PROTOCOL

Surgical Treatment for Ischemic Heart (STICH) Failure Trial

U01 HL69015

Authors: Eric J. Velazquez, Kerry L. Lee, Daniel B. Mark, Christopher M. O'Connor, Robert M. Califf, Robert H. Jones

Document Type: Clinical Study Protocol

Document Status: Version 04

Date: June 16, 2006

Number of Pages: 41

List of Figures

List of Tables

2.2 Secondary Objectives

1. To assess the clinical utility of the measurements by CMR of LV shape, size and function for predicting the benefit of a specific treatment strategy.
2. To assess the clinical utility of measurements by nuclear cardiology and/or echocardiography testing of myocardial ischemia, viability, and LV size and function for predicting preferential benefit associated with a specific treatment strategy.
3. To determine if the use of baseline ECHO derived assessments of LV systolic and diastolic function, hemodynamics, and valvular regurgitation identify subgroups that derive either substantial benefit or harm from surgical intervention and how CABG compared to MED and CABG + SVR compared to CABG differentially effects these parameters over time.
4. To test the hypothesis that levels of neurohormonal and pro-inflammatory cytokine activation will be useful in identifying that group of patients who will most likely benefit from either CABG and/or CABG + SVR.
5. To compare health-related quality of life, total medical costs, and incremental cost effectiveness between CABG and MED and between CABG with or without SVR.

2.3 Other Key Parameters

Mortality: 30-day all-cause; cardiovascular; all-cause death or LVAD or heart transplantation.
Morbidity: All-cause hospitalization; days of all-cause hospitalization; hospitalization for HF; days of hospitalization for HF; emergency department visits for HF.
Cardiac Procedures: PCI; any cardiac surgery; LVAD; heart transplantation; ICD.
Cost: total medical costs; cost-effectiveness.
Functional Capacity: quality of life; six-minute walk distance (absolute and change); treadmill duration (absolute and change).
Physiologic: LVEF (absolute and change); ESVI (absolute and change); neurohormonal levels (absolute and change); cytokine levels (absolute and change).

3. Background

3.1 Public Health Burden of Heart Failure

American Heart Association statistics suggest that HF affects 4.7 million patients in the United States and is responsible for approximately one million hospitalizations and 300,000 deaths annually.[1] The total annual costs associated with this disorder have been estimated to exceed $22 billion.[2] CAD is the cause of HF in the majority of patients, and HF is the only mode of CAD presentation associated with increasing incidence and mortality.[3] Ischemic HF may develop after a known myocardial infarction or through a process of chronic ischemia that leads to cardiac dilatation, cellular hypertrophy, apoptosis, and extracellular matrix remodeling.[4] The increasing prevalence of HF largely derives from enhanced survival from improved therapies for other manifestations of CAD, especially acute myocardial infarction. However, survival is usually accompanied by some degree of myocardial injury that is the basis for development of HF.

3.2 Development of Surgical Ventricular Reconstruction Operation

Reimer and Jennings showed that experimental coronary occlusion results in a wavefront of cell death moving from the subendocardial zone across the wall to involve progressively more of the transmural thickness of the ischemic zone.[5] Early reperfusion salvages subepicardial but not subendocardial myocardium. The decline in prevalence of thin-walled LV aneurysm that followed introduction of aggressive initial management of acute coronary syndromes suggests that the ischemic process also can be interrupted in patients before it reaches the transmural stage. With later healing of infarcted myocardium, scarring is maximal in the subendocardial region and is interlaced with normal myocardium in diminishing amounts toward the epicardial surface. Ventricular wall or chamber imaging commonly shows an akinetic zone that gradually blends with myocardium with increasingly normal function in contrast to the dyskinetic region with a discrete neck typical of a left ventricular aneurysm.

LV ventricular aneurysmectomy appears to reverse HF by lowering LV wall stress, but these linear amputations of dyskinetic scar commonly deform the LV cavity into a box-like shape. Intra-cavitary reconstruction techniques have been developed for repairing defects left by aneurysm resection that reduced LV cavity size.[6] This surgical ventricular reconstruction (SVR) strategy has more recently been applied not only to patients with dyskinetic scar but also to those with only akinetic myocardial segments.[7] At the time of cardiac operation, the epicardium of these akinetic zones may appear normal and palpable thinning is often minimal in the arrested, decompressed heart. This appearance derives from preservation of a rim of normal myocardium covering the myocardial fibrosis and contrasts with the leather-like appearance and thinness typical of a LV aneurysm.

Unlike the LV aneurysmectomy that removed myocardial scar or the Batista operation that reduced LV size by indiscriminate removal of portions of the LV wall, the SVR operation mechanically decreases the circumference of the zone of endocardial scar through an incision in normal epicardium. The surgical repair uses the intrinsic scar or an extrinsic patch to absorb excess linear wall tension from the adjacent myocardium. Decrease in wall stress appears to reduce the tendency for continued gradual expansion of the akinetic zone. The endocardial repair acutely decreases LV size, thereby acutely enhancing function in myocardial regions remote from the repair.[8] A recent registry report of 439 patients in 11 centers found a 5.1% operative mortality and 88% 18-month survival for patients who underwent CABG and SVR suggesting that this surgical strategy is safe and at least equivalent to CABG alone.[9] This operation is now sufficiently mature to justify this proposed randomized trial to evaluate whether appropriately performed SVR truly adds value to CABG in HF patients with regional dysfunction.[10]

3.3 Need for a Randomized Trial of Medical and Surgical Therapies

Along the broad spectrum of severity of ischemic HF, specific clinical information, such as severe angina or left main coronary artery stenosis, may clearly indicate the need for surgical therapy for some patients. However, a large number of patients fall into a gray zone without clear evidence for benefit from either medical or surgical therapy. For these patients, evidence supporting choice between therapies was never strong and has only been confused by recent studies showing improved outcomes with both therapies. Patients for whom equipoise of anticipated benefit now exists between modern medical and surgical therapy represent the broad population who are appropriate candidates for a randomized trial to provide the context for assessing the value of three therapeutic strategies: 1) MED alone; 2) MED + CABG; and 3) MED + CABG + SVR.

No randomized trial has ever directly compared long-term benefits of surgical and medical treatment of patients with ischemic HF. In the 1969-1972 developmental era of CABG, high surgical mortality rates were observed in patients with HF. Therefore, initial randomized trials comparing CABG to medical treatment conducted from 1972 to 1978 excluded most of these patients from participation. In a subgroup of 160 CASS Trial patients with LVEF <.50, ten-year survival was 61% in the 82 medically-treated patients and 79% in the 78 patients who underwent CABG (p = .01).[11] This survival advantage of CABG was not related to the presence or severity of HF or angina symptoms. A meta-analysis by Yusuf[12] combined individual patient data from the CASS Trial with those enrolled in the six other early randomized trials. Only 191 (7.2%) of the 2,649 total patients had an EF <.40, and only 106 (4.0%) of these patients who were primarily symptomatic with angina also had HF symptoms. CABG improved survival among all patients with proximal LAD stenosis, three-vessel, or left main coronary artery disease regardless of LV function. In these patients with a survival benefit from CABG, a low EF increased the absolute benefit but did not change the relative benefit of CABG. A literature search of 326 published reports on results of CABG in patients with HF or LV dysfunction identified three well-designed cohort studies.[13] Mortality benefit of CABG over medical therapy was 10 and 20 lives per 100 patients at three years in two of the three studies and 29 lives per 100 patients at five years in the third study.

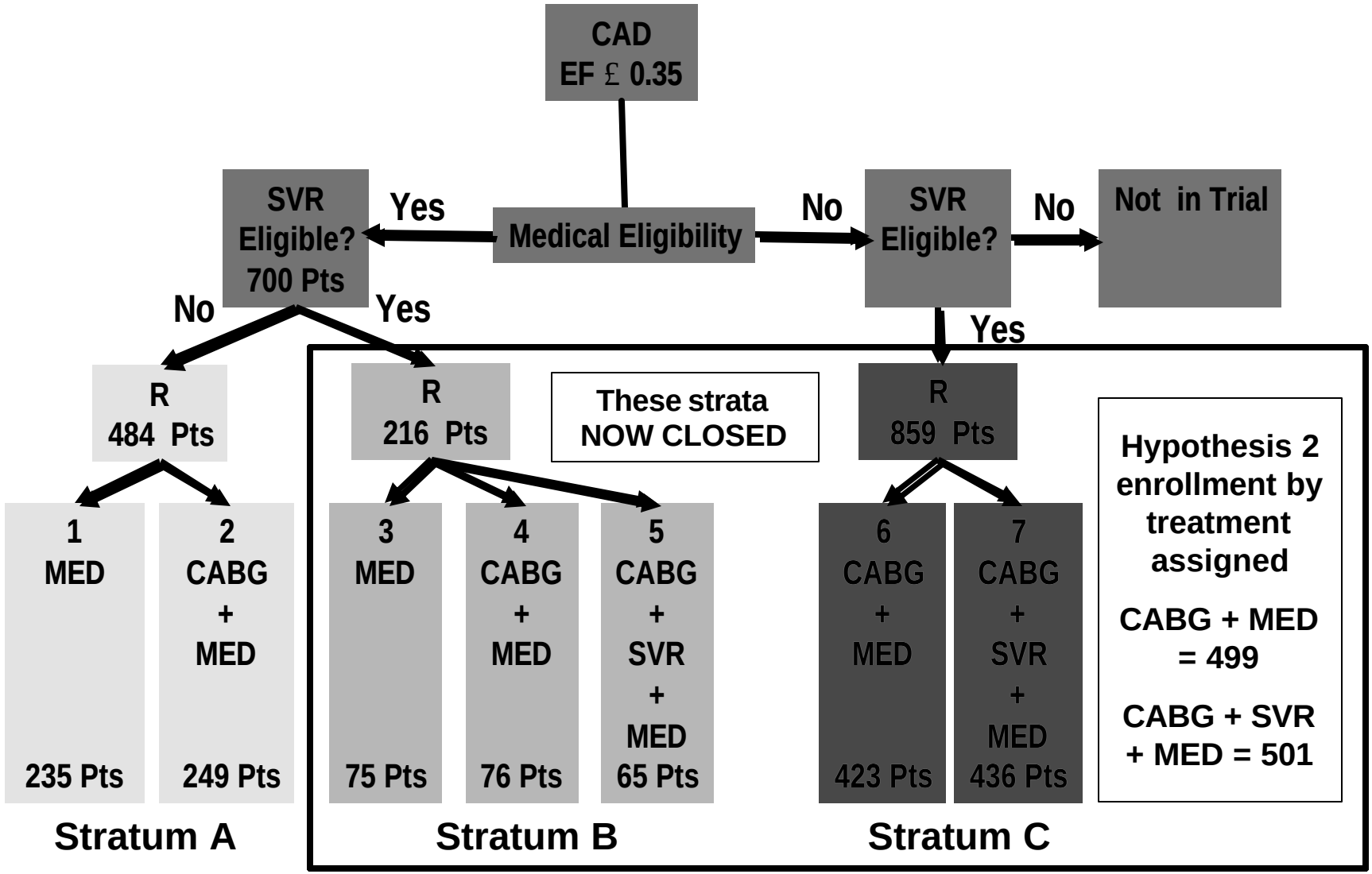
Despite this apparent benefit, data from the Duke Databank for Cardiovascular Diseases (DDCD) show that more than 71% of patients with characteristics that would qualify for inclusion in the STICH Trial receive medical therapy, 22% receive CABG, and 7% receive PCI. Moreover, this distribution of treatments has remained constant over the past decade at an institution with an aggressive use of myocardial revascularization procedures. Perhaps CABG is not more widely used in patients with coronary angiograms that suggest survival benefit because of concern for postoperative complications, such as stroke, that detract from its survival advantage. In addition, the general medical community often over estimates surgical risks and medically manages the high-risk patients without thoroughly characterizing the presence and extent of CAD.

The therapy called medical treatment in all randomized trials and most observational comparisons with CABG really only reflected the natural history of CAD since it rarely included drugs now known to be life prolonging, such as ACE inhibitors, beta blockers, lipid lowering, and anti-platelet agents. Refinement of operative and postoperative surgical management also has steadily improved CABG results over the past two decades. The paucity of modern data available to clinicians who must daily make high-risk management decisions in patients with HF emphasizes the need for a properly-designed randomized comparison of these therapies commonly used in clinical practice.

3.4 Evaluation of Noninvasive Studies

Increasingly over the past three decades, information describing cardiac structure, perfusion, hemodynamics, and metabolism obtained from noninvasive cardiac imaging studies has been used to guide management decisions for patients with HF. Although this anatomic and physiologic information often adds value to clinical care, an accepted strategy has not evolved that tailors the testing sequence to specific presenting features of individual patients to efficiently identify the treatment strategy most likely to improve outcomes. Consensus on proper use of cardiac imaging studies has been hindered by absence of clinical studies that objectively evaluate the independent treatment-related prognostic information of these tests obtained using standardized methods on patients for whom the test has minimal influence on choice of therapy.

FIGURE 1. Patient Stratum and Treatment Assignment at Time of Completion of Hypothesis 2 Enrollment on January 24, 2006



4.3 Study Subject Enrollment

4.3.1 Screening

Screening of patients referred must address clinical eligibility criteria and assess coronary anatomy and left ventricular function. The best available and most recently performed LV assessment by either contrast, gated SPECT, echocardiogram, or CMR ventriculogram read at the clinical sites will identify patients meeting LVEF entry criteria and be used to evaluate SVR eligibility. Baseline STICH Trial protocol tests that have been performed for clinical reasons in potential STICH patients in compliance with the STICH core laboratory protocols may be used for defining trial eligibility or meeting baseline testing requirements, thereby decreasing inconvenience to the patient and unnecessary cost to the STICH Trial from duplicate testing. After the screening process identifies a patient who meets STICH Trial entry criteria, an initial evaluation form is used to document information needed to confirm eligibility of the patient and to define the appropriate stratum for randomization.

4.3.2 Randomization

After eligible patients give informed consent, site personnel will initiate the randomization process by calling an always available toll-free telephone number to connect with the IVRS at the Coordinating Center that has a redundant voice back-up system. Once eligibility is established, a site-specific unique identifier will be assigned to the patient. A permuted block randomization, stratified by clinical site and by the three Strata A, B, or C shown in Figure 1, will be used with a block size not known by the investigators. Use of this randomization process insures proper trial enrollment and promptly provides the Coordinating Center with the basic patient enrollment information needed to monitor enrollment performance. Study design specifies enrollment of at least 2,000 randomized patients into the STICH Trial.

4.3.3 Baseline Evaluation of Randomized Patients Prior to Assigned Treatment

All patients eligible and consenting to be randomized will have had baseline clinical evaluation information entered on the initial evaluation form. Table 2 summarizes baseline evaluation studies for each stratum. Where logistically feasible, a blood sample for neurohormone, cytokine, and genetic studies is strongly encouraged. An echocardiogram will be sent to the ECHO Core Laboratory on all patients. A myocardial viability assessment by either or both radionuclide or echocardiographic techniques is strongly encouraged. A baseline quality of life assessment and six-minute walk test will be performed. A modified Bruce or bicycle stress test may be obtained on patients in Strata A and B who are able to exercise. This may be performed in conjunction with a stress sestamibi perfusion study. When possible, a baseline CMR or gated SPECT ventriculogram will be obtained in all patients. In situations when obtaining one of these two studies would preclude patient entry into the trial, the echocardiogram may serve as the baseline assessment of LV function.

Table 2. Baseline Evaluation Studies

Patient Characteristics	Stratum A	Stratum B	Stratum C
All Consenting Patients	• ECHO to Core • QOL to Core • EuroQOL to Core • CMR, Gated SPECT, or ECHO for LV function	• ECHO to Core • QOL to Core • EuroQOL to Core • CMR, Gated SPECT, or ECHO for LV function	• ECHO to Core • QOL to Core • EuroQOL to Core • CMR, Gated SPECT, or ECHO for LV function
When Feasible	• NCG Blood to Core • Myocardial Viability to Core	• NCG Blood to Core • Myocardial Viability to Core	• NCG Blood to Core
Patients Able	• 6-minute Walk • Exercise Stress	• 6-minute Walk • Exercise Stress	• 6-minute Walk • Exercise Stress

4.4 Initiation and Completion of Initial Treatment

Intensive medical treatment (MED) will be initiated after completion of baseline evaluation in all randomized patients. Patients randomized to CABG, with or without SVR, will undergo operation as soon as possible after completion of protocol-mandated baseline studies, but within 14 days in all patients.

The standard of best practice for each of the three possible treatments will be reviewed in the context of relevant published literature at regular intervals throughout the study by the Medical and Surgical Therapy Committees, and the original treatment protocol may be modified by the Data and Safety Monitoring Board (DSMB) if necessary. Definitions of the three therapies to be used at the start of the trial appear justified by current evidence. In hospitalized patients, initial treatment will be considered to be completed at hospital discharge or at 30 days if hospitalization continues. In patients randomized to medical therapy and treated as outpatients, initial treatment will be considered to be completed when the medical regimen is considered stable or at 30 days. The Initial Treatment Form (Medical or Surgical) should reflect all events between initiation and completion of the initial therapy period.

4.4.1 Treatment Groups

The three treatment groups are:

1. MED: MED for HF is now defined to include angiotensin-converting enzyme inhibitors (ACE-I), or angiotensin receptor blockers in patients who do not tolerate ACE inhibitors, beta blockers, spironolactone, and an antiplatelet agent such as aspirin or clopidogrel adjusted to optimal doses within 30 days of randomization unless contraindications are documented. HMG-CoA reductase inhibitors, diuretics, and digitalis use will be individualized to patient-specific indications. Physicians will be encouraged to manage concomitant conditions appropriately (diabetes, hypertension, hyperlipidemia, angina, atrial arrhythmias) and will be asked to avoid the use of non-steroidal anti-inflammatory drugs as well as most anti-arrhythmic agents and most calcium-channel blocking drugs. Electrophysiologic testing and the implantation of cardioverter-defibrillators and biventricular pacing should be used in compliance with standard guidelines. MED therapy will be monitored closely and modified when indicated at a frequency no less than the required follow-up clinic visits.

2. CABG: CABG will be performed using at least one internal mammary graft in all patients unless this conduit is unavailable or has inadequate flow. Patients with secondary mitral regurgitation identified by echocardiography who are judged to need mitral valve repair may undergo this procedure combined with CABG with or without SVR. Therefore, use of CABG or CABG + SVR throughout this protocol implies that mitral repair for regurgitation secondary to ischemic HF etiology may be included in the operative strategy at the discretion of the operating surgeon.

3. CABG + SVR: The two criteria used by the Surgical Therapy Committee to define the acceptable range of specific operative maneuvers essential to be considered an acceptable technical SVR operation for the STICH Trial will be any ventricular reconstruction method that consistently results in: 1) a low operative mortality; and 2) an average EF increase of ≥10% and average LVESVI decrease of ≥30% as assessed on the four-month postoperative CMR measurement. SVR may be performed with or without cardioplegic arrest. In general, the SVR is begun after completion of the CABG and in a beating heart. The region of dysfunction identified by preoperative CMR will be incised regardless of the epicardial appearance. The border separating the dysfunctional from contracting myocardium identified by palpation will guide placement of sutures and a patch, if necessary, to pursestring the endocardial scar and thereby decrease LV size without distorting LV shape. For patients undergoing surgery, intensive medical treatment will resume as soon as permitted by post-procedure recovery.

4.4.2 Interruption or Discontinuation of Treatment

A patient is considered randomized when the patient identification number has been assigned by the IVRS. Every effort must be made to ensure that patients remain in the study and on medical therapy for the duration of the study. An interruption of medical therapy may occasionally be required. If a temporary interruption occurs, the Coordinating Center Clinical Hot Line is available through the IVRS for consultation or notification about any study-related issues. Every attempt to reinitiate optimal medical therapy should be made throughout the duration of the study. The reinitiation of medical therapy is not subject to a time limit. Patients with a temporary or permanent discontinuation of any or all medications integral to optimal medical therapy should continue the visit schedule and undergo all protocol specified studies. Each patient entering the study must be followed until study completion whether or not the patient receives the randomized treatment or medical therapy is temporarily interrupted or permanently discontinued. A patient will be considered lost to follow-up only after exhausting all means of contact. The status of the patient at the last visit or contact will be used for the final analysis. The vital status of patients who withdraw consent or are lost to follow-up will be followed by a national death registry when possible.

4.5 Study Subject Follow-up

At each clinic visit at four-month intervals for the first year after study entry and at six-month intervals thereafter, a brief history and physical examination will be performed and follow-up forms will be completed. Follow-up data to be collected at each of these contacts will include an assessment of HF and angina symptoms and interval medical events including hospitalization, major cardiac procedures, and tests. Where logistically feasible, a follow-up blood sample is strongly encouraged at the four-month follow-up visit. Patients completing a treadmill test at baseline will have this test repeated at two years. Detailed quality of life assessment will be performed at the four-month visit, one year, and annually thereafter during follow-up. Table 3 summarizes the schedule of follow-up studies in the STICH Trial. Telephone contact will be maintained with the patient unwilling or unable to return to the clinical site and their local care providers to assure that intensive medical therapy is continued.

Table 3. Follow-up Evaluation Studies

Interval	Patient Characteristics	Stratum A	Stratum B	Stratum C
4 Months	All Patients	ECHO to Core EuroQOL to Core	ECHO to Core EuroQOL to Core V-gram to Core (CMR, SPECT, or Echo)	ECHO to Core EuroQOL to Core V-gram to Core (CMR, SPECT, or Echo)
	When Feasible	NCG Blood to Core	NCG Blood to Core	NCG Blood to Core
	Patients Able	6-minute Walk	6-minute Walk	6-minute Walk
12 Months	All Patients	EuroQOL to Core	EuroQOL to Core	EuroQOL to Core
	Patients Able	6-minute Walk	6-minute Walk	6-minute Walk
24 Months	All Patients	ECHO to Core EuroQOL to Core	ECHO to Core EuroQOL to Core V-gram to Core (CMR, SPECT, or Echo)	ECHO to Core EuroQOL to Core V-gram to Core (CMR, SPECT, or Echo)
	Patients Able	6-minute Walk Exercise Stress	6-minute Walk Exercise Stress	6-minute Walk Exercise Stress
36 Months	All Patients	EuroQOL to Core	EuroQOL to Core	EuroQOL to Core
	Patients Able	6-minute Walk	6-minute Walk	6-minute Walk
48 Months	All Patients	EuroQOL to Core	EuroQOL to Core	EuroQOL to Core
	Patients Able	6-minute Walk	6-minute Walk	6-minute Walk

4.6 Efficacy Assessments

4.6.1 Primary Efficacy Parameters

Documentation for occurrences of death or hospitalization will be required for submission to the Endpoint Committee. The Endpoint Committee will adjudicate causes of death and hospitalization based upon pre-specified definitions and procedures for this study.

1. All-cause mortality (time to death).
2. Survival free of hospitalization for cardiac cause.

4.6.2 Secondary Efficacy Parameters

1. All cause mortality at 30 days.
2. Survival free of hospitalization for heart failure.
3. Cardiovascular mortality (defined as sudden death, or death attributed to recurrent myocardial infarction, heart failure, a cardiovascular procedure, stroke, or other cardiovascular etiology).
4. Cardiovascular mortality and morbidity.
5. Survival free of revascularization.
6. Survival free of cardiac transplantation or LVAD implantation.
7. All-cause (unplanned and elective) hospitalization.

- To initially review the study protocol and commission a specific regular process for evaluation of STICH trial data.
- To meet no less than twice yearly for the duration of the trial to monitor the trial progress using confidential data summarizing trial safety and outcomes and to prepare a summary report of their review to the Executive Committee.
- To advise the Executive Committee on the implications of HF treatment advances on the conduct of the STICH trial.
- To approve protocol modifications if warranted.
- To advise the NHLBI directly regarding recommendations for trial modification or early trial cessation.

The efficacy and safety analyses will be performed semi-blinded (i.e., treatments A, B and C) by the Statistical Data Coordinating Center. The DSMB statistician will possess a copy of the treatment codes for unblinding purposes if deemed necessary by the DSMB. The study may be amended, or stopped early, or a treatment arm may be discontinued should any of these be deemed necessary based upon DSMB recommendation. The chairman of the DSMB will discuss the recommendations of the DSMB with the NHLBI project officer who will inform the Executive Committee. The chairman of the Executive Committee will notify the Steering Committee of any significant proposed amendment to the study or recommended discontinuation of the study. The roles, responsibilities and procedures of the DSMB are summarized in the DSMB Manual.

4.7.2 Operational Committees

The Executive Committee, with approval of the Steering Committee, is empowered to appoint operational committees to assist with trial work. Active committees are described in the Manual of Operations.

5. Data Management

5.1 Data Collection

5.1.1 Data Forms

Data collection forms used in this study are tabulated in the manual of operations with summaries of their content and purpose. The front page of each form will identify the date and version of the form, and provide brief summary instructions for data collection. The subject's study number and initials will be included on every page of each form. The study subject number will reflect both a site specific code, the subject's sequence number for that site and a check digit. Each form will require the signature of the physician investigator and the study coordinator with the date of completion.

5.1.2 Manual of Operations

The manual of operations provides more detail regarding study objectives, scope, design and operational policies and procedures than the clinical protocol. The manual can be accessed on the STICH Trial website (www.stichtrial.org).

5.1.3 STICH Trial Website

The STICH Trial website provides printable updated versions of the protocol and manual of operations. The website provides access to a complete directory of participating physician investigators and study coordinators at all clinical sites. The website provides a rich source of current information to investigators with secure access. Free access is available to the public to a portion of the site containing general information.

5.2 Database Management and Quality Control

Database management and quality control for this study are the responsibility of Duke Clinical Research Institute, Durham, NC, USA. Structured data elements from the CRFs are entered into the STICH database and reviewed using double data entry for verification. Coexistent diseases and adverse events are coded using a standard coding dictionary, MEDDRA 1.5. Concomitant medications are coded using a standard medication dictionary, WHO DRL. Information entered into the database is systematically checked by data management staff, using error messages generated from validation programs and database listings. Obvious errors are corrected by data management center personnel. Other errors, omissions or questions are entered on data query forms, which will be returned to the investigational site for resolution. After the investigator response is received at the data management center, the resolutions will be entered into the database. A copy of the signed data query form is kept with the CRFs. Quality control audits of all key safety and efficacy data in the database are made at designated times during the study. All clinical site patient-related reimbursement is prompted by completion of data forms with appropriate responses to all data elements.

6. Statistical Plan

6.1 Sample Size Determination - Overview

Several design factors and research objectives have been considered in developing appropriate sample size estimates for the study. First, patient enrollment has been determined so there would be a sufficient number of endpoints to provide a high degree of confidence (power $\geq$ 90%) for testing both primary hypotheses. Second, important secondary endpoints, including cardiovascular mortality, short-term (30-day) mortality, and measures of quality-of-life have also been considered. Third, we considered it important for the overall sample to be large enough to permit a prudent examination of treatment effects in selected subgroups of patients where surgical intervention might be particularly advantageous, or where the question of a treatment benefit from surgical intervention is particularly relevant. Important pre-specified subgroups of interest in this study include those defined by age, gender, race, HF class, ejection fraction, and the strata outlined in Figure 1. Finally, the sample size has been determined to provide a reasonable level of confidence of detecting clinically important therapeutic effects even in the event that current projections of event rates and treatment differences prove to be optimistic.

6.1.1 Sample Size and Power Considerations

A combination of survival data from recent HF trials that reflect patient outcomes when treated with intensive medical therapy (including beta blockers and ACE inhibitors) and historical data from the Duke Cardiovascular Database provide useful information on the range of patient outcomes that would be expected among the HF patients enrolled in STICH. Based on this information, we project that for testing the coronary revascularization hypothesis, the mortality rate in the medically treated arm of STICH will be approximately 25% after 3 years of follow-up. For the LV reconstruction hypothesis, we project that mortality or cardiac hospitalization will occur in the CABG (without SVR) arm with a 3-year rate of 50% or higher. Recognizing, however, that the actual event rates in STICH may vary somewhat from these estimates, we calculated sample size requirements for several different combinations of event rates and power levels in order to examine the sensitivity of the required sample size to different event rates and outcome scenarios that might conceivably arise in this trial.

In the absence of detailed life-table follow-up data on patients similar to those who will be enrolled in the trial and treated with intensive medical therapy, the following approach was used for calculating sample size requirements for the coronary revascularization hypothesis. (A similar approach was used for the LV reconstruction hypothesis, although the event rates differed because the two primary hypotheses are based on different endpoints). For a specified benchmark outcome rate in the medically treated (control) group (e.g., an overall event rate at 3 years of 25%), event rates at other time points during follow-up were estimated by assuming an exponential survival distribution. We then postulated that in the CABG arm, there would be a specific reduction of this benchmark rate (e.g., a 25% reduction). Event rates for the CABG group at various time points during follow-up were also estimated assuming an exponential survival distribution. Since the two groups will be compared with respect to length of survival using the log-rank test,[29] or equivalently, the Cox proportional hazards model,[29-31] the one-tailed formula of Schoenfeld[32] was modified to allow calculating sample size requirements based on the use of two-tailed tests.

Schoenfeld's method is based on first calculating the total number of events required in the combined treatment groups. By estimating the proportion of patients who will have an event by the end of the study, the total number of patients required for the trial can be determined. The necessary number of events depends on the level of power desired, the significance level (which we have chosen to be α=0.05), and the ratio of the hazard function for patients receiving surgery to the hazard function for patients in the medically treated arm.

Once the number of events has been determined, the proportion of patients who will experience an event during the patient follow-up period can be estimated using the method in Section 2.1 of Schoenfeld.[32] The total number of patients required is simply the number of events divided by the proportion of patients who will experience an event. A key issue in determining an adequate sample size for this trial is the magnitude of the event reduction we could expect to achieve in the CABG arm compared to the medical arm. Based on what is known about the effects of CABG in other patient populations, it is reasonable to postulate that surgery will be sufficiently effective to reduce overall mortality by 25%. A mortality reduction of this magnitude would be highly important from a clinical and public health standpoint, given the large population of patients in this country and throughout the world who suffer from moderate to severe HF secondary to CAD.

Sample Size for Hypothesis 1: Coronary Revascularization (all-cause mortality)

The three-year mortality rate was assumed to be 25% in the medically-treated arm in the design of the Hypothesis 1 component of STICH. The ischemic cohort of SCD-HeFT randomized to medical therapy (the control arm of that trial) had a three-year mortality of 28.5%. Although some differences exist between the SCD-HeFT and STICH populations (e.g., SCD-HeFT enrolled all NYHA class II or class III heart failure patients, whereas 85% of STICH patients are class II or class III), the differences are not expected to result in markedly lower mortality among the STICH medical cohort. A 25% reduction in mortality with CABG is hypothesized in the design of STICH. The level of statistical power chosen for assessing treatment differences in mortality is 90%.

Although STICH was originally designed to enroll a fixed number of patients, the number of patients (the sample size) and the length of time they would be followed were driven by the number of events required to achieve a statistical power of 90%. Due to the slower than anticipated enrollment in Hypothesis 1, the optimal use of available resources for the trial mandated that we carefully choose the number of patients and the duration of their follow-up to accrue the number of events required to achieve the desired level of statistical power. Approximately 400 events are required to achieve 90% power for detecting the hypothesized mortality reduction.

Statistical Power Tradeoffs. We have examined the tradeoff between the number of patients enrolled and the length of follow-up required to achieve a given level of statistical power, as reflected in Table 4.

TABLE 4. Hypothesis 1 – Statistical Power in Relation to Number of Patients and Length of Follow-up

Total	Average Length of Follow-up				
Patients	4.0 yrs	4.5 yrs	5.0 yrs	5.5 yrs	6.0 yrs
900	69%	73%	77%	80%	83%
1,000	73%	77%	81%	84%	87%
1,100	77%	81%	84%	87%	90%
1,200	81%	84%	87%	90%	>90%

For the primary mortality endpoint in Hypothesis 1, 80% and 90% power can be achieved under any of the scenarios outlined in Table 5.

TABLE 5. Hypothesis 1 – Enrollment, Follow-up, and Statistical Power

Total Patients	Average Follow-up Required to Achieve	
	80% Power	90% Power
900	5.5 yrs (Dec. 2010)	7.0 yrs (June 2012)
1,000	5.0 yrs (June 2010)	6.5 yrs (Dec. 2011)
1,100	4.5 yrs (Dec. 2009)	6.0 yrs (June 2011)
1,200	4.0 yrs (June 2009)	5.5 yrs (Dec. 2010)

If only 900 patients are enrolled, follow-up that is long enough to achieve 90% power will extend the study beyond what the budget will allow. An enrollment of 1,000 patients, our minimum acceptable sample size, prolongs the study rather substantially in order to achieve 90% power. A more desirable option as far as achieving the desired 90% level of power while not unduly extending follow-up is to enroll 1,200 patients and extend follow-up by 2 additional years (through 2010).

Sample Size for Hypothesis 2: LV Reconstruction (mortality or hospitalization for a cardiac reason):

For the LV reconstruction hypothesis where the primary endpoint is death (all cause) or hospitalization for a cardiac reason, the three-year event rate in the control arm (patients treated with CABG without SVR) is expected to be in the range of 50% or higher. If the control rate at 3 years is 50%, 387 patients per arm will provide 90% power for detecting a 20% reduction in the incidence of the primary endpoint, and 460 patients per arm will provide 90% power for detecting a 20% reduction if the control rate should be as low as 45%.

To achieve a robust sample size for testing the LV reconstruction hypothesis, 1,000 patients will be enrolled. Below is a summary of what this number of patients will provide the study.

1. Power >90% for detecting a *20%* reduction in the primary endpoint with CABG + SVR if the event rate at 3 years in patients treated with CABG (without SVR) is 45% or higher. *(Thus we have high power for detecting a conservative estimate of benefit if the control arm event rate is consistent with what is expected based on other studies (or even somewhat less than expected based on other studies).*
2. Power $\geq$ 85% for detecting a *20%* improvement if the 3-year control event rate should be as low as 40%. *(Thus we have high power for detecting a conservative estimate of the benefit if the control arm event rate is lower than expected.)*
3. Adequate power for detecting clinically important effects in subgroups of interest.

Secondary Endpoints:

Cardiac Mortality

Since a very high proportion of the deaths in this patient population will be cardiovascular, the power for detecting differences in this secondary endpoint will differ little from that outlined above for the primary endpoint of total mortality in assessing the coronary revascularization hypothesis.

6.1.2 Treatment Crossovers

The sample size calculations previously outlined do not include any adjustment to compensate for the fact that some patients enrolled in STICH may crossover from one treatment arm to another. Although participating clinical centers will be instructed concerning the importance of treatment compliance, and strict guidelines will be established regarding any changes to a patient's assigned therapy, we recognize that in certain instances a downstream change may be clinically indicated. Although all therapy changes and reasons for them will be documented, they are considered part of the initial treatment strategy, and groups will be compared with respect to the initially assigned treatment. Nonetheless, we have attempted to estimate the extent to which crossovers or dropouts might occur, and how such changes in therapy could affect the sample size/power figures cited above. From personal communication with investigators from the Bypass Angioplasty Revascularization Investigation (BARI), and from reviewing data in the Duke Cardiovascular Database, there is ample experience to suggest that the percent of consenting patients who are randomized to CABG, but don't actually receive CABG surgery, will be low (1-2%). We expect 5% crossovers from CABG without SVR to CABG with SVR and 5% of the patients randomized to SVR to not actually undergo the SVR procedure.

Crossover is most likely to occur among patients randomized to MED who, because of deteriorating symptoms or evidence of ischemia during follow-up, are felt to need revascularization with CABG. We project that up to 10% of MED patients may crossover to CABG within the first year following randomization, and up to 15% may crossover to CABG within three years. Some of the MED patients will receive percutaneous intervention. Indeed there may be patients in all three arms of the trial who undergo percutaneous intervention during follow-up, although this is more likely to occur among patients randomized to MED. Percutaneous intervention is not regarded as a treatment crossover, but rather as part of the downstream care associated with any of the treatment strategies in the trial.

Even if we assume that dropouts and treatment crossovers would reduce the effective sample size by as much as 15-20% (the impact is not expected to be that large), the randomization of 1,000 patients for testing the coronary revascularization hypothesis will still provide > 80% power for detecting a 25% reduction in mortality if the event rate in MED patients is consistent with our projection (25% at three years). For the LV

reconstruction hypothesis, the planned sample size of 1,000 patients will provide 90% power for detecting a 20% reduction in the primary endpoint even with liberal estimates for the amount of crossover if the three-year event rate for mortality or cardiac hospitalization is 50% or higher.

6.2 Statistical Analysis

Statistical analysis will be performed at the Statistical and Data Coordinating Center at Duke University. Although the methodologic approaches and operational details of the data analysis will be coordinated by the study biostatisticians, the major analyses of the study data will be highly collaborative, involving both statisticians and physicians to ensure appropriate interpretation of the data. All major treatment comparisons between the randomized groups in this trial will be performed according to the principle of "intention-to-treat;" that is, subjects will be analyzed (and endpoints attributed) according to the treatment arm to which patients were randomized, regardless of subsequent crossover. Statistical comparisons will be performed using two-sided significance tests. Additional perspective regarding the interpretation of the data will be provided through extensive use of confidence intervals[33] and graphical displays.

6.2.1 Background and Demographic Characteristics

Baseline demographic and clinical variables, including relevant descriptors from the history and physical examination; risk factors; comorbidity; HF functional class, angiographic assessment of the extent of CAD, left ventricular function, and descriptors obtained from the various noninvasive tests will be summarized for each randomized arm of the study. Descriptive summaries of the distribution of continuous baseline variables will be presented in terms of percentiles (e.g., median, 25th and 75th percentiles), while discrete variables will be summarized in terms of frequencies and percentages.

Statistical comparisons of treatment groups with respect to baseline characteristics will be limited to selected variables and disease factors known to influence prognosis. These variables will include age; sex; race; previous myocardial infarction; prior revascularization; descriptors of comorbidity; HF functional class; anatomic severity of disease; and left ventricular function (ejection fraction). For comparisons of treatment groups with respect to continuous baseline variables, emphasis will be given to nonparametric procedures such as the Wilcoxon rank sum test, or Kruskal-Wallis nonparametric analysis of variance.[34] Group comparisons with respect to discrete baseline variables will use the conventional chi-square test.

6.2.2 Concomitant Therapy

Summary statistics will be provided as appropriate. No formal analyses are planned.

6.2.3 Efficacy Evaluation

Coronary Revascularization Hypothesis

For the coronary revascularization hypothesis, the major statistical assessment will involve a comparison of the CABG arm (without SVR) versus the medically treated arm with respect to all-cause mortality. The log-rank test[29] (sometimes called the Mantel-Haenszel test for survival data[35]), which is a special case of the more general Cox proportional hazards regression model,[29-31] will be the primary analytic tool for assessing mortality differences between these two treatment arms. The log-rank test (and the Cox model) can accommodate varying lengths of patient follow-up, and not only uses information for each patient as to whether an event occurred, but also takes into account how long from study entry the patient survived before the endpoint occurred. The log-rank test also accommodates "censored" survival times, which arise because many patients will be alive when the analyses are performed, and the length of time they will survive

without an event is known only to be greater than the length of their current follow-up. The log-rank test is a well-established approach for providing an overall comparison of the entire survival curves.

Using this procedure, the analysis strategy will be to perform the treatment comparison stratifying by whether or not patients are SVR eligible (i.e., by whether patients are in Stratum A, or in Stratum B; see Figure 1). The analysis is thus reflective of the fact that the randomization will be stratified according to whether patients are in Stratum A versus Stratum B. This comparison will constitute the primary analysis to assess the effect of CABG on overall mortality. Kaplan-Meier survival estimates[36] based on the primary endpoint of all-cause mortality will be calculated for each treatment group to display the outcome results graphically.

Even though the primary comparison outlined above will contrast treatment groups that are determined by randomization, supplementary analysis involving other covariate adjustment will be performed with the Cox model. Such adjustment will be limited to a relatively small, prospectively defined set of patient characteristics that are known *a priori* to have a strong prognostic relationship with mortality. This adjustment will serve as a prelude to supplementary analyses examining differential treatment effects. The adjustment variables will include (in addition SVR eligibility) age, sex, race, HF class, history of myocardial infarction, previous revascularization, number of significantly diseased major coronary arteries, and ejection fraction. Cox model analyses may also be performed using geographic region as a stratification factor.

If the data provide evidence of an overall difference in outcome between treatment groups, we will further examine whether the therapeutic effect is similar for all patients, or whether it varies according to specific patient characteristics. In particular, we will focus on whether the relative therapeutic benefit differs according to patient age, sex, race, HF class, ejection fraction, and whether patients are eligible for SVR. These issues will be addressed formally with the Cox model by testing for interactions between treatment and these specific baseline variables.

Although the analysis will include an examination for treatment interactions as indicated above, treatment effects for the primary endpoint as characterized by the CABG:MED hazard ratio (with 95% confidence intervals) will be calculated and displayed for several prospectively-defined subgroups of patients defined by age, gender, race, SVR eligibility, HF class, and ejection fraction. These comparisons will be carefully interpreted in conjunction with the formal interaction tests.

To supplement the conventional significance testing and confidence interval approaches that will constitute the major analyses for this trial, we will provide additional perspective on the assessment of treatment effects using some established Bayesian approaches to the analysis of clinical trial data. The application of Bayesian methods to clinical trials has recently received considerable attention in the statistical and clinical trials literature. In part, its appeal stems from the fact that what consumers of clinical trial results often want to know is the likelihood (probability) that the treatment is actually beneficial, or the likelihood that the treatment has a clinically important benefit. Such probability assessments are directly obtainable from a Bayesian analysis. Spiegelhalter et al.[37] demonstrated how one can derive clinically useful information such as an estimate of the probability that the hazard ratio (CABG: medical therapy) is less than some specified value (e.g., 0.90) and an estimate of, for example, the probability that CABG therapy is "clinically equivalent" to intensive medical therapy, i.e., that the hazard ratio is within some interval close to 1.0 (such as 0.90 to 1.10). We will supplement the conventional statistical presentations discussed above by computing Bayesian probabilities that CABG therapy is beneficial and that it has a clinically important benefit. For these computations, a flat (non-informative) prior distribution for the CABG:MED hazard ratio will be assumed. We will use readily available S-Plus functions to perform the Bayesian calculations, or the

Cambridge group's BUGS program (Bayesian Inference Using Gibbs Sampling[38]) to derive a posterior distribution from a full Bayesian analysis.

LV Reconstruction Hypothesis

For the LV reconstruction hypothesis, the key statistical assessment will involve a comparison of CABG plus SVR versus CABG without SVR with respect to the composite endpoint of death (all cause) or hospitalization for a cardiac reason. The log-rank test will also be the primary analytic tool for this assessment. In this case, the test will be performed stratifying by whether the patients were eligible for either medical or surgical therapy versus whether surgery was indicated (i.e., by whether the patients were in Stratum B or Stratum C as depicted in Figure 1. Again, Kaplan-Meier estimates will be calculated and plotted for each treatment group to display the outcome results graphically. In this case, the Kaplan-Meier curves will display event-free survival, i.e., survival free of any cardiac hospitalization.

Again, the primary analysis will be supplemented by Cox model analyses involving other covariate adjustment. In addition to the stratification factor specified above for the primary LV reconstruction hypothesis, the small, prospectively-defined set of patient characteristics listed above as covariates for the coronary revascularization hypothesis will also be used in supportive analyses of the LV reconstruction hypothesis.

If the data provide evidence of an overall difference in outcome between treatment groups, we will determine whether the therapeutic effect is similar for all patients, or whether it varies according to specific patient characteristics. In particular, we will assess whether the relative therapeutic benefit differs according to patient age, sex, race, HF class, ejection fraction, and whether or not the patients are eligible for either medical or surgical therapy. This will be addressed formally with the Cox model by testing for interactions between treatment and these specific baseline variables.

We will provide additional perspective on the assessment of treatment effects with respect to the LV reconstruction hypothesis through calculating the hazard ratio for CABG with SVR relative to CABG without SVR and 95% confidence intervals (overall and in prospectively defined subgroups), and also through performing supplementary Bayesian analysis as outlined above for the coronary revascularization hypothesis.

Analysis of Secondary Endpoints

Secondary endpoints that will be evaluated in this trial include (1) cardiovascular mortality; (2) all-cause mortality within 30 days; (3) quality of life; (4) cost of care and cost effectiveness; (5) morbidity; (6) functional measures, including six minute walk distance and exercise duration on a treadmill test; and (7) physiologic measures from noninvasive testing, neurohormonal levels, and cytokine levels. Data analyses for each of these endpoints are discussed below.

Cardiovascular Mortality

This endpoint is most relevant for the assessment of the effects of coronary revascularization, and therefore will be used in comparisons of medical therapy vs. CABG (without SVR). The analysis will be similar to that outlined for the primary endpoint, using time from enrollment until death from a cardiovascular cause or censoring as the response variable, and assessing treatment differences using the Cox proportional hazards model. Kaplan-Meier survival curves will be computed to graphically display the survival experience of the treatment groups as a function of time from randomization.

Mortality within 30 days

The analysis for this endpoint will also be similar to that outlined for the primary mortality endpoint except that 30-day outcome will be treated as a binary endpoint (due to its short-term nature), and the logistic regression model will be used for treatment comparisons rather than the Cox proportional hazards model. These analyses will be performed for both MED versus CABG and for CABG + SVR versus CABG (without SVR) in order to fully characterize and compare procedural mortality. Thirty-day mortality rates will be reported for each treatment arm along with odds ratios and 95% confidence intervals for CABG:MED and for CABG + SVR:CABG (without SVR).

Quality of Life and Cost of Care

The important quality of life and cost of care endpoints analysis plans are addressed in detail in the EQOL Core Laboratory manual. These important endpoints include health status measures from the Medical Outcomes Study 36 Item Short Form (SF-36); quality of life and cardiac symptoms as assessed by the disease-specific Kansas City Cardiomyopathy Questionnaire and Seattle Angina Questionnaire; a measure of depressive symptoms based on the Center for Epidemiologic Studies-Depression (CES-D) scale; and utilities assessed by patient interview using the EuroQol.

Morbidity

Morbidity in this trial will be assessed by examining the incidence of several different complications and clinical events that occur during the follow-up period. These events will include complications associated with each treatment arm such as the frequency and duration of hospitalization for HF, and additional cardiac procedures (heart transplantation, LVAD, ICD, and stroke). We outline below the approach that we will take for analyzing one of these important complications, for example hospitalization for HF, noting that specific components of this approach can also be applied to the other non-fatal complications. The approach involves several related yet different analyses to comprehensively assess and fully characterize differences among the treatment arms with respect to this outcome.

One characterization of this secondary endpoint will be in terms of the time from randomization until a patient experiences their first hospitalization for HF during their follow-up in the trial. The methodology for analyzing censored failure-time data outlined previously for the primary endpoint will thus be applied to compare treatment groups with respect to this outcome. We point out that this analysis must be interpreted cautiously, however, because hospitalization as a non-fatal event may appear to occur with a lower incidence in one treatment group simply because that arm had a higher death rate. Obviously after patients die, they can no longer be hospitalized for their HF. To deal with this complexity, we will perform further analyses of this adverse complication by considering hospitalization for HF and death as a combined endpoint. We will thus consider the time until the first occurrence of either death (any cause) or hospitalization for HF, and treatment differences will be assessed using the Cox proportional hazards model. In support of this analysis, Kaplan-Meier event-free survival curves for each treatment arm will be calculated to graphically display event rates over time for this composite endpoint.

The third approach for analyzing this endpoint will take advantage of the fact that the two components of the composite endpoint (death or hospitalization) have a natural rank-ordering in terms of their severity. The treatments can thus be compared with an approach that not only uses the time until the first occurrence of one of these component outcomes, but also factors in the severity of the outcome. The approach we will employ, developed by Berridge and Whitehead,[39] combines the Cox proportional hazards model in conjunction with an ordinal severity of event model to potentially increase power for the treatment comparisons. The severity of event portion is a continuation ratio-type ordinal logistic model. An extended hazard function is defined by multiplying the regular Cox hazard function by the probability of the event being of severity j from the ordinal model, where j references the various severity categories. In this analysis, death will obviously be

considered as the most severe event, and hospitalization for HF as least severe. Within this framework, one can test whether one of the interventions under study in this trial prolongs the time to the first occurrence of either of these events, whether the therapy decreases the severity of the first event that occurs, or (with 2 degrees of freedom), whether the therapy has either effect.

Finally, since some patients may experience hospitalization multiple times, a comprehensive comparison of the treatment arms with respect to this outcome should take into account the longitudinal pattern in the repeated occurrences of such an event. In clinical situations where there may be multiple events per subject, it is desirable to be able to apply time-to-event methods such as the proportional hazards model which explicitly allow for varying lengths of patient follow-up and appropriately handle censored observations. A major issue in extending proportional hazards regression models to this situation, however, is intra-subject correlation (i.e., where multiple events occur within the same patient, those events will be correlated). Fortunately, there is active, ongoing methodological research on the application of survival models to this situation.[40-43] In order to provide a more complete and comprehensive analysis of the hospitalization data, the approach we will use in STICH is based on extensions of the proportional hazards model that accommodate multiple events per patient. The specific approach makes use of cluster modifications of so-called *sandwich* estimates of the variance-covariance matrix of the regression coefficients, thus providing standard errors of the regression coefficients that take into account the correlations among multiple event times within a given patient.[41-43] Specialized software functions for performing such analyses are available in S-Plus.[43]

Because of the problems already mentioned in analyzing hospitalization by itself, we will use the multiple-event methodology outlined above, which not only allows multiple events per patient of the same type (e.g., multiple hospitalizations), but also accommodates events of different types. Thus we will model both death and the multiple episodes of rehospitalization using multiple-events Cox model analysis.[41-43] By synthesizing the results from these different but complementary approaches (that is, by considering hospitalization in terms of a single event per patient, in terms of multiple events per patient, and hospitalization combined with death), we will be able to provide a comprehensive assessment of treatment differences with respect to hospitalization, and as part of the analyses, identify and assess other clinical factors that are associated with this important secondary outcome.

The approaches outlined above for analyzing hospitalization for HF can also be applied to other clinical outcomes and adverse events that will be of interest in this trial, such as stroke and other morbidity outcomes enumerated above.

Functional Capacity Outcomes

These outcomes will include distance covered in a six-minute walk test and exercise duration on a treadmill test. Of interest will be both the absolute magnitude of these measures as obtained during follow-up and also the change from baseline to the follow-up measurement. Treatment group differences will be statistically assessed using the nonparametric Wilcoxon rank sum test.

Physiologic Measures from Noninvasive Tests (Core Laboratory Data)

There will be considerable noninvasive test data acquired in the trial, including cardiac magnetic resonance imaging, echocardiography, and nuclear cardiology (radionuclide) studies. The schedule for obtaining these various tests is outlined in Tables 2 and 3 and generally involves the acquisition of data at baseline, at four months following study entry, and after two years of follow-up. This information will be valuable in addressing the mechanisms responsible for differences or similarities in patient outcomes observed among the randomized arms in the trial. Furthermore, with treatment groups defined by randomization and tests systematically performed and uniformly interpreted, bias in the choice of therapy and in the choice of which tests are obtained is eliminated. The data from these various tests may be very important for identifying

subsets of patients where there are clear benefits of myocardial revascularization, and perhaps other subsets where intensive medical therapy is an attractive choice. Even where revascularization appears to be a superior strategy, there may be some patients who do and others who do not benefit from SVR. The noninvasive information acquired in the trial will therefore be very useful in defining a pretreatment strategy for performing a sequence of tests that will optimize the choice of therapy.

The analysis of the data from these tests will consist of several parts: (1) a detailed assessment of the prognostic relationships of the baseline test data to clinical outcomes, and of the relative prognostic importance of each test; (2) an examination of differential treatment effects as a function of the baseline noninvasive test data; (3) a comparison between treatments of the changes in test results from baseline to four months as a characterization of early mechanistic effects of therapy; (4) an assessment of the four-month measures and the change from baseline to four months as predictors of subsequent outcomes; (5) a comparison of treatments with respect to the two year test outcomes and the change from baseline to two years as viewed from the perspective of an endpoint, and also as it might relate to late clinical outcomes; (6) an analysis comparing treatment arms with respect to the trajectory of key noninvasive test measures, considering all three measurement times (baseline, four months, two years).

For the analyses in part (1), we will use the Cox proportional hazards model to examine individual and joint relations between baseline test results and length of survival (or event-free survival in the case of the LV reconstruction endpoint). For continuous baseline measures such as ejection fraction and ESVI, we will examine the shape and strength of the relationship by use of a flexible model-fitting approach involving cubic spline functions (cubic polynomials).[44-46] The resulting functions will be graphically and statistically examined to assess the model assumptions and to judge the linearity of the relationships. Where relations are nonlinear, their shape will be characterized using spline functions. Determining how variables should be modeled will be an important step in characterizing prognostic relations and identifying which variables are most strongly related to long-term outcome. Once the important predictors from each individual test are identified, we will contrast the relative prognostic importance and the independent contribution of the various tests by jointly considering multiple tests. The modeling strategies to be used and the approach to evaluating the yield of each test are similar to those we have employed in previous studies.[47,48] Because there will be over 400 deaths occurring in the study population during the period of follow-up, the sample size (number of events) will adequately support these extensive analyses.

For step (2), we will use the key features identified in step (1) from each noninvasive test, and statistically test for a differential treatment effect using the Cox model. This will involve testing for interactions between treatment and specific baseline test measures (e.g., ESVI, measures of myocardial viability and myocardial perfusion).

For step (3), the four-month test results and the change from baseline to four months in the prognostic factors identified in step (1) above will be descriptively summarized, and treatments will be compared using the Wilcoxon rank sum test in the case of continuous or ordinal measures, and the chi-square test for categorical variables.

For step (4) where we are dealing with post-randomization information and attempting to judge its usefulness for predicting subsequent outcomes, the analysis becomes more complex. We will use two approaches. First, we will assess the relations of the four-month noninvasive test data to subsequent outcomes (outcomes after four months) in a fashion similar to that described above for the analysis of the baseline test data. Second, we will include events starting from randomization, but consider the four-month test results as intervening or time-dependent data, and factor this into the Cox model using time-dependent covariates. In both of these analyses, we will seek to address the issue of whether the four-month test results

or the change from baseline to four months provides independent predictive information beyond that contained in the baseline test data, and thereby assess the utility of doing the follow-up studies.

For step (5), treatment arms will be compared with respect to two-year test outcomes (considering the test results as an endpoint) using a similar approach as described for the four-month data in step (3). We will also attempt to judge the utility of the two-year test data in predicting subsequent outcomes as in step (4) above, recognizing, however, that the complexities of this analysis will require very cautious interpretation.

For step (6), we will characterize the trajectory of key prognostic measures from the tests by considering all three measurement times. In considering the longitudinal nature of these data, there are three major statistical issues that must be considered in making treatment comparisons: (a) the data may not be normally distributed, (b) the longitudinal measures within subjects will be correlated, and (3) there will be "missing" data, resulting from patients in whom it is not possible to obtain all three measurements (because they die or fail to return for other reasons to enable the follow-up measurements to be made). To address the first two issues, we will use the generalized estimating equation (GEE) approach to model estimation and testing proposed by Liang and Zeger.[49] The software for implementing this method is now available in a SAS macro-routine accessible to the CC. For the third issue above (missing data), we will, of course, make every effort to minimize the amount of missing data. However, because there will inherently be missing data, we will incorporate several strategies and assess the sensitivity of our conclusions to the missing data approach employed. The alternatives that will be considered include (1) analyze only the patients with complete data; (2) assign "worst case" scores to missing observations; (3) carry forward previously observed data to observation periods that are missed later; and (4) use likelihood-based methods to impute missing data. A description of these methods and their limitations is given in the review by Heyting.[50]

By synthesizing the results from these different but complementary approaches, we will be able to provide a comprehensive assessment of treatment differences with respect to the key mechanistic noninvasive test data, judge whether there is a preference for one particular treatment depending on the results of the noninvasive tests, and develop a recommendation for performing the sequence of tests that will optimize the choice of therapy.

With regard to physiologic measures such as neurohormonal levels and cytokine levels provided by the Neurohormone/Cytokine/Genetic Core Laboratory, there will be interest in both the absolute levels at follow-up and also in the change from baseline to follow-up. Treatment group differences will be statistically assessed using the Wilcoxon rank sum test applied to these various measures.

Multiple Comparisons

Although there are two primary aims, and there will be some overlap of the patients used for testing the two primary hypotheses [the overlapping patients will be those in Stratum B who are randomized to CABG (without SVR), projected to be approximately 75 patients], the statistical test for each primary hypothesis will be performed at the 0.05 level of significance. We justify this choice on the basis that the two hypotheses are addressing two very distinct questions. Indeed, the endpoints for the two questions are different, and were it not for the overlapping patients, the study would basically resemble two separate trials.

With two primary hypotheses and the various secondary endpoints that have been outlined, we recognize that there is a multiplicity of analyses to be performed, which leads to an increased probability that at least one of the comparisons could be "significant" by chance. There are adjustments (e.g., based on the Bonferroni inequality) that can be used to preserve the overall type I error level. To attempt to adjust for the effects of the repeated significance testing that will occur as part of the interim monitoring (discussed below), plus adjust for the multiplicity of analyses, would require that small significance levels be used for

every comparison. Instead of formally adjusting the significance level for the fact that there are two primary hypotheses and numerous secondary comparisons, we will be conservative in the interpretation of the analyses, taking into account the degree of significance, and looking for consistency across endpoints. The actual p-value for each comparison will be reported (not simply whether significance is achieved) to aid in the overall interpretation. Also, the Bayesian interpretations discussed above (along with so-called credible intervals) will assist in providing an appropriate interpretation of the study results. We have also pre-specified the primary and secondary outcome variables to help guard against the multiple testing problems.

6.2.4 Safety Evaluation

The assessment of safety is based mainly on the frequency of the pre-defined safety parameters and serious adverse events suspected by the investigator to be related to study treatment. Other safety data (e.g., vital signs) will be considered and summarized as appropriate.

Serious adverse events suspected by the investigator to be related to study treatment will be summarized for each treatment group by presenting the number and percentage of patients having any serious related adverse event, having a serious related event in each body system and having each individual serious related adverse event.

6.2.5 Interim Analysis

For ethical reasons, an interim examination of key safety and endpoint data will be performed at regular intervals during the course of the trial. The primary objective of these analyses will be to evaluate the accumulating data for an unacceptably high frequency of negative clinical outcomes in any of the treatment arms. In addition, however, the interim monitoring will also involve a review of the control arm event rates, patient recruitment, compliance with the study protocol, submission of data forms, and other factors which reflect the overall progress and integrity of the study. The results of the interim analyses will be carefully and confidentially reviewed by a Data and Safety Monitoring Board (DSMB).

It is anticipated that the DSMB will meet at approximately six-month intervals to review the accumulating data. Prior to each meeting, the Statistical Data Coordinating Center will conduct the desired statistical analyses and prepare a summary report that will be carefully reviewed by the DSMB. The extracted data files and analysis programs for each DSMB report will be archived and maintained at the Statistical Data Coordinating Center for the life of the study. Reports will be presented describing the progress of patient enrollment, the rates of compliance with therapy, and the frequency of protocol violations. The Statistical Data Coordinating Center will also report on the number of forms received, the number of forms entered in the computer, the number of forms queried, the number of queries completed, the number of forms reviewed through on-site monitoring, and the number of forms receiving the final, completed entry in the double data-entry system.

These interim safety and efficacy reports introduce well-recognized statistical problems related to the multiplicity of statistical tests performed on an accumulating set of data.[51,52] As a solution to the problem of repeated tests, we propose to adopt for use in STICH a group sequential method similar to that proposed by O'Brien and Fleming[53] as a guide in interpreting interim analyses. This procedure requires large critical values early in the study, but relaxes (i.e., decreases) the critical value as the trial progresses. Because of the conservatism early in the trial, the critical value at the final analysis is near the "nominal" critical value. Hence the sample size requirements with this group sequential procedure remain essentially the same as the conventional fixed sample size estimate. The actual method for this interim monitoring that will be employed in STICH is the general approach to group sequential testing developed by Lan and DeMets[54] for which neither the number of looks nor the increments between looks must be pre-specified. Rather, the Lan-DeMets approach only requires specification of the rate at which the Type I error (which in this trial will be

chosen to be α=0.05) will be "spent". This procedure allows "spending" a little of α at each interim analysis in such a way that at the end of the study, the total Type I error does not exceed 0.05. One such spending function generates boundaries that are nearly identical to the O'Brien-Fleming boundaries. It is this approach that we propose to use in STICH, namely two-sided, symmetric O'Brien-Fleming[53] type boundaries generated using the flexible Lan-DeMets[54] approach to group sequential testing. Since the number of looks and the increments between looks need not be pre-specified, it allows considerable flexibility in the monitoring process for accommodating additional comparative examinations of the data in response to concerns of the DSMB that may arise during the course of the trial. Assuming that the DSMB will conduct its first formal data review in the latter half of the first year of recruitment, and then continue those reviews approximately every 6 months thereafter through the patient recruitment period (three years) and the follow-up phase (three years), there will be approximately eight to ten reviews of the data. With nine interim analyses approximately equally spaced in time, the Lan and DeMets "spending function" that approximates the O'Brien-Fleming stopping boundaries involves a very stringent alpha level (0.00001) for declaring significance at the first interim analysis. At the subsequent interim analyses, the required significance levels will be somewhat less stringent. The requirements for significance at each interim analysis, depending on exactly when the analysis occurs, can be computed with the Lan-DeMets methodology, for which we have suitable computer software. The final analysis can be undertaken with a significance level of approximately 0.04, relatively close to the nominal 0.05 level. This approach to interim monitoring will be carried out in parallel for both of the primary study hypotheses.

The analytic approach that will be used at the interim analyses for assessing treatment differences will be the time-to-event analysis methods described above, except that interpretation of statistical significance associated with treatment comparisons of key study endpoints will be guided using the group sequential stopping boundaries outlined above.[53-55] The appropriateness of using the log-rank test (or equivalently the Cox model) in the group sequential framework has previously been well established.[56-59] For each of these interim analyses, the critical value of the test statistic and the corresponding p-value required for significance in that particular analysis will be presented so that significance can be assessed precisely. If significantly large and important treatment differences are observed at any of the interim analyses, the DSMB may recommend that randomization of patients be stopped, or that the design and conduct of the trial be appropriately modified. Judgment concerning the continuation or termination of the study will involve not only the degree of statistical significance observed at the interim analysis, but also the likelihood of achieving significance should enrollment continue to the originally projected sample size. As an aid in this latter assessment, the Statistical Data Coordinating Center will supplement the group sequential analyses outlined above with calculations of conditional power based on the method of stochastic curtailment.[60-62]

This procedure evaluates the conditional probability that a particular statistical comparison will or will not be significant at the end of the trial at the α level used in the design, given the hypothesized treatment difference and the data obtained to date. Conditional power for the primary comparisons associated with both primary hypotheses will be computed and provided to the DSMB as part of the interim study reports.

The approach to interim monitoring outlined above will be carried out in parallel for both of the primary study hypotheses. In addition, for the LV reconstruction hypothesis where the primary endpoint is a composite of death or cardiac hospitalization, it will also be important to carefully monitor the mortality component of this endpoint. Thus mortality rates and associated confidence intervals for each arm (CABG and CABG + SVR) will also be monitored at the interim reviews to ensure that the safety of patients enrolled in the trial is not compromised.

7. Ethics and Good Clinical Practice

This study must be carried out in compliance with the protocol and in accordance with STICH standard operating procedures. These procedures are designed to ensure adherence to Good Clinical Practice, as described in the following documents:

1. ICH Harmonized Tripartite Guidelines for Good Clinical Practice 1996.
2. Directive 91/507/EEC, The Rules Governing Medicinal Products in the European Community.
3. US 21 Code of Federal Regulations dealing with clinical studies (including parts 50 and 56 concerning informed consent and IRB regulations).
4. Declaration of Helsinki, concerning medical research in humans (Recommendations Guiding Physicians in Biomedical Research Involving Human Subjects, Helsinki 1964, amended Tokyo 1975, Venice 1983, Hong Kong 1989, Somerset West 1996).

The investigator agrees, when signing the contract with the Coordinating Center, to adhere to the instructions and procedures described in the protocol and thereby to adhere to the principles of Good Clinical Practice that it conforms to.

7.1 Institutional Review Board/Independent Ethics Committee

Before implementing this study, the protocol, the proposed informed consent form and other information to subjects, must be reviewed by a properly constituted Institutional Review Board/Independent Ethics Committee (IRB/IEC). A signed and dated statement that the protocol and informed consent have been approved by the IRB/IEC must be given to the Coordinating Center before study initiation. The name and occupation of the chairman and the members of the IRB/IEC must be supplied to the Coordinating Center. Any amendments to the protocol, other than administrative ones, must be approved by this committee.

7.2. Informed Consent

The investigator must explain to each subject (or legally authorized representative) the nature of the study, its purpose, the procedures involved, the expected duration, the potential risks and benefits involved and any discomfort it may entail. Each subject must be informed that participation in the study is voluntary and that he/she may withdraw from the study at any time and that withdrawal of consent will not affect his/her subsequent medical treatment or relationship with the treating physician.

Eligible patients will be approached for consent to enter a specific stratum determined by MED and SVR eligibility. This informed consent should be given by means of a standard written statement, written in non-technical language. The patient should read and consider the statement before signing and dating it, and should be given a copy of the signed document. If written consent is not possible, oral consent can be obtained if witnessed by a signed statement from one or more persons not involved in the study, mentioning why the patient was unable to sign the form. No patient can enter the study before his/her informed consent has been obtained. Once randomized, patients will be asked to provide a separate informed consent for genotyping.

The informed consent forms are part of the protocol, and must be submitted by the investigator with it for IRB/IEC approval. The Coordinating Center will supply proposed informed consent forms, which comply with regulatory requirements and are considered appropriate for the study. Any changes to the proposed consent form suggested by the investigator must be agreed to by the Coordinating Center before submission to the IRB/IEC, and a copy of the approved version must be provided to the Coordinating Center after IRB/IEC approval.

8. References

1. http://www.americanheart.org/statistics/pdf/HSSTATS2001/.0.pdf
2. O'Connell JB, Bristow MR. Economic impact of heart failure in the United States: Time for a different approach. *Heart Lung Transplant* 1993; 13:S107-S112.
3. Kaesemeyer WH. Holding smokers accountable for heart disease costs. *Circulation* 1994; 90:1029-1032.
4. Braunwald E, Bristow MR. Congestive heart failure: Fifty years of progress. *Circulation* 2000; 102:IV-14—IV-23.
5. Reimer KA, Jennings RB. The "Wavefront Phenomenon" of myocardial ischemic cell death. II. Transmural progression of necrosis within the framework of ischemic bed size (myocardium at risk) and collateral flow. *Lab Invest* 1979; 40:633-644.
6. Jatene AD. Left ventricular aneurysmectomy. Resection or reconstruction. *J Thorac Cardiovasc Surg* 1985; 89(3):321-321.
7. Dor V, Saab M, Coste P, Kornaszewska M, Montiglio F. Left ventricular aneurysm: a new surgical approach. *J Thorac Cardiovasc Surg* 1989; 37(1):11-19.
8. Di Donato M, Sabatier M, Dor V, Gensini GF, Toso A, Maioli M, Stanley AWH, Athanasuleas C, Buckberg G. Effects of the Dor procedure on left ventricular dimension and shape and geometric correlates of mitral regurgitation one year after surgery. *J Thorac Cardiovasc Surg* 2001; 121:91-96.
9. Athanasuleas CL, Stanley AWH, Buckberg GD, Dor V, DiDonato M, Blackstone EH, and the RESTORE Group. Surgical anterior ventricular endocardial restoration (SAVER) in the dilated remodeled ventricle following anterior myocardial infarction. *J Am Coll Cardiol* 2001; 37:1199-1209.
10. Jones RH. Is it time for a randomized trial of surgical treatment of ischemic heart failure? Invited Editorial. *J Am Coll Cardiol* 2001:1210-1213.
11. Alderman EL, Bourassa MG, Cohen LS, Davis KB, Kaiser GG, Killip T, Mock MB, Pettinger M, Robertson TL for the CASS Investigators. Ten-year follow-up of survival and myocardial infarction in the randomized Coronary Artery Surgery Study. *Circulation* 1990; 82:1629-1646.
12. Yusuf S, Zucker D, Peduzzi P, Fisher LD, Takaro T, Kennedy JW, Davis K, Killip T, Passamani E, Norris R, Morris C, Mathur V, Varnauskas E, Chalmers TC. Effect of coronary artery bypass grafting on survival: Overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. *Lancet* 1994; 344:563-570.
13. Baker DW, Jones R, Hodges J, Massie BM, Konstam MA, Rose EA. Management of Heart Failure. III. The role of revascularization in the treatment of patients with moderate or severe left ventricular systolic dysfunction. *JAMA* 1994; 272:1528-1534.
14. Eitzman D, al-Aouar Z, Kanter HL, et al. Clinical outcome of patients with advanced coronary artery disease after viability studies with positron emission tomography. *J Am Coll Cardiol* 1992; 20:559-565.
15. Di Carli MF, Davidson M, Little R, et al. Value of metabolic imaging with positron emission tomography for evaluating prognosis in patients with coronary artery disease and left ventricular dysfunction. *Am J Cardiol* 1994; 73:527-533.
16. Lee KS, Marwick TH, Cook SA, et al. Prognosis of patients with left ventricular dysfunction with and without viable myocardium after myocardial infarction: relative efficacy of medical therapy and revascularization. *Circulation* 1994; 90:2687-2694.
17. Gioia G, Powers J, Heo J, Iskandrian AS. Prognostic value of rest-redistribution tomographic thallium-201 imaging in ischemic cardiomyopathy. *Am J Cardiol* 1995; 75:759-762.
18. Vom Dahl J, Altehoefer C, Sheehan FH, et al. Effect of myocardial viability assessed by technetium-99m sestamibi SPECT and fluorine-18-FDG PET on clinical outcome in coronary artery disease. *J Nucl Med* 1997; 38:742-748.
19. Afridi I, Grayburn PA, Panza JA, Oh JK, Zoghbi WA, Marwick TH. Myocardial viability during dobutamine echocardiography predicts survival in patients with coronary artery disease and severe left ventricular systolic dysfunction. *J Am Coll Cardiol* 1998; 32:921-926.

20. Cuocolo A, Petretta M, Nicolai E, et al. Successful coronary revascularization improves prognosis in patients with previous myocardial infarction and evidence of viable myocardium at thallium-201 imaging. *Eur J Nucl Med* 1998; 25:60-68.
21. Pasquet A, Robert A, D'Hondt AM, Dion R, Melin JA, Vanoverschelde JL. Prognostic value of myocardial ischemia and viability in patients with chronic left ventricular ischemia. *Circulation* 1999; 100:141-141.
22. Senior R, Kaul S, Lahiri A. Myocardial viability on echocardiography predicts long-term survival after revascularization in patients with ischemic congestive heart failure. *J Am Coll Cardiol* 1999; 33:1848-1854.
23. Chaudhry FA, Tauke JT, Alessandrini RS, Vardi G, Parker MA, Bonow RO. Prognostic implications of myocardial contractile reserve in patients with coronary artery disease and left ventricular dysfunction. *J Am Coll Cardiol* 1999; 34:730-738.
24. Michels VV, Moll PP, Miller FA. The frequency of familial dilated cardiomyopathy in a series of patients with idiopathic dilated cardiomyopathy. *N Engl J Med* 1992; 326:77-82.
25. Leiden JM. The genetics of dilated cardiomyopathy: emerging clues to the puzzle. *N Engl J Med* 1997; 337:1080-1081.
26. Olson TM, Michels VV, Thibodeau SN, Tai YS, Keating MT. Actin mutations in dilated cardiomyopathy, a heritable form of heart failure. *Science* 1998; 280:750-752.
27. Dilsizian V, Rocco TP, Freedman NMT, Leon MB, Bonow RO. Enhanced detection of ischemic but viable myocardium by the reinjection of thallium after stress-redistribution imaging. N Engl J Med 1990;323:141-146.
28. Sciagra R, Bisi G, Santoro GM, et al. Comparison of baseline-nitrate technetium-99m sestamibi with rest-redistribution thallium-201 tomography in detecting viable hibernating myocardium and predicting postrevascularization recovery . J Am Coll Cardiol 1997;30:384-391.
29. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data.* New York, NY: John Wiley & Sons, Inc; 1980.
30. Cox DR. Regression models and life-tables (with discussion). J Royal Statist Soc B 1972;34:187-220.
31. Breslow NE. Covariance analysis of censored survival data. Biometrics 1974;30:89-99.
32. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. Biometrics 1983;39:499-503.
33. Borenstein M. The case for confidence intervals in controlled clinical trials. Contro Clin Trials 1994;15:441-428.
34. Lehmann EL. *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden-Day Inc.; 1975.
35. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 1966;50:163-170.
36. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Statist Assn 1958;53:457-481.
37. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. J Royal Stat Soc 1994;157:357-416.
38. Thomas A, Spiegelhalter DJ, Gilks WR. BUGS: A program to perform Bayesian inference using Gibbs sampling. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM. *Bayesian Statistics*. Oxford: Clarendon Press; 1992:837-842.
39. Berridge DM, Whitehead J. Analysis of failure time data with ordinal categories of response. Stat Med 1991;10:1703-1710.
40. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. J Am Statist Assoc 1993;88:811-820.
41. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Statist Assoc 1989;84:1065-1073.
42. Lee EW, Wei LJ, Amato D. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Klein JP, Goel PK, eds. *Survival Analysis, State of the Art.* Rotterdam: Kluwer Academic Publishers; 237-247.

43. Therneau TM, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag; 2000.
44. Harrell FE Jr, Pollock BG, Lee KL. Graphical methods for the analysis of survival data. Proc 12th Annual Conference. SAS Users Group International 1987;12:1107-1115.
45. Harrell FE Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Nat Cancer Inst 1988;80:1198-1202.
46. Smith PL. Splines as a useful and convenient statistical tool. Am Statistician 1979;33:57-62.
47. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543-2546.
48. Califf RM, Pieper KS, Lee KL, Van de Werf F, Simes RJ, Armstrong PW, Topol EJ for the GUSTO-I Investigators. Prediction of 1-year survival after thrombolysis for acute myocardial infarction in the Global Utilization of Streptokinase and TPA for Occluded Coronary Arteries Trial. Circulation 2000;101:2231-2238.
49. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13-22.
50. Heyting A, Tolboom JTBM, Essers JGA. Statistical handling of drop-outs in longitudinal clinical trials. Statistics in Medicine 1992;11:2043-2061.
51. Armitage P, McPherson CK, Rowe BC. Repeated significance tests on accumulating data. J R Stat Soc Series A 1969;132:235-244.
52. McPherson K. Statistics: the problem of examining accumulating data more than once. N Engl J Med 1974;290:501-502.
53. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979;35:549-556.
54. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659-663.
55. Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: Ramifications and guidelines for practitioners. Biometrics 1987;213-223.
56. Tsiatis, AA. The asymptotic joint distribution of the efficient scores tests for the proportional hazards model calculated over time. Biometrika 1981; 68:311-315.
57. Tsiatis AA. Repeated significance testing for a general class of statistics used in censored survival analysis. J Am Stat Assoc 1982;77:855-861.
58. Gail MH, DeMets DL and Stud EV. Simulation studies on increments of the two-sample log rank score test for survival data, with application to group sequential boundaries. Edited by R. Johnson and J. Crowley. In *Survival Analysis. Monograph Series 2.* Hayward, California: IMS Lecture Notes, 1982;287-301.
59. DeMets DL, Gail MH. Use of logrank test and group sequential methods at fixed calendar times. Biometrics 1985;41:1039-1044.
60. Halperin M, Lan KKG, Ware JH et al. An aid to data monitoring in long-term clinical trials. Controlled Clin Trials 1982;3:311-323.
61. Lan KKG, Simon R and Halperin M. Stochastically curtailed tests in long-term clinical trials. Commun Stat., Sequential Analysis. 1982;1:207-219.
62. Ware JH, Muller JE, Braunwald E. The futility index, an approach to the cost-effective termination of randomized clinical trials. Am J Cardiol 1985;78:635-643.