

## HFSA Working Group

# Designs for Mechanical Circulatory Support Device Studies

JAMES D. NEATON, PhD,<sup>1</sup> SHARON-LISE NORMAND, PhD,<sup>2</sup> ANNETINE GELIJNS, PhD,<sup>3</sup> RANDALL C. STARLING, MD,<sup>4</sup>  
DOUGLAS L. MANN, MD,<sup>5</sup> AND MARVIN A. KONSTAM, MD<sup>6</sup>

*Minneapolis, Minnesota; Boston, Massachusetts; New York, New York; Cleveland, Ohio; Houston, Texas*

### ABSTRACT

**Background:** There is increased interest in mechanical circulatory support devices (MCSs), such as implantable left ventricular assist devices (LVADs), as “destination” therapy for patients with advanced heart failure. Because patient availability to evaluate these devices is limited and randomized trials have been slow in enrolling patients, a workshop was convened to consider designs for MCS development including alternatives to randomized trials.

**Methods and Results:** A workshop was jointly planned by the Heart Failure Society of America and the US Food and Drug Administration and was convened in March 2006. One of the panels was asked to review different designs for evaluating new MCSs. Randomized trials have many advantages over studies with no controls or with nonrandomized concurrent or historical controls. These advantages include the elimination of bias in the assignment of treatments and the balancing, on average, of known and unknown baseline covariates that influence response. These advantages of randomization are particularly important for studies in which the treatments may not differ from one another by a large amount (eg, a head-to-head study of an approved LVAD with a new LVAD). However, researchers have found it difficult to recruit patients to randomized studies because the number of clinical sites that can carry out the studies is not large. Also, there is a reluctance to randomize patients when the control device is considered technologically inferior. Thus ways of improving the design of randomized trials were discussed, and the advantages and disadvantages of alternative designs were considered.

**Conclusions:** The panel concluded that designs should include a randomized component. Randomized designs might be improved by allowing the control device to be chosen before randomization, by first conducting smaller vanguard studies, and by allowing crossovers in trials with optimal medical management controls. With use of data from completed trials, other databases, and registries, alternative designs that include both a randomized component (eg, 2:1 allocation for new device versus control) and a nonrandomized component (eg, concurrent nonrandomized control, historical control, or a comprehensive cohort design) should be evaluated. This will require partnerships among academic, government, and industry scientists. (*J Cardiac Fail* 2007;13:63–74)

**Key Words:** Devices, Study design, Control groups.

*From the*<sup>1</sup>*University of Minnesota School of Public Health, Minneapolis, Minnesota;*<sup>2</sup>*Harvard Medical School and School of Public Health, Boston, Massachusetts;*<sup>3</sup>*Department of Medicine, Columbia University, New York, New York;*<sup>4</sup>*Department of Cardiology, Cleveland Clinic Foundation, Cleveland, Ohio;*<sup>5</sup>*Baylor College of Medicine, Houston, Texas and*<sup>6</sup>*Cardiovascular Division, Tufts-New England Medical Center, Boston, Massachusetts.*

Manuscript received October 26, 2006; revised manuscript received December 13, 2006; revised manuscript accepted December 15, 2006.

Reprint requests: James D. Neaton, PhD, Department of Biostatistics, University of Minnesota 2221 University Avenue SE, Room 200, Minneapolis, MN 55415.

The opinions expressed in this article are those of the authors and do not necessarily represent those of the editor or the Heart Failure Society of America. All writing group members were required to complete and submit a Faculty Disclosure Questionnaire before the workshop was held, March 30–31, 2006. The workshop was sponsored by the HFSA.

1071-9164/\$ - see front matter

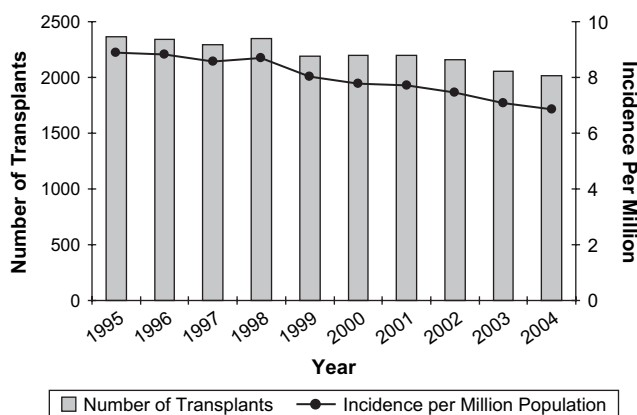
© 2007 Elsevier Inc. All rights reserved.

doi:10.1016/j.cardfail.2006.12.003

Mechanical circulatory support devices (MCSDs), such as implantable left ventricular assist devices (LVADs), were initially evaluated and approved to support patients with advanced heart failure while they awaited heart transplantation. Use as a temporizing device to “bridge” critically ill patients to cardiac transplant is increasing because of limited heart donor availability. For example, there were only 2125 transplants during 2005,<sup>1</sup> far below the estimated 20,000 to 30,000 patients each year who could benefit from this procedure. Barr et al reported that the percentage of patients awaiting heart transplantation for more than 2 years increased from 23% in 1994 to 49% in 2003.<sup>2</sup> As waiting time increases, the likelihood of clinical deterioration prior to transplant driving desirability of VAD implant, also increases. Figure 1 illustrates the decrease in the number of heart transplants that have occurred annually.<sup>3</sup>

There is also increased interest in MCSDs as “destination” therapy (ie, implant without intent to transition to alternative therapy) for patients who are not candidates for transplantation.

Since January 2002, the International Society of Heart Lung Transplantation has been collecting information from centers worldwide known to perform MCSD implantation. As of December 2004, 655 patients had received a MCSD, of whom 542 received an LVAD.<sup>4</sup> Five hundred and thirteen patients received an MCSD as a bridge to cardiac transplantation, and 78 patients received a MCSD for destination therapy, whereas the remaining patients receiving a MCSD as bridge to recovery or for reasons not specified. Therefore, it appears that MCSDs are used primarily as a bridge to transplantation and relatively few patients receive MCSDs as destination therapy. However, after low complication rates and prolonged survival rates with newer MCSDs are demonstrated, the number of patients who receive MCSDs as destination therapy is likely to increase. In the meantime, patient availability to evaluate new devices with potential technological improvements for destination therapy is limited.



**Figure 1.** Number of heart transplants and incidence per million population, 1995–2004. Source: 2005 OPTN/SRTR Annual Report, Tables 11.4 and 11.5 (available at [http://www.ustransplant.org/annual\\_reports](http://www.ustransplant.org/annual_reports)).

The Heart Failure Society of America convened a workshop with the US Food and Drug Administration (FDA) on March 30–31, 2006, to consider designs for MCSD development, including alternatives to randomized controlled trials. This report summarizes the discussion of that workshop, and is divided into 8 parts: 1) background; 2) randomized trials; 3) no controls; 4) concurrent nonrandomized controls; 5) historical controls; 6) combination of randomized and nonrandomized controls; 7) study design summary; and 8) conclusions.

## Background

### LVADs as Bridge to Transplant

There are several FDA-approved devices for use as a bridge-to-transplant: Thoratec Corporation’s HeartMate I Left Ventricular Assist System (LVAS) Series—implantable pneumatic (IP) and vented electric (VE/XVE); Thoratec ventricular assist device (paracorporeal VAD and implantable VAD); WorldHeart’s Novacor LVAS; and SynCardia Systems’ CardioWest temporary Total Artificial Heart (TAH-t).

Data from a multicenter study of 34 patients were used to describe the safety and effectiveness of the HeartMate 1000 IP LVAS.<sup>5</sup> Sixty-five percent of patients receiving the device underwent transplantation and 80% of these patients were discharged from the hospital after transplantation. The transplantation rate for 6 concurrent, nonrandomized control patients who met entry criteria but who did not receive the device was 50%; however, all 6 control patients, including the 3 who were transplanted, died within 77 days of meeting the inclusion criteria. A subsequent report described the experience of 75 IP LVAS patients and 33 nonrandomized control patients.<sup>6</sup> Fifty-three (71%) IP LVAS patients survived to transplantation compared with 12 (36%) patients in the control group. The average interval between enrollment and either transplantation or death was 76 days for those in the IP LVAS group (range: <1 to 344 days) and 12 days for the control group (range: 1 to 72 days). A 95% confidence interval (CI) for the percent of IP LVAS patients surviving to transplantation based on the binomial probability model is 61% to 81%.

Data from a study with historical controls conducted at 24 centers have been reported for the VE LVAS device.<sup>7</sup> Twenty-nine percent of VE LVAS-treated patients (82/280) died before receiving a transplant; 188 VE LVAS patients (67%) survived to transplant (95% CI based on binomial probability model: 61.6% to 72.6%); and 10 patients elected to have the device removed before transplantation. By comparison, 67% of patients in the historical control group died (32/48). One-year posttransplant survival was 84% (VE LVAS) versus 63% (controls).

The safety and effectiveness of the CardioWest TAH-t was evaluated in a nonrandomized study that used historical controls.<sup>8</sup> Eighty-one patients received the TAH-t and their survival was compared with 35 control patients who met the same entry criteria. The major outcome variable was

survival to transplant. Survival to transplant was 79% (64 of 81 patients) (95% CI 68–87%) for those who received the device and 46% for control patients. One-year survival rates were 70% for those who received the device (95% CI 63–77%) and 31% among controls.

Experience with LVADs for bridge to transplantation is much more extensive now. At a Consensus Conference in 2000 at which progress in mechanical support was reviewed, the authors noted that 300 to 400 patients each year are implanted with bridging devices. They noted that for the more than 3000 patients who have been implanted with bridging devices, 60% to 70% received a transplant.<sup>9</sup>

None of the bridging devices was evaluated in a randomized controlled trial. As with heart transplantation, obvious survival advantages in initial experiences with LVADs used as bridge devices compared with apparently comparable nonrandomized controls (historical and concurrent) made it difficult to conduct a randomized trial. Based on the data that have been amassed on bridging devices, the FDA defined an objective performance criterion for survival to transplant of 65% to 70%, which was presented at the CardioWest TAH-t Circulatory System Device Panel on March 17, 2004.<sup>10</sup> This performance goal was established in advance of the Panel meeting by the FDA and does not include confidence limits (ie, the precision with which the survival percent must be demonstrated).

### LVADs as Destination

Increasingly, LVADs are being considered as permanent implants (destination devices). Only one device is approved for destination therapy (HeartMate VE).

A trial, Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure (REMATCH), established the superiority of an LVAD (HeartMate VE) as destination therapy compared with medical therapy.<sup>11</sup> In REMATCH, survival and quality of life with the HeartMate VE LVAS (n = 68) versus optimal medical management (n = 61) were evaluated in patients who were ineligible for heart transplantation. Patients had symptoms of New York Heart Association Class IV heart failure for at least 90 days despite attempted therapy with angiotensin-converting enzyme inhibitors, diuretics, and digoxin. Patients also had to have a left ventricular ejection fraction of 25% or less, peak oxygen uptake of no more than 12 mL·min<sup>-1</sup>·kg<sup>-1</sup> or a continued need for intravenous inotropic therapy. If patients were receiving  $\beta$ -blocker therapy, it should have been used at least 60 of the 90 days before randomization. The trial was designed to continue until a total of 92 deaths occurred. At the time of the final analysis,<sup>11</sup> 41 deaths had occurred in the device group and 54 in the optimal medical management group (95 total) giving a relative risk (RR) of 0.52 (95% CI 0.34–0.78). Even with a high frequency of serious adverse events from infection, bleeding, and device malfunction, LVAD recipients experienced better quality of life than their medical therapy counterparts.<sup>11</sup> With extended follow-up, the survival rates

for LVAD versus optimal medical management patients were 52% versus 28% at 1 year and 29% versus 13% at 2 years (log-rank:  $P = .008$ ). There was a significant improvement in survival for LVAD-supported patients who enrolled during the second half of the trial as compared with the first half ( $P = .03$ ), and this improvement was not seen with medically managed patients.<sup>12</sup> This latter observation suggests that with greater clinical experience with the device, patient outcomes may improve. This point is discussed later because it has relevance to the use of historical control groups.

For another device, the Novacor LVAS, results of a nonrandomized study were reported at the American Heart Association meetings in 2005.<sup>13</sup> Investigation of Non-Transplant Eligible Patients who are Inotrope Dependent (INTrEPID) was a multicenter nonrandomized feasibility study of 55 patients (37 device and 18 control). One-year survival was 27% for those receiving the LVAD and 11% for controls who received optimal medical management. After the INTrEPID feasibility study, WorldHeart initiated a pivotal randomized study, Randomized Evaluation of the Novacor LVAS In a Non-Transplant population (RELIANT), comparing the Novacor LVAS to the HeartMate XVE. This trial is ongoing.

The FDA requires randomized trials for the approval of LVADs used as destination devices. For example, the ongoing RELIANT trial is comparing the Novacor LVAS with the HeartMate device.<sup>14</sup> The HeartMate XVE device is serving as the control and the goal of the trial is to show that the Novacor LVAS and HeartMate XVE devices are equivalent. By so doing, one could infer that the Novacor LVAS device is superior to optimal medical management based on the REMATCH study provided that the patient population studied, the end point used, and other trial conditions are the same as in the REMATCH study. The RELIANT study has a goal of 300 patients (200 on the Novacor LVAS and 100 on the HeartMate device).

Mancini and Burkhoff recently reviewed mechanical devices under development for patients with heart failure.<sup>15</sup> They called for technologic improvements in LVADs because of adverse effects related to infection, stroke, bleeding, and device failure.

There are several other ongoing trials of new LVADs, but because the patient population eligible for destination LVAD therapy is small, these trials have had difficulty enrolling patients.

### Devices Versus Drugs

There are a number of differences between device and drug development for heart failure.<sup>9,16</sup> One such difference is the possibility of progressive device modifications that are made based on experience. Such modifications may be made after completion of a trial or even during the course of a trial. Small modifications that are unlikely to diminish device efficacy or increase the likelihood of adverse effects, implemented after completion of a trial, may be

considered for incorporation into clinical use without repeating a full-scale clinical trial. When small modifications to a device are made during the course of a trial, it may be acceptable to pool results from pre- and postmodification time periods, although subgroup analyses to explore the consistency of results across the 2 groups should be performed. Of course, “small” requires definition. Unlike drugs in which “minor” chemical changes may yield unforeseen pharmacologic effects, the impact of device design changes may be more predictable based on *in vitro* bench testing and tests in animal models.

Another difference between drug and device studies is that devices have greater up-front risks. In preclinical or animal studies, these risks sometimes can be quantified. Nonetheless, studies may be necessary to rule out certain safety risks. In addition, in trials of implantable, high-risk devices versus optimal medical management, morbidity, and mortality hazards may not be proportional over the follow-up period.<sup>17</sup> This may necessitate studies with longer follow-up to establish whether benefits seen in the long-term outweigh short-term risks associated with implantation.

Other important differences are the inability to blind outcome assessments in some device studies (eg, LVADs), the importance of the skill (technical) and experience (surgical) of the surgical implanting and medical team, and the smaller patient pool available for research. Devices are easier to track than drugs after approval if appropriate registries are developed, maintained, and periodically audited. A common design issue for both device and drug studies is that optimal patient management is continually evolving. Thus historical information may not provide reliable estimates of event rates.

Given these differences and the particular difficulties facing clinical trials of devices, what options are there for conducting device trials in ways that yield valid results in an expeditious manner?

### Randomized Trials

Substantial experience in many therapeutic areas indicates that randomized trials provide the most reliable data for establishing the safety and efficacy of a new intervention. Three often cited, important advantages to carrying out a randomized study are:<sup>18</sup>

1. With appropriate blinded randomization,<sup>19</sup> bias is eliminated in the allocation of participants to treatment groups (ie, patients are not selected—consciously or unconsciously) to receive a particular treatment.
2. Known and unknown covariates are balanced on average so that no treatment has an unfair advantage.
3. Randomization ensures the validity of the statistical tests of significance that are used to compare treatments.

The first 2 of these advantages were described by Bradford Hill more than 50 years ago.<sup>20</sup> Fisher described the statistical advantages of randomization several years before that.<sup>21</sup> In a review article on randomization, Armitage noted

that Hill emphasized the first 2 points because the results would be more credible because one could not be accused by critics of having set up personally biased groups for comparison.<sup>22</sup> Hill’s emphasis on credible results is important. Investigations of new treatments are costly. At the end of the study, it is important to be able to convince others of the validity of the findings.

Another advantage of randomization is the opportunity to apply what is considered the best possible treatment for the controls (eg, optimal medical management or another device). Of course, this is a potential advantage of any prospectively defined controlled study, even if not randomized.

For head-to-head comparisons of LVADs, the first 2 advantages of randomization are important because if the incremental change in effectiveness (eg, survival) is only moderate with a new LVAD, bias and confounding arising from nonrandomized studies can dominate and, as a consequence, estimates of the new device’s effectiveness will be wrong. MacMahon and Collins make this point with respect to use of observational studies of treatments and indicate that only if the treatment effect is very large should nonrandomized studies be considered to assess their effectiveness.<sup>23</sup>

Randomized studies are also more reliable approaches for assessing the safety of devices. In the absence of randomization, it may be difficult to determine whether an excess of a particular adverse effect within a treatment group is due to the device, implantation technique, or underlying illness.

Because LVADs have been approved for both bridge and destination treatment, future trials of new LVADs in similar target populations will probably need to be head-to-head comparisons with an approved LVAD as control. These may be superiority studies or could be designed as noninferiority studies. Noninferiority trials should include a sufficient number of patients to ensure that claims of noninferiority are valid and that it is highly likely the new LVAD is superior to medical therapy alone. Considerations in the choice of the noninferiority margin to use in such trials have been described.<sup>24,25</sup> In general, the noninferiority margin must be set at a level to retain a prespecified proportion of the previously observed benefit of the control device over optimal medical management. The comparison with optimal medical management is indirect (must be imputed) because the trial does not include such an arm.<sup>26</sup> For example, in REMATCH the RR (device/optimal medical management) for mortality was 0.52 (95% CI 0.34–0.78). In a future study, comparing a new device with the HeartMate VE LVAS, if the noninferiority margin were set to ensure that there would be no more than a 50% loss of effectiveness based on this point estimate, this would result in using an upper bound of the confidence interval for the RR of (new device/VE LVAS device) of  $1.42 = 0.74$  (26% reduction)/0.52 (48% reduction). Of course the upper bound for noninferiority would be smaller (1.14) and the sample size would be much greater if the upper bound of the confidence interval (0.78) was used instead of the point

estimate. The definition for noninferiority will have to be carefully considered by the investigators and sponsor. Because the noninferiority margin is set by using historical data concerning the effectiveness of the control device, it is important that the patients to be enrolled in the trial are similar to those previously studied and that other trial conditions are the same so that the effectiveness of the control device among patients in the trial of the new device can be assumed to be similar to that in the historical trial used for planning.

Several recommendations were considered for improving the design of randomized trials for use in evaluating heart failure devices.

1. Follow prespecification of the control device to be used. This is most relevant for noninferiority studies of new devices. If there is more than 1 control device that could be used, before randomization patients and investigators could be allowed to choose which device to use as control if the control group is randomly assigned. After randomization to the control group, the prespecified device would be used. For such a trial, it would be important to carry out a prospectively defined subgroup analysis of the control device chosen although power for this comparison would be limited.
2. Conduct a smaller preliminary trial (a vanguard study) before conducting a definitive survival study. With this approach, it may be possible to use some of the data from the vanguard trial as part of the subsequent definitive trial. Possible outcomes in the vanguard study would include cardiopulmonary exercise testing, days alive outside of hospital, and general health status.
3. Allow crossovers from optimal medical management to the device under study and design the study accordingly. This is most relevant for lower risk populations in which LVADs are being considered. With this approach, when a patient in the control group (optimal medical management) meets certain prespecified criteria indicating that a LVAD should be implanted, he or she would be offered it. Sample criteria include reaching the point of inotrope dependence or perhaps a projected 1-year mortality of more than 30% to 40% per the Seattle Heart Failure score.<sup>27</sup> The questions posed by such a design is one of timing—whether immediate use of the device is preferable to deferred use of the device. Because of the crossovers, this approach would likely lead to larger sample sizes if an intention-to-treat analysis was followed for morbidity and mortality outcomes; however, the target population could be considerably broadened for study. As an intermediate endpoint, the criteria used for deferral of device implantation could be used if the criteria applied equally to those immediately receiving the device.

During the workshop several obstacles to enrolling patients in randomized trials were identified: 1) the number of clinical sites that can carry out the research is not large;

2) surgeons are reluctant to randomize patients if they feel that the control device is technologically inferior; 3) some investigators are uncomfortable with the idea of randomizing seriously ill patients; and 4) the patient population available for randomized trials is limited. A National Heart, Lung, and Blood Institute Working Group reviewed clinical research on cardiovascular surgery including research on mechanical circulatory support devices and identified similar barriers. The working group identified three common challenges in the conduct of trials of cardiac surgical procedures: 1) low enrollment of study participants; 2) difficulty in randomizing patients; and 3) crossover from the assigned arm to another arm.<sup>28</sup> This working group noted aggressive marketing of innovative procedures led to patient refusals and that bias on the part of individual surgeons and cardiologists impeded randomization. The working group recommended “sustained and sophisticated appeals to the physician community...for the best possible scientific methodologies” and more interactions between physicians and surgeons with study participants during which the rationale and goals of studies were explained.

Four alternatives to the use of randomized controls to evaluate LVADs include: 1) no controls; 2) historical controls (literature versus registry database or medical record), including use of objective performance criterion criteria; 3) concurrent, nonrandomized controls; and 4) a combination of randomized and nonrandomized controls. Many of the concerns about randomized studies are not unique to the study of LVADs.

### No Controls

Byar, in a commentary on AIDS trials, identified 5 conditions, all of which must be met to justify an uncontrolled study.<sup>29</sup>

- There is no treatment to use as a control;
- Untreated patients will experience substantial morbidity and mortality during the course of the study;
- The drug (in our situation the device) is not so toxic that the harm done could easily outweigh the benefit;
- There is sufficient experience with untreated disease (eg, historical data) to permit unambiguous evaluation of trial results;
- The scientific rationale for the treatment is sufficiently strong that a positive result would be widely believed.

He noted that the last 2 conditions are the most difficult and that is where the debate should be focused. In many situations, it may not be in the best of interest of patients, even seriously ill patients, for the researcher to carry out an uncontrolled study of a new treatment. Many promising treatments do not pan out when studied in randomized controlled trials. For example, milrinone showed promise based on a small uncontrolled study of 20 patients with advanced heart failure<sup>30</sup>; however, later, in the Prospective Randomized Milrinone Survival Evaluation (PROMISE) trial, it was found to increase mortality compared with

standard therapy of digoxin, diuretics and angiotensin-converting enzyme inhibitors.<sup>31</sup>

### Concurrent Nonrandomized Controls

Use of contemporaneous controls permits a standard protocol to be used for defining inclusion and exclusion criteria and for defining and reviewing effectiveness and safety outcomes. In this design, the control group can be specified using recently acquired evidence. Thus such designs are generally superior to use of historical controls. A disadvantage of concurrent controls compared to historical controls is that there may be many more historical controls available for comparison. This will become increasingly the case as new registries are established.

As with use of historical controls, studies with concurrent, nonrandomized controls are subject to confounding by indication.<sup>23</sup> That is, it may not be clear as to how patients were chosen for the device and control groups, and their selection may be related to the outcome of interest. This could bias estimates of the effectiveness of the device in either direction. As previously noted, in trials that compare 2 devices with one another, and for which treatment effects are not expected to be large, control of bias is a major design consideration. In such trials, there is no safe alternative to randomization.

### Historical Controls (Literature-Based, Medical Records, or Registry Databases)

Historical controls draw on data obtained or provided from an observational study or clinical trial done in the past and having well-defined diagnostic, data collection, and follow-up procedures. Meinert notes that historical controls can be defined with varying amounts of rigor.<sup>32</sup> If historical controls are used, the protocol for their definition and for comparison with the new device must be precisely described.

Major disadvantages associated with the use of historical controls have been summarized<sup>33</sup> and include the following.

1. The comparability of patients in the historical group and new device group cannot be ensured.
2. It may not be possible to reproduce precisely inclusion and exclusion criteria and, related to this, time zero for assessing time to important outcome events may not be well-defined.
3. Major outcomes for safety and effectiveness may not be identically defined.
4. It is difficult to have an independent end point review committee evaluate safety and effectiveness outcomes blinded to the treatment group.
5. Concomitant medical treatment may differ between controls and those given the new device in ways that influence prognosis.
6. Outcomes of patients may improve over time as a consequence of greater clinical experience with the device

(there was some evidence of this in the REMATCH study).<sup>12</sup>

7. If the historical database is limited, there may be considerable imprecision associated with the historical control effectiveness rates.

The first disadvantage is a major one. If there are important differences between the historical data and the new data, the wrong answer may be obtained. For example, data used to design the REMATCH study indicated that 2-year mortality for patients receiving medical management would be 75%.<sup>17</sup> In the trial, 2-year mortality for those implanted with an LVAD was 77%; the control mortality was 92%.<sup>11</sup> Use of historical data would have resulted in an inaccurate assessment of the survival benefit of the LVAD. Examples also exist in other areas of research. Byar et al describe 2 large series of patients with prostate cancer selected by the same criteria but having very different survival.<sup>34</sup> An often cited advantage of use of historical controls is that all patients receive the new treatment that is thought to be superior. As the example of milrinone indicates, this is not always the case.<sup>31</sup>

If historical controls are used, it is generally preferable that they be identified in databases at the same institutions participating in the study of the new device. It may also be preferable to use more than 1 historical control group. There are limited historical data for use of LVADs as destination. A new registry, Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS), has begun enrollment. This database may provide useful historical data in a few years.<sup>35</sup>

### Methods for Controlling Confounding Factors With Nonrandomized Controls

A number of methods have been developed to control for important prognostic variables so that causal effects can be estimated from nonrandomized studies. A popular approach is the use of propensity scores.<sup>36,37</sup> Such an approach may be applicable with nonrandomized controls that are either historical or concurrent. An advantage of propensity scores compared with usual multivariable modeling is their transparency.<sup>37</sup> With a propensity score approach, baseline covariates are used to estimate the probability of using one treatment (eg, experimental device) or the other (eg, control). These probabilities can then be used to stratify patients. Device and control outcomes can be compared within strata or adjusted for the propensity score. Of course, this approach only yields an unbiased estimate of the treatment difference if one can assume that there are no important, unobserved differences between the treatment groups (ie, that all of the important covariates and their interactions have been considered in predicting who receives the experimental device and who does not).

Even when that might be the case, propensity scores (as well as other approaches for obtaining adjusted treatment effects) may not work well when the covariate differences between those receiving the experimental device and

control are extensive. Yue describes an experience with a medical device clinical study in which a propensity score based analysis did not succeed because of excessive imbalance in observed covariates.<sup>38</sup> The example considered was a multicenter nonrandomized study with 133 patients in the investigational device arm and 67 patients in the control arm. At the completion of the propensity score modeling, the distributions of the propensity scores in the 2 treatment groups were examined and the 2 distributions barely overlapped. For example, after patients were grouped according to propensity score quintiles, it was found that the first quintile contained 58% of control patients but only 1 patient in the treatment arm, whereas the fifth quintile contained 30% of patients in the treatment arm but no control patients. Consequently, a commonly used propensity score-based analysis, treatment comparison within propensity score quintiles, was impossible to implement.

### Combination of Randomized and Nonrandomized Controls

It is hazardous to rely totally on nonrandomized controls to evaluate a new treatment. Peto et al suggested that a reasonable compromise would be to randomize in the ratio of 2:1 (eg, device:control). With this approach, two-thirds of the patients would still be available for comparison with historical controls.<sup>39</sup> This might be a reasonable approach to consider if data from a well-designed registry were available. In some situations, trial sample size might be reduced, without loss of power, by using the historical data in conjunction with the randomized controls.<sup>40</sup> Pocock proposed 6 conditions that must be met to use historical controls and randomized controls together.

1. The historical controls must have received a precisely defined standard treatment that is the same as the treatment for the randomized controls.
2. The historical controls must have been part of a recent clinical study that contained the same requirement for patient eligibility.
3. The methods of treatment evaluation must be the same.
4. The distribution of important patient characteristics in the historical group must be similar to those in the trial.
5. The previous study must have been performed in the same organization with largely the same clinical investigators (same surgeons for LVADs).
6. There must be no other indications leading one to expect differing results between the randomized and historical controls (eg, differing rates of accrual).

To eliminate temporal bias, the use of concurrent nonrandomized controls is a better strategy than the use of historical controls. Pocock's criteria can be formally translated into distributional assumptions regarding the outcomes. For example, the investigators may be willing to assume that the average outcome among the concurrent nonrandomized controls is exchangeable with the average outcome among the randomized controls. This particular assumption amounts to combining the nonrandomized control

effectiveness estimate with the estimate from the randomized controls taking into account the observed or assumed heterogeneity. Exchangeability is a strong assumption, but, if reasonable, then it would be natural to use registries and past studies to obtain information about the variability of the true effectiveness endpoint.

Rather than assuming exchangeability of estimates from nonrandomized controls with randomized controls, investigators may assume that the average outcome in the nonrandomized control group is the same as that in the randomized controls plus some bias. It may be reasonable to assume that the degree of bias in the nonrandomized control group is, on average, zero, but this could vary from patient to patient by some amount  $\sigma_b^2$ . Prior opinion regarding the value of  $\sigma_b^2$  must be specified, however. Standard procedures can be used to obtain a final estimate of the true effectiveness end point.

Another approach is to discount the influence of the nonrandomized controls through the use of a power function,  $p$ . The choice of  $p$  ranges from 0, corresponding to totally discounting the nonrandomized data, to 1, corresponding to regarding nonrandomized evidence as equal to the randomized evidence. The power function effectively adjusts the number of events in the nonrandomized control group from  $n$  to  $n \times p$ . As with other strategies that use nonrandomized information, the investigators must specify  $p$  in advance. Last, investigators may assume that use of the nonrandomized control data gives the same measure of the treatment effect as the randomized controls. This assumption amounts to assuming that patient-level data can be pooled. This is a particularly strong assumption not likely to be met in practice.

### Comprehensive Cohort Design

Machin used the approach proposed by Pocock to study the inclusion of some patients in an ongoing trial that were not randomized and received only 1 of the protocol treatments.<sup>41</sup> For example, some centers in the study might randomize their patients to the new device or control treatment; other centers would only treat patients with the new device or only with the control treatment. Machin considered a variety of assumptions concerning number enrolled and the potential for bias and concluded that there was little to be gained by the inclusion of sites that were not participating in the randomization. Machin also noted that with such a design, a center that would have otherwise participated in the randomized part of the study might use the design as an excuse to opt out of randomization.

This approach studied by Machin has been referred to as a "comprehensive cohort" design by others.<sup>42–44</sup> Some researchers proposed this design for following patients who were not enrolled in the trial, but who were eligible, to assess nonconsent bias, that bias introduced when subjects in the randomized trial differ from the eligible population with

respect to characteristics that are associated with the magnitude of the treatment effect.<sup>42,43</sup> This approach involved using a parallel nonrandomized trial that is otherwise identical to the randomized trial but consists of those patients who were eligible for randomization but instead chose their own treatments. These authors felt that such an approach could bolster the generalizability of the results from the randomized trial population.<sup>43,44</sup> An important advantage of this approach is that data collection for the randomized and non-randomized components of the study is done via the same or very similar protocols.

Related to this work, Prentice et al describe an analytic approach for the joint analysis of clinical trial and observational data under a very similar protocol.<sup>45</sup> The kind of design considered by Prentice et al might arise if a sponsor conducted both a randomized clinical trial and a non-randomized study of a new device using the same protocol and at the same sites. Prentice et al compared the effects of hormone replacement therapy on different outcomes in the randomized trial and the observational study and estimated the residual (after adjustment for important covariates) confounding in the observational study by use of an interaction term of treatment with randomization status.

The different labels attached to the interaction term by Olschewski<sup>43</sup> and Prentice<sup>45</sup> illustrates a potential problem with interpretation if there is a significant interaction. Prentice et al refer to the interaction (ie, a measure of whether the RR for treatment versus control is different in the randomized and nonrandomized cohorts) as a measure of “residual confounding,” suggesting that any difference is likely the result of unmeasured confounders in the non-randomized study.<sup>45</sup> In the article by Olschewski et al, the authors indicate that the interaction term is a measure of the internal validity of the randomized trial, suggesting that a difference between the randomized and nonrandomized studies may result from the patients in the trial not being representative of the target population.<sup>43</sup> In practice, residual confounding and poor external validity are possible reasons for an interaction, necessitating that each be considered.

If an approach like that proposed by Prentice et al or by Olschewski et al is used, the interaction test must be appropriately powered. That is, an appropriate number of patients must be included in both the trial and the nonrandomized study to reliably assess whether they can be pooled. Olschewski et al outline an approach for considering the power of the interaction test.<sup>43</sup> For example, 1 approach for a study of a destination device would be to power the interaction test to ensure that the difference is smaller than that observed in the randomized trial of the control device and optimal medical management. For an LVAD, considering the REMATCH study results, a difference of  $\ln(0.52) = -0.65$  in the log hazard ratios (or some fraction of this) between the randomized trial and the nonrandomized trial would have to be ruled out with 95% confidence.

Another approach that could be used would be to ensure that the randomized component of the study was adequately

powered to detect the hypothesized hazard ratio for the primary end point but at a higher than usual type I error rate (eg, 0.2). Such an approach has been advocated by Parides et al<sup>46</sup> for destination trials. They suggested shifting some of the burden of proof to postmarketing studies or to combining destination therapy data with bridge to transplant data. In this case, if a comprehensive cohort design were used and the randomized component yielded a *P* value of .20 and the combined randomized and nonrandomized components yielded a *P* value < .05, and the results were consistent (no interaction), the combined evidence might be acceptable.

Before considering designs such as the comprehensive cohort, data from completed trials could be used to assess how improved risk prediction models for heart failure patients perform.<sup>27</sup> It will be important in such designs to measure important prognostic factors similarly for randomized and non-randomized patients. Applications of new risk prediction models to completed device trials such as has been done with studies of bypass surgery may improve confidence in these approaches.<sup>43</sup>

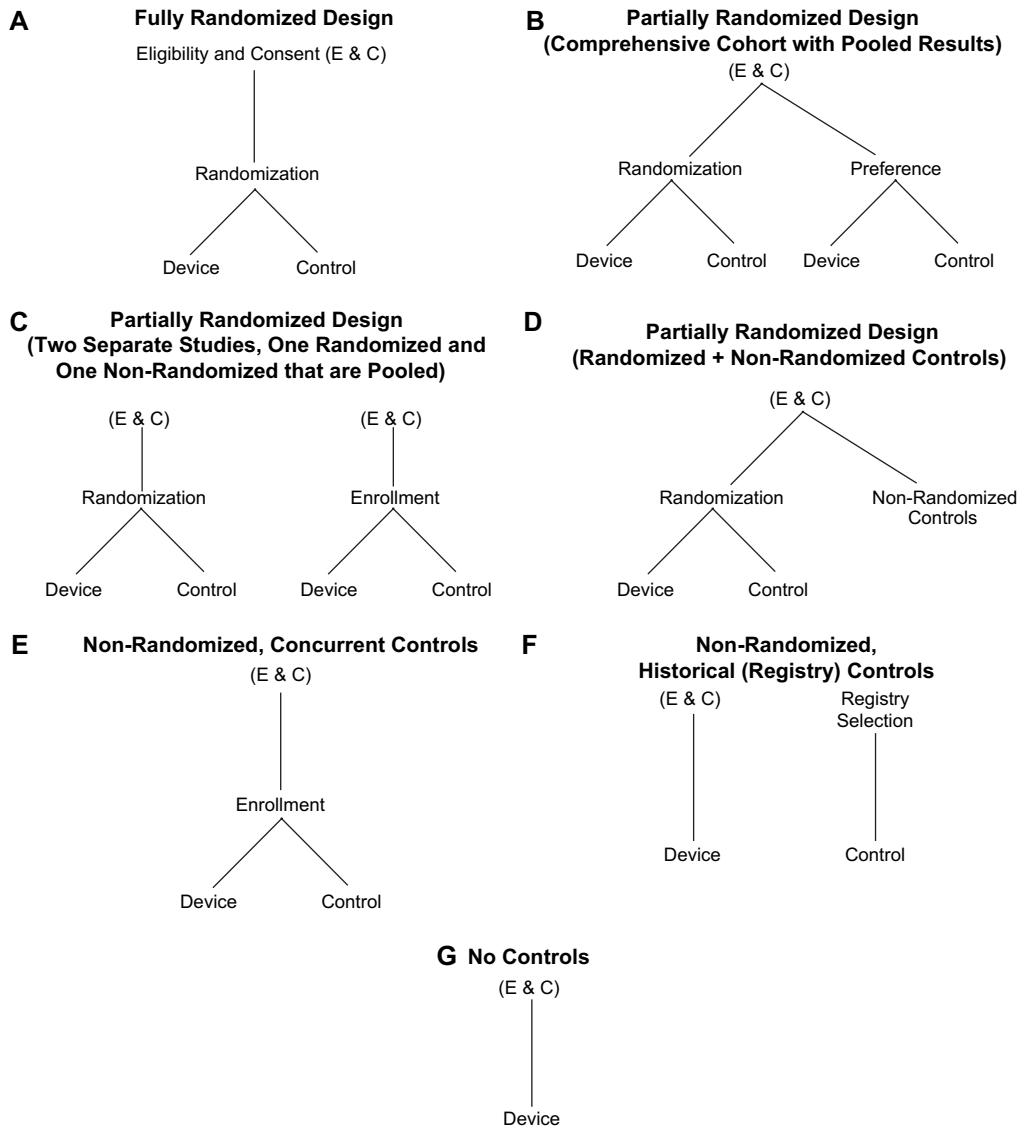
## Study Design Summary

Figure 2 depicts the various designs considered during the workshop. Table 1 lists control groups that could be used for a new MCS. The designs are presented and discussed below in an order corresponding to what designs are likely to result in the most reliable evidence.

### Fully Randomized Design

Such designs have consistently shown to provide the most reliable evidence on the safety and effectiveness of a new treatment (Fig. 2A). In studies of new MCSs for destination therapy, the control arm should be an approved MCS and the study should be designed either as a superiority or as a noninferiority study. For the latter, power must be adequate to rule out the possibility that more than 50% of the effectiveness of the already approved device is not lost. After several MCSs are approved, the control arm could be broadened to allow patients and clinicians to choose the control MCS. In such a design, the preference for the type of control MCS should be specified before randomization, so that subgroup analyses protected by randomization can be carried out by control device prespecified. A recent hypertension trial illustrates this concept.<sup>47,48</sup> In the Controlled Onset Verapamil Investigation of Cardiovascular Endpoints (ie, CONVINCENCE) study, a calcium channel blocker was compared with “standard of care,” where standard of care was the investigators’ choice of a  $\beta$ -blocker or diuretic.

For patients who are less severely ill, the control arm should be optimal medical management and the trial should be designed as a superiority study. In such a trial, provision should be made to provide control patients an MCS if



**Figure 2.** Seven illustrative designs for comparing a new mechanical circulatory support device with control. Some designs may be mixed (eg, concurrent plus historical controls).

they meet prespecified criteria that correspond to the indication for destination therapy.

### Partially Randomized Design (Comprehensive Cohort)

The original motivation for the comprehensive cohort design was to assess the external validity of data from randomized trials by following patients who chose not to be randomized but who received the study treatments (Fig. 2B). The design has the potential advantage of obtaining a more precise estimate of the treatment effect with use of nonrandomized data to supplement the randomized data when the 2 sources of data provide similar results after adjustment for confounding factors. This design has been used for coronary artery bypass surgery,<sup>43</sup> a study of adenoidectomy in children,<sup>42</sup> a trial of HIV treatments,<sup>48</sup> a study of breast cancer treatments,<sup>49</sup> and a study of

postmenopausal hormone treatment.<sup>45</sup> An advantage of this design is that data for randomized and nonrandomized patients are collected using an identical protocol by the same group of clinical sites. Ideally, the randomized component of this type of study design should be powered for selected outcomes (ie, even in the presence of a significant interaction that could indicate residual confounding), certain outcomes could be evaluated solely in the randomized part of the study design. A disadvantage of this design is that it may discourage participation in the randomized part of the study.

### Partially Randomized Design (Separate Randomized and Nonrandomized Studies)

In many cases there are data from both randomized trials and nonrandomized studies on the effectiveness of a treatment but the data are collected using separate protocols

**Table 1.** Possible Control Arms for Destination Device

Control Arm	Comment
Optimal medical management	For trials of patients in which a device is not licensed
Optimal medical management with crossover to device	For trials of patients in which a device is not licensed but who may meet criteria for a device during the study
Devices	For trials of patients in whom a device is already approved
Choice of device	For trials of patients in whom more than 1 device is already approved and are considered equivalent

(Fig. 2C). A disadvantage of pooling data from 2 separate studies is that baseline data on prognostic factors, outcome assessment, and clinical sites might vary. If sponsors use this approach (eg, a randomized trial in one set of clinical sites and a nonrandomized study in another set of sites, the data collection protocols [baseline prognostic factors and outcomes] should be identical). A protocol should be prepared for combining the 2 studies before they are begun.

#### Partially Randomized Design (Randomized and Nonrandomized Controls)

Figure 2D depicts the design described by Peto, Pocock, and Machin.<sup>39–41</sup> The nonrandomized controls ideally would be concurrent and followed according to the same protocol. If historical controls are used (eg, data from a registry), then the 6 conditions specified by Pocock need to be addressed.<sup>40</sup> The allocation ratio in the randomized component may be set to assign more patients to the new device than the control (eg, 2:1). The randomized component of the study should be powered to address selected outcomes.

#### Nonrandomized, Concurrent Controls

Figure 2E shows the design less satisfactory than a fully or partially randomized study (eg, comprehensive cohort) for the reasons stated previously. The potential for confounding by indication is a major concern and to control that potential bias randomization must be employed at least in part.

#### Nonrandomized, Historical Controls

With the availability of registries this design is likely to be frequently considered (Fig. 2F). In addition to data in registries for destination devices, data from bridge-to-transplant studies may also be of use in evaluating an LVAD to be used as destination as suggested by Parides et al.<sup>46</sup> As noted previously, there are many limitations to the sole use of historical information.<sup>33</sup> The potential for confounding by indication and for improvements in optimal medical management indicate that a better approach would be to consider a combination of historical and randomized controls (Fig. 2D).

#### No Controls

This design is almost never satisfactory (Fig. 2G). A comparison group always needs to be considered.

#### Conclusions

Several study designs are possible for the evaluation of new MCSs and all are likely to be used. In this article, we have reviewed the strengths and weaknesses of different approaches for the evaluation of MCSs.

Randomized designs have many strengths and they are likely the only reliable approach to studies in which moderate treatment differences, if any, are likely. This is true of many new treatments and particularly true for studies involving an “active” control (eg, another device that is already approved). Several recent examples illustrate the problem of relying on nonrandomized studies to assess the effectiveness of treatments.<sup>50–56</sup> However, after a new device is used in an uncontrolled study, if results are presumed to be positive, it is difficult to mount a randomized study. For this reason, several years ago, Chalmers argued that randomization should begin with the very first patient. At that point, randomization is most ethical.<sup>57</sup> Presently, randomized studies are hampered by preconceived notions about the efficacy of new devices, difficulty in consenting seriously ill patients, and by the availability of a sufficient number of patients. Although it might not be possible to randomize the first patient to a new device because of regulatory considerations, Chalmers advice should be construed as initiating randomization of patients as soon as possible after the necessary feasibility studies have been conducted.

Registry data will be available in the near future. If the registry is well-designed so that standardized information is collected on a well-defined set of patients with implants and follow-up is excellent for important outcomes, it will be invaluable for defining target populations for future research and for studying predictors of morbidity and mortality. However, caution will have to be exercised in using the patient data from the registry as historical controls for evaluating the effectiveness of new devices. Byar, MacMahon, and Pocock have reviewed the potential problems.<sup>23,33,40</sup> One approach to gaining acceptance and establishing the validity of the use of historical data for evaluating a new device is to compare data from completed randomized trials with data from the registry. For example, can the problem that could have arisen in REMATCH with use of data from patients that were used to design the trial<sup>17</sup> (historical controls) be avoided by better selection of the nonrandomized control group and by appropriate adjustment for confounding factors? Hak et al compare experimental and nonexperimental studies of influenza vaccines and provide an excellent overview of confounding by indication and of analytic methods to control it.<sup>58</sup> Similar retrospective analyses of device trials may help guide how best to control for confounding factors and how prospective studies that

use both randomized and nonrandomized controls should be designed.

Designs that include a randomized component (eg, 2:1 allocation for device versus control) and a nonrandomized component may be reasonable compromises. If comprehensive cohort designs are used or designs with both randomized and nonrandomized controls are used, it may be possible to power the randomized part of the trial for some intermediate outcomes such as cardiopulmonary exercise test results or general health status, and power the combined randomized and nonrandomized study for rarer outcomes such as serious complications and survival. For the outcomes in which both randomized and nonrandomized data are used, it will be important to be able to rule out important interactions (eg, a large difference between the randomized and nonrandomized comparison).

In summary, several scientific and practical issues must be considered in the design of MCSDs. These include study design issues such as the appropriate control group, power, and the potential for bias, and feasibility issues such as patient availability and investigator interest. Many years ago, Bradford Hill noted that “there is no need in the search for precision to throw common sense out of the window.”<sup>59</sup> That advice remains relevant today for many areas of medicine including the design of studies for new MCSDs. To move this field forward, collaborative efforts of industry, government, and academic scientists are needed to facilitate retrospective analyses of completed trials and registry data to assess the utility of different designs (eg, fully and partially randomized) and analytic methods for controlling for confounding factors. These same stakeholders should use these retrospective analyses to develop model designs for prospective studies of new devices.

### Acknowledgment

The authors wish to thank the Division of Cardiovascular Devices, US Food and Drug Administration, for their critique and comments in preparing this article. This manuscript represents the professional opinion of the authors and although the FDA provided comments, this is not an official document, agency guidance or policy of the US government, the Department of Health and Human Services, or the FDA, nor should any official endorsement be inferred.

### References

- Rosamond W, Glegal K, Friday G, et al. Heart disease and stroke statistics—2007 update. A report from the American Heart Association Statistics Committee and Stroke Statistics Committee. *Circulation* 2007;115:e69–e171.
- Barr ML, Bourge RC, Orens JB, et al. Thoracic organ transplantation in the United States, 1994–2003. *Am J Transplant* 2005;5:934–49.
- 2005 Organ Procurement and Transplantation Network (OPTN)/Scientific Registry of Transplant Recipients (SRTR) Annual Report. Available from: [http://www.ustranplant.org/annual\\_reports/](http://www.ustranplant.org/annual_reports/). Accessed February 8, 2007.
- Deng MC, Edwards LB, Hertz MI, et al. Mechanical circulatory support device database of the International Society for Heart and Lung Transplantation: Third annual report—2005. *J Heart Lung Transplant* 2005;24:1182–7.
- Frazier OH, Rose EA, Macmanus Q, et al. Multicenter clinical evaluation of the HeartMate 1000 IP left ventricular assist device. *Ann Thorac Surg* 1992;53:1080–90.
- Frazier OH, Rose EA, McCarthy P, et al. Improved mortality and rehabilitation of transplant candidates treated with a long-term implantable left ventricular assist system. *Ann Surg* 1995;3:327–38.
- Frazier OH, Rose EA, Oz MC, et al. Multicenter clinical evaluation of the HeartMate vented electric left ventricular assist system in patients awaiting heart transplantation. *J Thorac Cardiovasc Surg* 2001;122:1186–95.
- Copeland JG, Smith RG, Arabia FA, et al. Cardiac replacement with a total artificial heart as a bridge to transplantation. *N Engl J Med* 2004;351:859–67.
- Stevenson LW, Kormos RL. Consensus conference report. Mechanical cardiac support 2000: current applications and future trial design. *JACC* 2001;37:340–70.
- US Food and Drug Administration. Summary minutes: Meeting of the Circulatory System Devices Advisory Panel. Available from: [http://www.fda.gov/ohrms/dockets/ac/04/minutes/4029m1\\_summary.htm](http://www.fda.gov/ohrms/dockets/ac/04/minutes/4029m1_summary.htm). Accessed January 30, 2007.
- Rose EA, Gelljns AC, Moskowitz AJ, et al. Long-term use of a left ventricular assist device for end-stage heart failure. *N Engl J Med* 2001;345:1435–43.
- Park SJ, Tactor A, Piccioni W, et al. Left ventricular assist devices as destination therapy: a new look at survival. *J Thorac Cardiovasc Surg* 2005;129:9–17.
- Rogers JG, Pierson RN, Butler J, et al. Long-term results with Novacor® left ventricular assist device in non-transplant candidates: results of the INTREPID trial [abstract 1960]. American Heart Association, 2005. Available from: <http://www.abstractsonline.com/arch/RecordView.aspx?LookupKey=12345&recordID=20821>. Accessed February 8, 2007.
- World Heart. Clinical trials. Available from: [http://www.worldheart.com/clinical\\_information/clinical\\_trials.cfm](http://www.worldheart.com/clinical_information/clinical_trials.cfm). Accessed January 30, 2007.
- Mancini D, Burkhoff D. Mechanical device-based methods of managing and treating heart failure. *Circulation* 2005;112:438–48.
- Konstam MA, Piña I, Linefeed J, Packer M. A device is not a drug. *J Cardiac Failure* 2003;9:155–7.
- Rose EA, Moskowitz AJ, Packer M, et al. The REMATCH Trial: rationale, design, and end points. *Ann Thorac Surg* 1999;67:723–30.
- Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trial*. New York (NY): Springer-Verlag; 1998.
- Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995;274:1456–8.
- Hill AB. The clinical trial. *Br Med Bull* 1951;7:278–82.
- Fisher RA. The arrangement of field experiments. *J Ministry Agric Great Br* 1926;33:503–13.
- Armitage P. The role of randomization in clinical trials. *Stat Med* 1982;1:345–52.
- MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity. II. Observational studies. *Lancet* 2001;357:455–62.
- Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. *Ann Intern Med* 2000;133:455–70.
- Pocock SJ. The pros and cons of noninferiority trials. *Fund Clin Pharmacol* 2003;17:483–90.
- Simon R. Bayesian design and analysis of active controlled clinical trials. *Biometrics* 1999;55:484–7.
- Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model. Prediction of survival in heart failure. *Circulation* 2006;113:1424–33.

28. Baumgartner WA, Burrows S, del Nido P, et al. Recommendations of the National Heart, Lung, and Blood Institute Working Group on Future Direction in Cardiac Surgery. *Circulation* 111:3007–13.
29. Byar DM. Design considerations for AIDS trial. *J Acquire Immune Defic Syndr* 1990;2(Suppl):S16–9.
30. Baim DS, McDowell AV, Cherniles J, et al. Evaluation of a new bipyridine inotropic agent—milrinone—in patients with severe congestive heart failure. *N Engl J Med* 1983;309:748–56.
31. Packer M, Carver JR, Rodeheffer RJ, et al. Effect of milrinone on mortality in severe chronic heart failure. The PROMISE Study Research Group. *N Engl J Med* 1991;325:1468–75.
32. Meinert CL. *Clinical trials dictionary*. Baltimore (MD): Johns Hopkins Center for Clinical Trials; 1996.
33. Byar D. Why data bases should not replace randomized clinical trials. *Biometrics* 1980;36:337–42.
34. Byar D, Simon R, Friedewald WT, et al. Randomized clinical trials. Perspectives on some recent ideas. *N Engl J Med* 1976;8:74–80.
35. Personal communication, James Kirklin, MD, Principal Investigator, INTERMACS, October 2006.
36. Rosenbaum P, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
37. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127:757–63.
38. Yue LQ. Statistical and regulatory issues with the application of propensity score analysis to non-randomized medical device clinical studies. *J Bio Pharm Stat*. 2007;17:1–13.
39. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976;34:585–612.
40. Pocock SJ. The combination of randomized and historical controls in clinical trials. *J Chron Dis* 1976;29:175–88.
41. Machin D. On the possibility of incorporating patients from non-randomising centres into a randomized trial. *J Chron Dis* 1979;32:347–53.
42. Marcus SM. Assessing non-consent bias with parallel randomized and nonrandomized clinical trials. *J Clin Epidemiol* 1997;50:823–8.
43. Olschewski M, Schumacher M, Davis KB. Analysis of randomized and non-randomized patients in clinical trials using the comprehensive cohort-follow-up study design. *Cont Clin Trials* 1992;13:226–39.
44. Schumacher M, Davis K. Combining randomized and nonrandomized patients in the statistical analysis of clinical trials. In: *Recent results in cancer research*, Vol. III. Berlin: Springer-Verlag; 1988.
45. Prentice RL, Langer RD, Stefanick ML, et al. Combined analysis of Women's Health Initiative observational and clinical trial data on postmenopausal hormone treatment and cardiovascular disease. *Am J Epidemiol* 2006;163:589–99.
46. Parides MK, Moskowitz AJ, Ascheim DD, Rose EA, Gelijns AC. Progress versus precision: challenges in clinical trial design for left ventricular assist devices. *Ann Thor Surg* 2006;82:1140–6.
47. Black HR, Elliott WJ, Grandits G, et al. Principal results of the Controlled Onset Verapamil Investigation of Cardiovascular Endpoints (CONVINCE) Trial. *JAMA* 2003;289:2073–82.
48. MacArthur RD, Chen L, Mayers DL, et al. The rationale and design of the CPCRA (Terry Beirn Community Programs for Clinical Research on AIDS) 058 FIRST (Flexible Initial Retrovirus Suppressive Therapies) trial. *Control Clin Trials* 2001;22:176–90.
49. Schmoor C, Olschewski M, Schumacher M. Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies. *Stat Med* 1996;15:263–71.
50. Rimm EB, Stampfer MJ, Ascherio A, et al. Vitamin E consumption and the risk of coronary heart disease in men. *N Engl J Med* 1993;328:1450–6.
51. Stampfer MJ, Hennekens CH, Manson JE, et al. Vitamin E consumption and the risk of coronary heart disease in women. *N Engl J Med* 1993;328:1444–9.
52. The Alpha-Tocopherol Beta Carotene cancer prevention study group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other causes in male smokers. *N Engl J Med* 1994;330:1029–35.
53. Khaw KT, Bingham S, Welch A, et al. Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. *European Prospective Investigation into Cancer and Nutrition*. *Lancet* 2001;357:657–63.
54. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20536 high-risk individuals: a randomized placebo-controlled trial. *Lancet* 2002;360:23–33.
55. Stampfer M, Colditz G. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 1991;20:47–63.
56. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results of the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321–33.
57. Chalmers TC. Letters to the Editor. When should randomization begin? *Lancet* 1968;1:858.
58. Hak E, Verheij Th JM, Grobbee DE, Nichol KL, Hoes AW. Confounding by indication in non-experimental evaluation of vaccine effectiveness: the example of prevention of influenza complications. *J Epidemiol Commun Health* 2006;56:951–5.
59. Hill AB. Medical ethics and controlled trials. *BMJ* 1963;i:1043–9.