TITLE: Gene Expression Profiling for the Diagnosis of Heart Transplant Rejection

AUTHOR: Jeffrey A. Tice M.D.
Assistant Professor of Medicine
Division of General Internal Medicine
Department of Medicine
University of California San Francisco

PUBLISHER: California Technology Assessment Forum

DATE OF PUBLICATION: October 13, 2010

PLACE OF PUBLICATION: San Francisco, CA
GENE EXPRESSION PROFILING FOR THE DIAGNOSIS OF HEART TRANSPLANT REJECTION

A Technology Assessment

INTRODUCTION

The California Technology Assessment Forum (CTAF) has been asked to update its review of the scientific literature on the safety and efficacy of gene expression profiling for the diagnosis of heart transplant rejection. The topic was last reviewed in October 2006. A randomized trial comparing use of the Food and Drug Administration (FDA) approved test, AlloMap, to endomyocardial biopsies was published in 2010. AlloMap remains the only FDA approved gene-expression test for heart transplant rejection monitoring, but there is another potential test under development in Canada. This update of the CTAF review will focus on AlloMap.

BACKGROUND

Heart transplantation

Outcomes following heart transplant have improved markedly since the procedure was first performed in 1967. Mortality rates in the year following transplant were as high as 44%. More recently, one year mortality rates in the Registry of the International Society for Heart and Lung Transplantation (ISHLT) have decreased from approximately 23% in the 1980’s to approximately 15% and continue to steadily improve. Median survival over the same period has increased from approximately eight years to greater than 11 years. Acute transplant rejection is a common problem resulting in significant morbidity and mortality. It accounts for approximately 7% of deaths in the first 30 days following transplant, 12% of deaths from day 31 through one year and approximately 10% from years one through three. Acute cellular rejection is the primary form of transplant rejection, although antibody-mediated or humoral rejection also contributes to morbidity in transplant recipients. Much of the improvement in long term mortality is thought to be due to refinements in the immunosuppressive medications that prevent transplant rejection.

After heart transplantation, patients are carefully monitored for signs of rejection. The incidence of rejection peaks at about one month after transplant and then rapidly declines. Biopsy evidence of rejection usually is present before other signs and symptoms of myocardial compromise. Cardiac transplant rejection is often asymptomatic. Thus, routine surveillance biopsies have been the mainstay of early detection and treatment of acute rejection. The usual surveillance course includes endomyocardial biopsy of the right ventricle.
weekly for the first month, once or twice monthly for six months, and then on an annual basis. Because late rejection is a rare event, some centers do not perform routine endomyocardial biopsies after one-year post-transplant in clinically stable patients.

In 1990, a consensus conference devised a standard system for evaluating rejection in heart biopsy specimens, the International Society for Heart and Lung Transplantation grading system.\textsuperscript{10} This scale ranges from 0 to 4, with 0 indicating no evidence of rejection and higher scores reflecting greater degrees of lymphocyte infiltration and myocyte necrosis (Table 1). Low grade rejection (ISHLT grade 1A, 1B, or 2) is generally not treated unless there is evidence of a decline in cardiac function. Pulse steroids are usually used to treat higher grades of rejection (ISHLT grade 3A, 3B, or 4) and more aggressive immunosuppressive therapies are reserved for rejection associated with hemodynamic compromise. A revised scale was agreed upon in 2004 and published in 2005,\textsuperscript{7} but it has not been used in most of the published studies evaluating alternatives to endomyocardial biopsy.

\textbf{Table: ISHLT Grading System for Acute Cellular Rejection}

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>None</td>
<td>Absent</td>
</tr>
<tr>
<td>1A</td>
<td>1R</td>
<td>Focal perivascular or interstitial</td>
<td>Absent</td>
</tr>
<tr>
<td>1B</td>
<td>1R</td>
<td>Multifocal or diffuse, sparse</td>
<td>Absent</td>
</tr>
<tr>
<td>2</td>
<td>1R</td>
<td>Single focus, dense</td>
<td>Present</td>
</tr>
<tr>
<td>3A</td>
<td>2R</td>
<td>Multifocal, dense</td>
<td>Present</td>
</tr>
<tr>
<td>3B</td>
<td>3R</td>
<td>Diffuse and dense</td>
<td>Present</td>
</tr>
<tr>
<td>4</td>
<td>3R</td>
<td>Diffuse and extensive; hemorrhage, edema, and vascular injury may be present</td>
<td>Present</td>
</tr>
</tbody>
</table>

Unfortunately, there is a high degree of inter-observer variability in the grading of the biopsy results\textsuperscript{11-13} and transplant rejection can occur in the setting of apparently normal biopsy results.\textsuperscript{14} Furthermore, sampling error from random biopsies can miss areas with the most severe/significant rejection. Thus, endomyocardial biopsy is an invasive and imperfect measure of rejection with risks for significant adverse events. The overall risk of complications is low (<2%), but serious complications such as cardiac perforation with tamponade, vascular injury, arrhythmias, and death can occur.\textsuperscript{15-17} Tricuspid regurgitation, occasionally requiring valve replacement, is a long term complication reported in transplant patients who undergo multiple right ventricular biopsies.\textsuperscript{18} There is active research attempting to identify less invasive and potentially more accurate methods to identify transplant rejection.
Echocardiographic\textsuperscript{19-24}, electrocardiographic\textsuperscript{25-27}, and magnetic resonance imaging (MRI)\textsuperscript{28} measures have been studied as early indicators of rejection, but none have proved sensitive or specific enough in validation studies when compared with endomyocardial biopsy. A number of biomarkers have been investigated in the search for a reliable serologic marker for transplant rejection. These include elevated troponin levels\textsuperscript{29-32}, brain natriuretic peptide (BNP)\textsuperscript{33, 34}, C-reactive protein\textsuperscript{35-38} and soluble interleukin-2 receptor levels\textsuperscript{39}. Recently, gene expression profiling of peripheral blood lymphocytes has generated the most excitement.\textsuperscript{40-42}

Gene expression profiling

Gene expression profiling refers to a number of different technologies that attempt to quantify the relative levels of messenger RNA (mRNA) for large numbers of genes in specific cells or tissues. The goal is to measure differences in the level of translation (expression) of different genes and utilize patterns of differential gene expression in order to characterize different biological states of the tissue. One potential value of this approach is the identification of genes and gene products associated with a disease process that were not previously known. In cancer biology, the technology has been used to try to differentiate between different subtypes of cancers\textsuperscript{43-49}, to identify tumors with good and bad prognoses\textsuperscript{43, 48, 50-56}, and to identify subgroups of tumors with a high likelihood of responding to one therapeutic regimen compared with another.\textsuperscript{57, 58} In transplant medicine, the focus has been on profiling gene expression in circulating white blood cells in order to identify early changes in the immune system that correlate with rejection of the transplanted organ.\textsuperscript{59}

The most common approach to gene expression profiling utilizes arrays of deoxyribonucleic acid (DNA) sequences bound to a surface like a glass slide. Often, tens of thousands of DNA sequences are organized on an individual microarray in an attempt to profile all of the 20-30,000 genes in the human genome. DNA from a test sample (tumor, white blood cells, normal tissue) is bound to fluorescent dye. Then, it is exposed to the surface of the microarray. Any sample DNA that matches DNA on the microarray (complementary sequences) is bound to the microarray at a specific location. The remaining sample is then washed away. The amount of DNA binding at each site is measured by the intensity of the fluorescent signal. Since the identity of the DNA at each site on the microarray is known, the degree of fluorescence can be correlated with the relative amount of RNA in the original sample.

Another approach to the measurement of gene expression is known as real-time, reverse-transcriptase polymerase chain reaction (RT-PCR). This approach uses the reverse transcriptase enzyme to generate complementary DNA (cDNA) from the mRNA in a sample. The cDNA is then amplified using PCR. This approach is most commonly used to quantify the relative amounts of a smaller set of genes as it is more precise and reproducible.
Gene expression experiments usually start with microarrays containing many thousands of genes and compare the profiles of tissue with and without certain characteristics in order to identify a smaller subset of genes that differentiate between the two states (rejection/no rejection; metastases/no metastases). This smaller subset of genes is then validated using new patient samples. Additional candidate genes based on known biological associations may also be included.

These experiments generate tens of thousands of data points, but because of the expense of microarrays and the difficulty obtaining appropriate tissue, the number of patients evaluated is often quite low. Much has been written about the statistical dangers of evaluating thousands of predictor variables in small datasets (multiple hypothesis testing, overfitting). It is essential that any pattern identified by such experiments be independently validated. Unfortunately, excitement about the results from initial experiments has often overwhelmed statistical caution. One recent paper re-evaluated the data from seven gene expression profiles of cancer prognosis and showed that five of them were likely to predict outcome no better than chance.

**AlloMap**

AlloMap is the only commercially available gene expression profile currently available for heart transplant patients. The developers of the diagnostic test hypothesized that peripheral blood mononuclear cells may contain information on the host response to the heart transplant and that this could be detected by measuring gene expression levels in these cells. A complex series of experiments (described below) identified eleven genes that distinguish transplant rejection from quiescence. The assay also measures expression levels of an additional nine housekeeping genes that serve as reference standards. The informative genes are primarily involved in T-cell activation and trafficking, the response to corticosteroids, and hematopoiesis. RT-PCR is used to measure the relative expression of these twenty genes in peripheral blood mononuclear cells. Then a proprietary algorithm is applied to the results to generate a score ranging from 0 to 40. The value of the score is then used to predict the likelihood of rejection. The exact cut-point for low risk of rejection varies depending on the time since the initial transplant. The test is marked as identifying the absence of acute cellular rejection. Thus, the primary diagnostic test statistic of interest is the negative predictive value of the test.

**Technology Assessment (TA)**

**TA Criterion 1:** The technology must have the appropriate regulatory approval.
Until recently, AlloMap and other Gene Expression Profiles were considered laboratory-developed tests, developed by a single clinical laboratory for use only in that laboratory. Laboratory-developed tests are exempt from FDA oversight. AlloMap was available as a laboratory-developed test from the manufacturer's CLIA certified laboratory starting in 2005.

However, on September 7, 2006, the FDA published draft guidance on planned regulation of In Vitro Diagnostic Multivariate Assays (IVDMIA). Complex tests combining data from multiple laboratory tests using a complex algorithm, like those derived from gene expression profiles, will be subject to FDA review in the future. In particular, those with direct implications for medical therapy will be considered Class III devices and will be subject to the Pre-Market Approval (PMA) process. AlloMap was approved by the FDA using the 510(k) process and as an IVDMIA in August 2008 for use in heart transplant recipients 15 years of age and older who are at least 55 days post-transplant to aid in the identification of patients with stable allograft function who have a low probability of moderate/severe acute cellular rejection at the time of testing in conjunction with standard clinical assessment.

**TA Criterion 1 is met.**

**TA Criterion 2:** The scientific evidence must permit conclusions concerning the effectiveness of the technology regarding health outcomes. For diagnostic tests, there is evidence that use of the test would result in improved medical management in a way that will benefit the patient.

The Medline database, EMBASE, Cochrane clinical trials database, Cochrane reviews database and the Database of Abstracts of Reviews of Effects (DARE) were searched using the key words ‘AlloMap,’ ‘xdx,’ ‘heart transplantation,’ and ‘gene expression profiling’. The updated search was performed for the period from January 2006 through September 2010 and identified 136 articles. The bibliographies of systematic reviews and key articles were manually searched for additional references. The abstracts of citations were reviewed for relevance, and all potentially relevant articles were reviewed in full. In order to be included in this systematic review, articles had to compare the results of gene expression profiling using AlloMap with the results of endomyocardial biopsies or describe clinical care guided by AlloMap. Ideally, randomized clinical trials would compare the clinical outcomes of cardiac transplant patients managed using standard monitoring, including endomyocardial biopsy, to those of patients managed using information from gene expression profiling. The search identified six new publications of observational studies and one randomized trial. Of note, three of the studies reported on subsets of patients from the CARGO study and two reported on the same series of patients from the Cleveland Clinic.
Level of evidence: 1, 3

TA Criterion 2 is met.

TA Criterion 3: The technology must improve the net health outcomes.

The primary outcomes of interest should be overall survival and intermediate measures such as ejection fraction, New York Heart Association functional status, Minnesota Living with Heart Failure quality of life questionnaire, and the 6-minute walking distance. An alternative method to screen for transplant rejection that significantly reduces the need to perform endomyocardial biopsies, while preserving survival and quality of life, would be a great advance. In the absence of such studies evaluating these outcomes, the sensitivity and specificity of gene expression profiling for the detection of rejection using the pathology results of endomyocardial biopsy specimens as the gold standard may be helpful.

Cardiac Allograft Rejection Gene Expression Observational (CARGO) Study

The Cardiac Allograft Rejection Gene Expression Observational (CARGO) study was a complex series of studies designed to develop and validate a parsimonious set of mRNA markers that could be measured in peripheral blood mononuclear cells to predict acute heart transplant rejection. The AlloMap test is based on the results of this study. Patients were enrolled prospectively beginning in September 2001. Slides from each patient were sent to a central pathology department for interpretation by a panel of three pathologists blinded to clinical data, though it is not clear if they were blinded to the original interpretation or to the results of the gene expression profiling. Out of 4917 samples from 629 patients, 827 samples from 273 patients were used in the development and validation of the diagnostic test. Patients were required to be at least five years old, at least 21 days post-transplant and post-rejection therapy, and at least 30 days post blood transfusion. The authors included all samples exhibiting rejection (ISHLT grade ≥ 3A by at least two of four pathologists) and a “representative” sample of ISHLT Grade 0 samples frequency matched on age, sex, race, use of induction center, use of cyclosporine or FK506, time since transplant, and clinical center. There were three phases to the study. Phase One used a custom microarray to evaluate the relative expression of 7,370 genes in 285 samples from 98 patients. Based on prior knowledge from literature reviews and correlation of gene expression levels with biopsy results, a set of 252 genes was selected for further evaluation. Phase Two used quantitative PCR to evaluate the 252 candidate genes in an additional 145 samples from 107 patients (36 grade ≥ 3A rejections and 109 grade 0 quiescent samples). Phases One and Two were not completely independent. A total of 39 samples from 31 patients were used in both Phase One and Phase Two. Through a complex series of machine learning algorithms, a linear discriminant classifier
was developed that utilized 11 genes. The final algorithm gives a score between 0 and 40, with higher scores reflecting a higher likelihood of transplant rejection. Phase Three validated the classifier in two additional sets of samples. The primary validation used 63 samples from 63 patients not included in Phases One or Two of the study. The secondary validation included these 63 samples (31 grade ≥ 3A; 32 grade 0) and an additional 184 samples, 30 of which were also used in Phase One of the study. The primary objective of the validation study was to test the hypothesis that the diagnostic score distinguishes between quiescence (ISHLT grade 0) and moderate to severe rejection (ISHLT grade ≥ 3A).

This review will focus on the validation study results. For additional details of the methods used in Phase One and Two, please see the original article and the supplemental methods that are available online. In the primary validation study, the score from the classifier was significantly higher in samples from patients with at least ISHLT Grade 3A rejection on biopsy compared with samples from patients with Grade 0 rejection (values not reported, p=0.0018). The investigators prospectively defined a score ≥20 as the threshold for rejection. Using this threshold, the test had a sensitivity of 84% (95% CI 66%-94%) and a specificity of 38% (95% CI 22%-56%). The larger secondary validation set gave similar results (sensitivity 76%, specificity 41%). In a post-hoc analysis, the investigators noted that the scores increased with time post-transplant, usually in association with decreasing the intensity of steroid therapy. The investigators suggest that optimal thresholds should be a score of 28 in the period from six months to one year and a score of 30 for patients more than one year post-transplant.

In an unplanned analysis, the investigators then evaluated the performance of the test in a representative set of 281 samples from 166 patients at least one year post-transplant (prevalent population study). This sample was chosen to avoid spectrum bias: the distribution of ISHLT grade in these biopsies was representative of patients at least one year post-transplant. Only nine of the 281 samples had ISHLT scores ≥ 3A. Using a threshold of 30, the test had a positive predictive value of 6.8% and a negative predictive value of 99.6%. It is instructive to note that a test that classifies everyone as not having rejection would have a negative predictive value of 96.8% in this test set. Positive and negative predictive values are very sensitive to the prevalence of disease in the population studied. By focusing on a group of patients with a very low probability of rejection, the investigators could be assured of a very high negative predictive value. This says very little about the clinical utility of the test.

The study offers hope that expression profiling of peripheral blood may be useful in some heart transplant patients. However, the results of the primary validation study, using the a priori threshold of 20, were disappointing. The sensitivity of the test for rejection was 84% (95% CI 66%-94%) and the specificity was only 38% (95% CI 22%-56%). Both estimates had wide confidence intervals reflecting the small sample size.
in each group. Furthermore, the validation study suffered from significant spectrum bias. Biopsy specimens with ISHLT Grades 1A, 1B, and 2 were excluded from the initial validation study. These grades represent approximately 40% of the specimens used in the prevalent population study described in the paper. Because the AlloMap test measures white blood cell activation, test results are likely to be altered by infection and by the form of immunosuppressive therapy used by the patient. Further studies will need to evaluate the test characteristics of AlloMap in patients with active infection and in patients treated with different immunosuppressive regimens. Finally, in a post-hoc analysis, the investigators realized that a threshold of 20 was too low and that different thresholds were needed for patients at different time points in their post-transplant course. Additional validation studies are needed using thresholds that are defined prior to the start of the study before we can have confidence in the clinical utility of the test.

Six new observational studies

None of the six new observational studies prospectively evaluated the sensitivity and specificity of AlloMap compared with endometrial biopsy. Three were additional publications using subsets of patients from the CARGO study exploring associations of the AlloMap score with Grade 1B acute cellular rejection or with the prediction of future rejection. Two other studies likely used the same set of patients from the Cleveland Clinic to explore potential associations of the AlloMap score with post-transplant ischemic injury and with coronary allograft vasculopathy. The fifth explored associations between the AlloMap score and routinely collected measures in cardiac transplant patients. These studies suggest potential additional uses for AlloMap score, but do not give us any additional information about the negative predictive value of the test or whether it can safely reduce the use of endomyocardial biopsy in the follow-up monitoring of patients who have received a heart transplant.

The first substudy using data from the CARGO participants included only patients with blood drawn at least 55 days post-transplant and more than 21 days since treatment for rejection: 265 of the 737 patients from the CARGO study centers. They reported the mean AlloMap scores for patients in each of the acute cellular rejection grades as defined by the three pathologist consensus panel. The scores for patients with severe rejection (Grade ≥ 3A) were significantly higher than the scores for Grades 0, 1A, and 2 (32.0 versus 25.3, 23.8, and 26.9 respectively, p<0.01 for each comparison). However, the AlloMap score for Grade 1B, 29.8, was not significantly lower than that for Grades 3A and higher (p=0.25). The authors suggest that some specimens graded as 1B may represent false negatives with diffuse monocyte infiltration and myocyte injury that is not clinically evident with light microscopy. Using the 2004 revision of the grading system, there was a stepwise increase in the AlloMap score: 25.3 for Grade 0; 26.9 for Grade 1R, and 32.0 for Grades ≥ 2R.
The second CARGO substudy was a nested case control study that included patients who were at least 30 days from either transplantation or treatment for rejection and were without signs or symptoms of rejection in the past 30 days. Both cases and controls had Grade 0 or 1A rejection at baseline. Cases were patients who developed severe acute cellular rejection within 12 weeks of their baseline visit. Controls remained free of Grade ≥ 2 rejection for at least 12 weeks and were frequency matched on “demographic and clinical factors” that were not further defined in the paper. There were 39 cases and 65 controls. There were highly significant differences in the baseline ISHLT biopsy grade at baseline (Grade 1A in 69% of cases versus 34% of controls, p=0.006). There was also a significant difference in the AlloMap score at baseline (27.4 versus 23.9, p=0.01) that remained significant after adjusting for race and biopsy grade. In the subset of patients who were less than 180 days post-transplant, the difference was more marked (28.4 versus 22.4, p = 0.0004) and none of the patients who went on to rejection had AlloMap scores less than twenty. The authors suggest that the AlloMap score may be useful in predicting future episodes of acute cellular rejection and in tailoring patients’ immunosuppressive medications. However, the results need validation in larger prospective studies.

The same research group expanded on these results using 127 patients from the CARGO study. They found that among patients with ISHLT Grade 0 or 1A rejection on endomyocardial biopsy and AlloMap scores ≤ 20, none progressed to severe rejection. Additionally, a gene score ≥ 30 was associated with progression to severe rejection for 58% of patients. They estimated that approximately 44% of heart transplant patients would have gene expression scores ≤ 20 or ≥ 30 during the first six months post-transplant. Again, these results require prospective validation.

The investigators at Cleveland Clinic explored the association between the AlloMap score and both post transplant ischemia and transplant-associated coronary allograft vasculopathy. The AlloMap score was higher in patients with early post-transplant ischemia (31.5, n=19) than in controls without evidence of ischemia (21, n=48, p<0.001). Similarly, the AlloMap score was higher in patients with coronary allograft vasculopathy (32.2, n=20) than in controls without evidence of coronary allograft vasculopathy (26, n=49, p=0.001). Both studies had too few patients to adequately control for potential confounding, but they do suggest possible reasons for false positive AlloMap scores.

Most recently, investigators at Columbia presented an exploratory analysis looking at 35 parameters routinely collected on 76 post-cardiac transplant patients and correlated them with the AlloMap score. Using a one-sided p-value, ten of the parameters were correlated with the AlloMap score. Of these, only the platelet count and the corrected QT interval (QTc) from electrocardiography remained significant in multivariate regression. The study used looked at a large number of associations with a low p-value.
threshold, so many of the associations that were identified may represent Type 1 errors. However, the QTc has been associated with acute cellular rejection in prior studies, so this is likely to be a true correlation. It is unclear what impact these findings will have on clinical practice.

The Invasive Monitoring Attenuation Through Gene Expression (IMAGE) Randomized Trial

The Invasive Monitoring Attenuation Through Gene Expression (IMAGE) study was a non-blinded randomized trial at 13 U.S. cardiac transplant centers designed to evaluate whether monitoring heart transplant patients with the AlloMap gene expression score was not inferior to monitoring based on endomyocardial biopsy. Patients were followed for two years after randomization. The threshold for non-inferiority was the upper bound of the one-sided 95% confidence interval for the hazard ratio for the AlloMap group would be less than 2.054. The initial inclusion criteria were patients ages 18 years and older with an ejection fraction of at least 45% who were between 12 and 60 months post transplant and who were without evidence of severe cardiac allograft vasculopathy or antibody mediated rejection. Patients were excluded if they had current signs or symptoms of cardiac dysfunction, therapy for ISHLT Grade 3A or higher rejection in the past two months, recent changes in immunosuppressive medications, blood transfusion in the past four weeks, or current use of corticosteroids at a dose equivalent to ≥ 20 mg / day of prednisone. The study was active from January 2005 through October 2009. Because of slow enrollment, in November 2007, the entry criteria were expanded to include patients from six to twelve months post-transplant.

The primary endpoint was a composite of a 25% or greater reduction in ejection fraction on echocardiography, rejection with hemodynamic compromise, retransplantation, or death. Patients randomized to the endomyocardial biopsy group received biopsies according to each center’s usual protocol and had gene expression testing done at each study visit, but the treating physicians were blinded to the gene expression profiling results. Patients randomized to the AlloMap group had their gene expression profile performed at each rejection surveillance visit. Patients with clinical or echocardiographic evidence of allograft dysfunction also underwent endomyocardial biopsy and other testing according to each center’s protocol. If the AlloMap score was greater than or equal to 30, then the patient underwent biopsy. If the biopsy result was ≥ 2R, the patient was treated for acute cellular rejection. Otherwise, the patient continued to receive routine surveillance with gene expression profiling. In November 2005, the protocol was amended to require biopsy for a score of 34 or greater. The original threshold score were based on data from the CARGO study described above. In comments received from the manufacturer, they note that the new test had not yet been used for any prospective patient management decisions. Early data from the prospective IMAGE trial found that the suggested thresholds from CARGO were too low to achieve the desired high
negative predictive value. This led to the protocol amendment and highlights the fact that the most appropriate way to use the test for clinical management of patients continues to evolve as better quality studies are performed.

The study randomized 602 subjects and the subjects had a median follow-up of 19 months. Their average age was 54 years and 82% were men. The two groups were well matched except for a lower proportion of black patients in the AlloMap group (8% versus 15%, p=0.01). The two-year rate of the primary outcome was similar in both groups (14.5% AlloMap versus 15.3% biopsy, HR 1.04, 95% CI 0.67 to 1.68). Thus, the study's prespecified criterion for non-inferiority was met. However, there was a higher rate of the primary outcome in black patients (18.3% versus 10.2%, p=0.07). The hazard ratio in the primary analysis was 1.13, 95% CI 0.70 to 1.84, after adjusting for black race. The two-year overall mortality did not differ between the two groups (6.3% versus 5.5%, p=0.82).

There were fewer endomyocardial biopsies performed in the AlloMap group (409 versus 1249; 0.5 biopsies per year versus 3.0 biopsies per year, p<0.001). Biopsy-related complications occurred in one patient in the AlloMap group (medication error – formalin given rather than lidocaine at the cannulation site: wound required debridement) and four patients in the biopsy group (two tricuspid valve insufficiency, one symptomatic pericardial effusion, one bleeding episode). Patient satisfaction scores in the AlloMap group increased from 6.86 at baseline to 8.15 at one year and 8.74 at two years. Patient satisfaction scores in the biopsy group remained stable from 6.74 at baseline to 6.64 at one year and 6.66 at two years (p for the difference between groups <0.001 at one and two years). There were no differences on the mental-health summary scores of the short-form 12 (SF-12) quality of life measures at any time point. However the AlloMap group had a lower physical health summary score at one year (44.7 versus 47.3, p=0.03). The difference was no longer significant at two years (45.1 versus 46.2, p=0.52).

It is commendable that the investigators undertook a randomized trial to test the utility of a cardiac transplant management strategy based on the AlloMap gene expression profile despite the difficulties given that only 3500 heart transplants are done annually in the world and that the patients lives depend on diagnosing severe rejection early enough to treat it successfully. That said, there are several concerns about the quality of the trial. The baseline imbalance in the distribution of race suggests that allocation concealment was inadequate and that there may have been selection bias during the randomization process. The investigators adjusted for this in their analyses, but the presence of potential selection bias in a randomized trial is concerning. In addition, the study was unblinded. This can impact both outcome identification as well as final endpoint adjudication. The magnitude of bias introduced by lack of blinding tends to be largest for subjective outcomes and may explain part of the difference in the patient satisfaction
scores between the two groups. However, it can impact co-interventions, such as the adjustment of patient’s immunosuppressive medications and the decision to biopsy based on borderline clinical symptoms. Data in the supplementary appendix to the NEJM article suggest that there were baseline differences in immunosuppression between the two groups, but that these differences diminished through the trial. The definition of a “negative” test continued to evolve in this study. When the test was initially developed the threshold was 20, then it was increased to 28 for patients six to twelve months post-transplant and 30 for patients more than a year post-transplant. In this study, the threshold was increased again to 34. This varying threshold across studies makes it difficult to interpret the entire body of literature on the test and highlight the continued evolution in thinking about the clinical utility of the test. Finally, the confidence intervals surrounding the primary estimate were wide: the point estimate from the adjusted analysis was a 13% increase in the primary outcome with reasonable estimates falling between a 30% decrease and an 84% increase in the risk of death, retransplantation, or significant cardiovascular compromise. If the true value is a 50% increased risk for the above events, many patients and their treating physicians would not choose monitoring with gene expression profiling.

The results should not be generalized to patients six to twelve months from transplantation. These patients represent less than 15% of the sample in the study and none of them had two years follow-up as the protocol modification including them in the trial occurred less than two years before the close of the trial.

**TA Criterion 3 is met for patients at least one year post-transplantation.**

**TA Criterion 4:** The technology must be as beneficial as any established alternatives.

The established alternative to gene expression profiling is endomyocardial biopsy. Both are used in the context of the patient’s clinical history, physical exam findings, and regular echocardiographic evaluation. The right heart catheterization performed as part of the biopsy procedure provide important hemodynamic information, such as cardiac filling pressures and cardiac output, which are also used to guide patient management. These data would not be available to clinicians making therapeutic decisions based on the AlloMap score. The high negative predictive value of the CARGO prevalent population sub-study suggested that patients at least a year post-transplant may be adequately assessed by gene expression profiling, but the majority of biopsies have been performed in transplant patients by one year. In addition, some authors have questioned the need for routine surveillance biopsies in this population and some sites do not routinely perform biopsies in patients who are clinically stable.
The IMAGE trial described above demonstrated non-inferiority of a management strategy guided by gene expression profiling to one guided by endomyocardial biopsy. There were significantly fewer biopsies performed and patient’s reported higher satisfaction with the gene expression profiling strategy. However, there were many concerns about the trial as detailed under TA criterion 3. The results should not be applied to patients less than one year post-transplantation. Furthermore, patients and treating clinicians need to be informed about the uncertainties surrounding the relative benefits and harms associated with a monitoring strategy that incorporates gene expression profiling. Finally, clinically stable patients may not require routine gene expression profiling or endomyocardial biopsy beyond one year in addition to the usual clinical monitoring for signs and symptoms of rejection as well as routine echocardiography. Given those caveats, the evidence supports gene expression profiling being as beneficial as the current standard, endomyocardial biopsy, when used as part of the routine post-transplant monitoring in stable patients at least one year post-transplant.

**TA Criterion 4 is met.**

**TA Criterion 5:** The improvement must be attainable outside the investigational setting.

The large CARGO trial reported data from eight transplant centers and the IMAGE trial from 13 centers, suggesting that the sample collection process can be handled reliably at centers with the expertise to perform heart transplants. All of the tests are required to be performed at one central laboratory. No data was presented on the reliability of the final test itself. Measures of the consistency of the test results when performed on the same sample would be useful.

**TA Criterion 5 is met.**

**CONCLUSION**

Heart transplant patients face significant risks for life-threatening rejection, particularly during the first year after transplant. Endomyocardial biopsies are performed according to a strict schedule in order to diagnose significant rejection as early as possible. The search for a less invasive marker of rejection has been a research priority for decades. Gene expression profiling offers the potential for a non-invasive test that may replace endomyocardial biopsy as the gold standard for transplant rejection in stable patients. The AlloMap gene expression profile has a high negative predictive value, but a low positive predictive value. Thus it may be useful to avoid biopsy in stable patients, but the high false positive rate precludes its use to definitively diagnose acute cellular rejection. Endomyocardial biopsies will still need to be performed in all patients with elevated AlloMap scores and all patients with clinical signs of rejection. The IMAGE trial provides data
supporting the non-inferiority of a monitoring strategy for heart transplant patients incorporating the AlloMap gene expression profile in lieu of routine endomyocardial biopsy. However, the data only support such strategies in patients more than a year post-transplant. More data are needed to confirm the tests utility earlier in the post-transplant period when the majority of endomyocardial biopsies are performed.

RECOMMENDATION

It is recommended that the use of gene expression profiling meets Technology Assessment Criterion 1 through 5 for safety, effectiveness and improvement in health outcomes when used to manage heart transplant patients at least one year post-transplant.

October 18, 2006
This is the second CTAF review of this technology

The CTAF panel voted to accept the recommendation as presented.
RECOMMENDATIONS OF OTHERS

Blue Cross Blue Shield Association (BCBSA)
The BCBSA Technology Evaluation Center (TEC) has not conducted a review of this technology.

Centers for Medicare and Medicaid Services (CMS)
At this time CMS does not have a published National Coverage Decision regarding the use of this technology. However, the lab providing the services does have a provider number.

American College of Cardiology California Chapter (CA ACC)
The CA ACC was asked to provide an opinion regarding this technology and to have a representative attend the meeting.

International Society of Heart and Lung Transplant (ISHLT)
The ISHLT 2010 Guidelines for the Care of Heart Transplant Recipients, Task Force 2: Immunosuppression and Rejection are available at:
https://www.ishlt.org/ContentDocuments/ISHLT/GL_Task_Force_2_080510.pdf
# Abbreviations Used in This Assessment:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTAF</td>
<td>California Technology Assessment Forum</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>ISHLT</td>
<td>International Society of Heart and Lung Transplant</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>BNP:</td>
<td>Brain Natriuretic Peptide</td>
</tr>
<tr>
<td>mRNA:</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>RT-PCR:</td>
<td>Reverse-transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>cDNA:</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>IVMIA:</td>
<td>In-Vitro Diagnostic Multivariate Assays</td>
</tr>
<tr>
<td>PMA:</td>
<td>Pre-Market Approval</td>
</tr>
<tr>
<td>DARE:</td>
<td>Database of Abstracts of Reviews of Effects</td>
</tr>
<tr>
<td>SAM:</td>
<td>Statistical Analysis of Microarrays</td>
</tr>
<tr>
<td>CARGO:</td>
<td>Cardiac Allograft Rejection Gene Expression Observational</td>
</tr>
<tr>
<td>IMAGE:</td>
<td>Invasive Monitoring Attenuation Through Gene Expression</td>
</tr>
<tr>
<td>PCR:</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>QTc:</td>
<td>Corrected QT interval</td>
</tr>
<tr>
<td>SF-12:</td>
<td>Short form 12</td>
</tr>
</tbody>
</table>


