considered investigational.

There is not enough research to show that gene expression profiling to predict rejection improves health outcomes for patients who have had a heart transplant. Therefore, the use of gene expression profiling, including the AlloMap\textsuperscript{®} test, to assist in the detection of heart transplant rejection is considered investigational.

REFERENCES

1. AlloMap website. [cited 05/23/2017]; Available from: www.allomap.com


16. BlueCross BlueShield Association Medical Policy Reference Manual "Laboratory Tests for Heart Transplant Rejection." Policy No. 2.01.68

### 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>81479</td>
<td>Unlisted molecular pathology procedure</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>86849</td>
<td>Unlisted immunology procedure</td>
</tr>
<tr>
<td>HCPCS</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>

*Date of Origin: January 2005*