The International Journal of Biostatistics

Volume 7, *Issue* 1 2011 *Article* 29

Combining Censored and Uncensored Data in a *U*-Statistic: Design and Sample Size Implications for Cell Therapy Research

Lemuel A. Moyé, *University of Texas Health Science Center at Houston*
Dejian Lai, *University of Texas School of Public Health*
Kaiyan Jing, *University of Texas School of Public Health*
Mary Sarah Baraniuk, *University of Texas School of Public Health*
Minjung Kwak, *National Heart, Lung and Blood Institute*
Marc S. Penn, *The Cleveland Clinic Foundation*
Colon O. Wu, *National Heart, Lung and Blood Institute*

Recommended Citation:
Moyé, Lemuel A.; Lai, Dejian; Jing, Kaiyan; Baraniuk, Mary Sarah; Kwak, Minjung; Penn, Marc S.; and Wu, Colon O. (2011) "Combining Censored and Uncensored Data in a *U*-Statistic: Design and Sample Size Implications for Cell Therapy Research," *The International Journal of Biostatistics*: Vol. 7: Iss. 1, Article 29.
DOI: 10.2202/1557-4679.1286

Combining Censored and Uncensored Data in a *U*-Statistic: Design and Sample Size Implications for Cell Therapy Research

Lemuel A. Moyé, Dejian Lai, Kaiyan Jing, Mary Sarah Baraniuk, Minjung Kwak, Marc S. Penn, and Colon O. Wu

Abstract

The assumptions that anchor large clinical trials are rooted in smaller, Phase II studies. In addition to specifying the target population, intervention delivery, and patient follow-up duration, physician-scientists who design these Phase II studies must select the appropriate response variables (endpoints). However, endpoint measures can be problematic. If the endpoint assesses the change in a continuous measure over time, then the occurrence of an intervening significant clinical event (SCE), such as death, can preclude the follow-up measurement. Finally, the ideal continuous endpoint measurement may be contraindicated in a fraction of the study patients, a change that requires a less precise substitution in this subset of participants.

A score function that is based on the *U*-statistic can address these issues of 1) intercurrent SCE's and 2) response variable ascertainments that use different measurements of different precision. The scoring statistic is easy to apply, clinically relevant, and provides flexibility for the investigators' prospective design decisions. Sample size and power formulations for this statistic are provided as functions of clinical event rates and effect size estimates that are easy for investigators to identify and discuss. Examples are provided from current cardiovascular cell therapy research.

KEYWORDS: U-statistic, clinical trials, score function, stem cells

Introduction

This manuscript develops a *U*-statistic that incorporates two nettlesome but unavoidable features for collecting continuous endpoint measures in modern cardiology clinical trials; 1) the occurrence of a clinical event (e.g., death) during the trial that precludes the measurement of the endpoint at the end of the study, and 2) the requirement through clinical circumstances that a less precise determination of the endpoint (e.g., echocardiographic determination) be substituted for a most precise determination (e.g., magnetic resonance imaging).

Endpoint selection is challenging in early human cardiovascular cell therapy clinical trials. Possible choices for the endpoint are the size of the heart damaged by a heart attack, known as the infarct region [Strauer BE, et al. (2002), Abdel-Latif A, et al. (2007)], or changes in the percent of blood ejected by the left ventricle with each heart beat, or left ventricular ejection fraction (LVEF) [Assmus B, et al. (2002)]. Recent attention has focused on other measures of left ventricular dysfunction e.g., left ventricular end-diastolic volume (LVEDV) (how large the left ventricle becomes at the peak of the cardiac cycle when it is full of blood), and left ventricular end systolic volume (LVESV) (how small the ventricle is after it has ejected its blood content [Penicka M, et al. (2007)]).

Continuous response variables (endpoints) provide necessary statistical power in well designed clinical experiments. However, since continuous endpoints require measurements at both baseline and during the follow-up period, clinical events can complicate the collection of these important measures. For example, the occurrence of an intervening significant clinical event (SCE) (e.g., death) precludes the follow-up measurement, reducing the precision of the overall measure of therapy effect by reducing the number of endpoint-evaluable subjects. In addition, the observation that there may be a greater proportion of subjects with an SCE in the control group than in the treatment group introduces a new informative censoring complication to the analysis. The informative censoring approach of Follmann, Wu, et. al. [Follmann D and Wu M. (1995)] provides a useful tool for analyzing data in the presence of informative censoring; however, there is no literature on trial design and sample size computations using the informative censoring procedure.

In addition, the most precise measure of the continuous endpoint may not be in the subject's best interest. For example, while some believe that cardiac magnetic resonance (cMR) is superior to echocardiographic measures of heart function [Grothues (2002)] cMR measures cannot be obtained in everyone. Subjects who have an implantable (metallic) device e.g., a pace maker, remain contraindications to cardiac magnetic resonance imaging. If a substantial fraction of the subjects who were recruited for the study had one of these devices implanted during their follow-up procedure, they will be unable to undergo cMR

at the end of the study. A statistical test that incorporates the most precise heart function measure when available, but uses the less precise measure in its absence would permit the study to use each of these subject's data, regardless of the measurement that is indicated by the subject's condition.

This manuscript discusses the development of a *U*-statistic to permit 1) the inclusion of a dichotomous endpoint (SCE) when its occurrence precludes measuring the continuous endpoint measure, and 2) the use of the most precise endpoint information when it is available and less precise information when it is not. The statistic's development and calibration are based on commonly used clinical research measures that physician scientists understand. Examples are provided from ongoing cardiovascular research.

1. Background

Coronary artery disease (CAD) remains the single largest killer of Americans, producing myocardial infarctions and heart failure (HF).[Rosamond W, et al. (2006)] Recent research has reduced coronary heart disease mortality [Ford ES, et al.(2007)]. However CAD remains a leading cause of HF. Seven million heart attack hospitalizations in the US have generated almost 5 million subjects living with HF who face end-stage HF with its 5-year mortality of approximately 50%.[Levy D, et al.(2002), Roger VL, et al. (2004)] The large burden faced by these subjects with limited options has spurred the investigation of alternative treatments. One potential treatment strategy is the use of bone marrow–derived mononuclear cells (BMMNCs), a source of stem cells that shows promise for the treatment of these subjects.

Phase II clinical trials in cell therapy research are underway in the United States. Studies in animal models have demonstrated that heart function can be significantly improved with bone marrow-derived stem cells following experimental heart attacks induced in animals [Orlic D, et al. (2001), Kocher AA, et al. (2001), Jackson KA, et al. (2001), Yoon YS, et al. (2005)]. Although data supporting significant heart regeneration in these preclinical studies has not been uniform [Murry CE, et al. (2004), Balsam LB, et al. (2004)], it has led to a number of clinical trials testing the strategy that delivery of a subject's own (or autologous) bone marrow-derived mononuclear cells (BMMNCs) into the infarct region following AMI may improve heart function [Schächinger V, et al. (2006), Janssens S, et al. (2006), Wollert KC, et al. (2004), Lunde K, et al. (2006)].

In light of the relative paucity of mechanistic studies into important questions, such as timing of cell delivery, the National Heart, Lung, and Blood Institute (NHLBI) established the Cardiovascular Cell Therapy Research Network (CCTRN) to accelerate research into the use of cell-based therapies for the management of cardiovascular diseases. The Network is simultaneously conducting

two trials in the acute myocardial infarction environment, TIME [Traverse, et. al. (2010)] LateTIME [Traverse, et. al. (2010)] and one trial in subjects with chronic heart failure and ongoing ischemia, FOCUS [Willerson, et. al. (2010)]. Each of these trials has the characteristics of 1) having continuous measure primary endpoints, 2) enrolling patients who can have SCE's that preclude the final endpoint measure, and 3) having the most precise of the primary endpoint unavailable in a substantial fraction of the population, requiring the use of a less precise measure available.

2. Methods

This method is based on the two-sample *U*-statistic [Kowalski J and Tu XM.] (2007)], a well established, nonparametric measure of effect based on an investigator-determined scoring mechanism. Our development is modeled after the *U*-statistic's implementation to score the occurrence of a combination of two discrete endpoints in a cardiovascular clinical trial [Moyé LA, et al. (1992), Moyé LA, (1991), Penicka M, et al. (2007)]. A recent use of this statistic in medical research has been its application to multivariate ordinal data [Wittkowski KM, et al. (2004)].

In its simplest adaptation, the *U*-statistic "builds itself up" from a prospectively selected scoring procedure. Let there be n observations in the control group. Let each of the n subjects in the control group have a continuous endpoint measure x_i, $i = 1, 2, 3, ..., n$. Similarly, let the primary endpoint measure for each of the m subjects in the active group be indexed by y_j, $j = 1, 2, 3, \ldots, m$.

The *U*-statistic requires a simple scoring mechanism, denoted by $\varphi_{i,j}$. This is the assignment of a score designed in this paper based on comparing the i^{th} subject in the control group with the j^{th} subject in the active group. The score may be as simple as $\varphi_{i,j} = 1$ if $x_i > y_j$; $\varphi_{i,j} = 0$ if $x_i = y_j$; or $\varphi_{i,j} = -1$ if $x_i < y_j$. Since each of the n control group subjects will be compared to each of the m active group subjects, there are nm comparisons. The *U*-score statistic, W_e is simply the average of these nm scores,

$$W_e = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij} \tag{1}$$

The normalized statistic based on these scores for a test of the null hypothesis (H_0) of no treatment effect versus the alternative hypothesis (H_a) of a change in the distribution of the y_j's based on the treatment is

$$TS = \frac{W_e - \mathrm{E}\left[W_e \mid H_0\right]}{\sqrt{Var\left[W_e \mid H_0\right]}} \tag{2}$$

Under mild regulatory conditions and adequate sample size, we assume that (2) follows a standard normal distribution, then we can compute the sample size from (assuming $n = m$)

$$N = 2n = 2\left[\frac{Z_{1-\alpha/2}\sqrt{v_0} \quad - \quad Z_{\beta}\sqrt{v_a}}{\mathrm{E}\left[W_e \mid H_a\right] - \mathrm{E}\left[W_e \mid H_0\right]}\right]^2 \tag{3}$$

where $v_0 = Var\left[W_e \mid H_0\right]$, $v_a = Var\left[W_e \mid H_a\right]$, α is the probability of a type I error, β is the probability of a type II error, and Z_c is the c^{th} percentile value from the standard normal distribution, Alternatively, power may be computed from

$$1-\beta = 1-\Phi_Z\left[\frac{Z_{1-\alpha/2}\sqrt{Var\left(W_e \mid H_a\right)} \quad - \quad \left(\mathrm{E}\left[W_e \mid H_a\right] - \mathrm{E}\left[W_e \mid H_0\right]\right)}{\sqrt{Var\left(W_e \mid H_a\right)}}\right] \tag{4}$$

where $\Phi_Z(z)$ is the cumulative distribution function of the standard normal distribution.

However, the adoption of this statistic requires a careful justification of the scoring mechanism required for the response variables (endpoints). The setting for our evaluations is that of a randomized clinical trial with both a control and an active group. We will construct the score statistic in two cases:

Case 1. A dichotomous right censored measure combined with a single continuous response variable.

Case 2. A dichotomous right censored measure combined with two continuous response variables to be used in a hierarchy determined by the precision of the two response variables.

The mathematics of Case 1 will be developed in detail, and then applied to the Case 2 scenario, which is the scenario that we face in the CCTRN network cell therapy studies.

Case 1: A dichotomous right censored measure combined with a single continuous response variable.

The investigators' goal is to compare the change in the measure of a single continuous response variable over time in the control group to the change in that same variable in the active group. Left ventricular ejection fraction (LVEF) is an example of a commonly measured continuous endpoint. LVEF is the percent of the blood in the left ventricle ejected at each beat (for subjects without heart disease LVEF is typically larger than 80%.). To assess changes over time in a

variable such as LVEF, there should be a measurement at baseline and at the end of the study. However, the investigators recognize that this goal may not be achievable in all subjects because of the occurrence of death or another SCE. We will assume that (as is the case with LVEF) an increase in the response variable over time corresponds to improved health status.

Let r be this continuous endpoint variable. Then, for the i^{th} subject in the control group, i = 1, 2, 3, …, n, let $d_i(x) = r_{i,2}(x) - r_{i,1}(x)$ be the change in this endpoint variable over the duration of the study. Assume that $d_i(x)$ has mean $\mu_{\Delta R}(x)$ and variance σ_x^2. Analogously, let $d_j(y) = r_{j,2}(y) - r_{j,1}(y)$ be the change in the endpoint measure for the j^{th} subject in the active group, which mean $\mu_{\Delta R}(y)$ and known variance σ_y^2. Under the null hypothesis of the study, $\mu_{\Delta R}(x) = \mu_{\Delta R}(y)$. If we assume that larger values of $\mu_{\Delta R}$ correspond to improved health, then under the alternative hypothesis, the researchers expect that $\mu_{\Delta R}(x) < \mu_{\Delta R}(y)$.

However, the occurrence of a significant event (SCE) (e.g., a death, a recurrent myocardial infarction (MI), can affect the follow-up measurement of the continuous variable. The hallmark of the SCE is that 1) its occurrence during the trial either precludes the follow-up measurement (as in the case of death), or perturbs the measurement to the point that the effect of therapy can be difficult to assess (e.g., the occurrence of an intercurrent heart attack), and 2) the SCE event rates in the randomized groups may themselves be related to the therapy effect. The occurrence of an intervening SCE (itself an underpowered evaluation in a small study) reduces the power of the LVEF measure by decreasing the number of subjects who survive to have the follow-up measurement.

In this case we define the scoring mechanism $\varphi_{i,j}$ as follows:

$\varphi_{i,j} = 1$ if both the i^{th} subject if the control group and the j^{th} subject in the active group experience an SCE during the study, and the time to event for the control group subject is less than the time to event for the active group subject.

$\varphi_{i,j} = 1$ if the i^{th} subject in the control group experiences an SCE during the study and the j^{th} subject in the active group does not experience an SCE during the study.

$\varphi_{i,j} = -1$ if both the i^{th} subject in the control group and the j^{th} subject in the active group experiences an SCE during the course of the study, but the time to event for the control group subject is greater than the time to event for the active group subject.

$\varphi_{i,j} = -1$ if the i^{th} subject in the control group does not experience an SCE during the study and the j^{th} subject in the active group does experience an SCE during the study.

$\varphi_{i,j} = c$ if neither the i^{th} subject in the control group nor the j^{th} subject in the active group experience an SCE during the study, and the change in the continuous measure r for the control group subject is less than the change in continuous measure for the active group subject.

$\varphi_{i,j} = c$ If neither the i^{th} subject in the control group nor the j^{th} subject in the active group experiences an SCE during the study, and the change in the continuous measure r_i for the control group subject is greater than the change in continuous measure for the active group subject, r_j.

$\varphi_{i,j} = 0$ otherwise.

Under this mechanism, the occurrence of an early SCE (e.g., a death) in one group is considered worse than a subject survival or a later occurring SCE in the other treatment group. If both subjects in the comparison have no SCE, then the change in the response variable is compared.

With some additional notation, the assignment of this scoring system permits the computation of the mean and variance of W_e under the null and alternative hypothesis.

Notation:

Define $C_{X(i)}(E, R)$ as the endpoint status of the i^{th} subject in the control group, and $C_{Y(j)}(E, R)$ as the endpoint status of the j^{th} subject in the active group. We will use this notation to allow us to capture either 1) the time to the occurrence of an SCE if one has occurred during the course of the trial, or 2) the change in the continuous variable if an SCE has not occurred.

If an SCE has occurred for the i^{th} subject in the control group, then $C_{X(i)}(E, R) = C_{X(i)}(+, R)$, and its value is the time to the occurrence of the SCE. Since the SCE has occurred during the course of the study, then $0 \le C_{X(i)}(+, R) \le T$ where T is the maximum time a subject is to be followed in the research protocol. If an SCE has not occurred, then $C_{X(i)}(E, R) = C_{X(i)}(-, R)$, and we set $C_{X(i)}(-, R)$ to equal the change in the continuous measure. Identical notation applies to the j^{th} subject in the active group, $C_{Y(j)}(E, R)$.

For example, if in a 180 day clinical trial, the 4^{th} subject in the control group died on day 117, then $C_{X(4)}(E, R) = C_{X(4)}(+, R) = 117$, the positive sign signifying that the SCE event occurred. Alternatively, if the 5^{th} subject in the active group survived the trial and experienced a six unit increase in the continuous response variable, then $C_{Y(5)}(E, R) = C_{Y(5)}(-, R) = 6$, the minus sign in $C_{Y(5)}(-, R)$ indicating that no SCE occurred during the study.

Using this notation and letting $\mathbf{1}_{X \in A}$ be the indicator function that takes the value of 1 when x is a member of set A and 0 otherwise, we can write the score function $\varphi_{i,j}$ as

$$\begin{aligned}\varphi_{i,j} &= \mathbf{1}_{\left[C_{x(i)}[+,R] < C_{y(j)}[+,R]\right]} + \mathbf{1}_{\left[C_{x(i)}[E,R]=C_{x(i)}[+,R] \cap C_{x(i)}[E,R]=C_{y(j)}[-,R]\right]} \\ &\quad - \mathbf{1}_{\left[C_{x(i)}[+,R] > C_{y(j)}[+,R]\right]} - \mathbf{1}_{\left[C_{x(i)}[E,R]=C_{x(i)}[-,R] \cap C_{x(i)}[E,R]= C_{y(j)}[+,R]\right]} \\ &\quad + c\mathbf{1}_{\left[C_{x(i)}[-,R] < C_{y(j)}[-,R]\right]} - c\mathbf{1}_{\left[C_{x(i)}[-,R] > C_{y(j)}[-,R]\right]}.\end{aligned} \tag{5}$$

For this function we can compute its expected value under both the null ($\mathbf{E}[\varphi_{i,j} \mid H_0]$) and alternative $\mathbf{E}[\varphi_{i,,j} \mid H_a]$ hypotheses (Appendix A).

Sample size computation

Assuming a normal distribution for W_e, we compute that,

$$TS = \frac{W_e - \mathbf{E}\left[W_e \mid H_0\right]}{\sqrt{\mathbf{Var}\left[W_e|H_0\right]}} \tag{6}$$

And from consideration of the type II error, we may write,

$$Z_\beta = \frac{Z_{1-\alpha/2}\sqrt{\mathbf{Var}\left[W_e|H_0\right]} - \mathbf{E}\left[W_e \mid H_a\right]}{\sqrt{\mathbf{Var}\left[W_e|H_a\right]}} \tag{7}$$

and

$$Z_\beta\sqrt{\mathbf{Var}\left[W_e|H_a\right]} = Z_{1-\alpha/2}\sqrt{\mathbf{Var}\left[W_e|H_0\right]} - \mathbf{E}\left[W_e \mid H_a\right]. \tag{8}$$

The variance terms in (8), are computed (Appendix B) to be

$$\mathbf{Var}\left[W_e|H_0\right] = V_0 = \frac{1}{k^2}\left[\frac{A_0}{n} - \frac{B_0}{n^2}\right] \text{ and } \mathbf{Var}\left[W_e|H_a\right] = V_a = \frac{1}{k^2}\left[\frac{A_a}{n} - \frac{B_a}{n^2}\right]$$

Where the constants A_0 B_0, and A_a, and B_a are functions of expectations $\varphi_{i,,j}$ under the null and alternative hypothesis. In this scenario, an exact solution for solution for the sample size n is available. If we write $m = kn$, where k is known (for example if there are an equal number of subjects in the active group as in the control group, then $k = 1$), then substituting for the variance term, we may write

$$Z_{1-\alpha/2}\sqrt{\frac{1}{k^2}\left[\frac{A_0}{n} - \frac{B_0}{n^2}\right]} - Z_{\beta}\sqrt{\frac{1}{k^2}\left[\frac{A_a}{n} - \frac{B_a}{n^2}\right]} = \mathbf{E}\left[W_e \mid H_a\right]. \quad (9)$$

Squaring both sides, simplifying, and squaring again, with expansion and further simplification produces the quartic equation

$$a_4 n^4 + a_3 n^3 + a_2 n^2 + a_1 n + a_0 = 0 \quad (10)$$

where

$$\begin{aligned}
a_4 &= k^4 \mathbf{E}^4\left[W_e \mid H_a\right] \\
a_3 &= -2\left(Z^2_{1-\alpha/2}\mathbf{E}^2\left[W_e \mid H_a\right]k^2 A_0 + Z^2_{\beta}\mathbf{E}^2\left[W_e \mid H_a\right]k^2 A_a\right) \\
a_2 &= Z^4_{1-\alpha/2}A_0^2 + Z^4_{\beta}A_a^2 + 2Z^2_{1-\alpha/2}Z^2_{\beta}A_0^2A_a^2 + 2Z^2_{1-\alpha/2}\mathbf{E}^2\left[W_e \mid H_a\right]k^2 B_0 \\
&\quad + 2Z^2_{\beta}\mathbf{E}^2\left[W_e \mid H_a\right]k^2 B_a - 4Z^2_{1-\alpha/2}Z^2_{\beta}A_0^2A_a^2 \\
a_1 &= -2Z^4_{1-\alpha/2}A_0B_0 - 2Z^4_{\beta}A_aB_a - 2Z^2_{1-\alpha/2}Z^2_{\beta}A_0B_a - 2Z^2_{1-\alpha/2}Z^2_{\beta}A_aB_0 \\
&\quad + 4Z^2_{1-\alpha/2}Z^2_{\beta}A_0B_a + 4Z^2_{1-\alpha/2}Z^2_{\beta}A_aB_0 \\
a_0 &= Z^4_{1-\alpha/2}B_0^2 + Z^4_{\beta}B_a^2 + 2Z^2_{1-\alpha/2}Z^2_{\beta}B_0B_a - 4Z^2_{1-\alpha/2}Z^2_{\beta}B_0B_a
\end{aligned} \quad (11)$$

Power can be more directly computed as

$$1-\Phi_Z\left[\frac{Z_{1-\alpha/2}\sqrt{\frac{1}{k^2}\left[\frac{A_0}{n} - \frac{B_0}{n^2}\right]} - \mathbf{E}\left[W_e \mid H_a\right]}{\sqrt{\frac{1}{k^2}\left[\frac{A_a}{n} - \frac{B_a}{n^2}\right]}}\right] \quad (12)$$

An asymptotic solution is also available (Appendix C)

$$\begin{aligned}
\mathbf{Var}\left[W_e \mid H_0\right] &= \frac{k+1}{kn}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right] \\
\mathbf{Var}\left[W_e \mid H_a\right] &= \frac{1}{kn}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right] + k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right] - (k+1)\mathbf{E}^2\left[\varphi_{ij}\right]\right)
\end{aligned} \quad (13)$$

Now substituting equations for the $\mathbf{Var}[\varphi_{i,j} \mid \mathrm{H}_0]$ and $\mathbf{Var}[\varphi_{i,j} \mid \mathrm{H}_a]$ from (13) into (8) to compute the sample size of the trial, we write

$$\begin{aligned} & Z_{\beta}\sqrt{\frac{1}{kn}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_a\right] + k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_a\right] - (k+1)\mathbf{E}^2\left[\varphi_{ij} \mid H_a\right]\right)} \\ & = Z_{1-\alpha/2}\sqrt{\frac{k+1}{kn}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right]} - \mathbf{E}\left[W_e \mid H_a\right] \end{aligned} \tag{14}$$

Noting that the total number subjects in the study is n control group plus kn in the active group, we can write

$$N = (k+1)\left[\frac{\left(Z_{1-\alpha/2}\sqrt{\frac{k+1}{kn}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right]} - Z_{\beta}\sqrt{\frac{1}{kn}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_a\right] + k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_a\right] - (k+1)\mathbf{E}^2\left[\varphi_{ij} \mid H_a\right]\right)}\right)}{\mathbf{E}\left[W_e \mid H_a\right]}\right]^2 \tag{15}$$

Power can be expressed as

$$1-\beta = 1-\Phi_Z\left[\frac{Z_{1-\alpha/2}\sqrt{\frac{k+1}{kn}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right]} - \mathbf{E}\left[W_e \mid H_a\right]}{\sqrt{\frac{1}{k}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_a\right] + k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_a\right] - (k+1)\mathbf{E}^2\left[\varphi_{ij} \mid H_a\right]\right)}}\right] \tag{16}$$

Case 2. Competing clinical measures of different precision

This second case adds one level of complexity to Case 1. The investigators' goal is to compare the change in a continuous measure over time in the control group to that of the active group. However in this circumstance the investigators have two competing assessments of the same endpoint continuous variable. For example, in the TIME study, while cMR measure of LVEF is the most precise measure, it will need to be substituted by echocardiographic determinations of ejection fraction. The first, denoted by the continuous variable r, is the most precise but is not available in all subjects. The second, denoted by s, is less accurate, but is available for everyone.

For this case, we can modify the scoring function from Case 1 adding the following conditions

$\varphi_{i,j} = d$ if neither the i^{th} subject in the control group nor the j^{th} subject in the active group experience an SCE during the study, the change in the variable r is not available for both control and active group subjects and the change in s in control group subject is less than the change in s in the active group subject.

$\varphi_{i,j} = -d$ if neither the i^{th} subject in the control group nor the j^{th} subject in the active group experience an SCE during the study, the change in the variable r is not available for both active and control group subjects and the change in s in control group subject is greater than the change in s in the active group subject.

The computations follow the development of Case 1. $\mathbf{E}[\varphi_{i,j}\ \varphi_{i,j'}]$ now requires 22 terms (Appendix D).

3. Results

A series of evaluations of this U-statistic for clinical measures of event rates and the effect of therapy on the continuous variable were carried out (Figures 1 and 2).

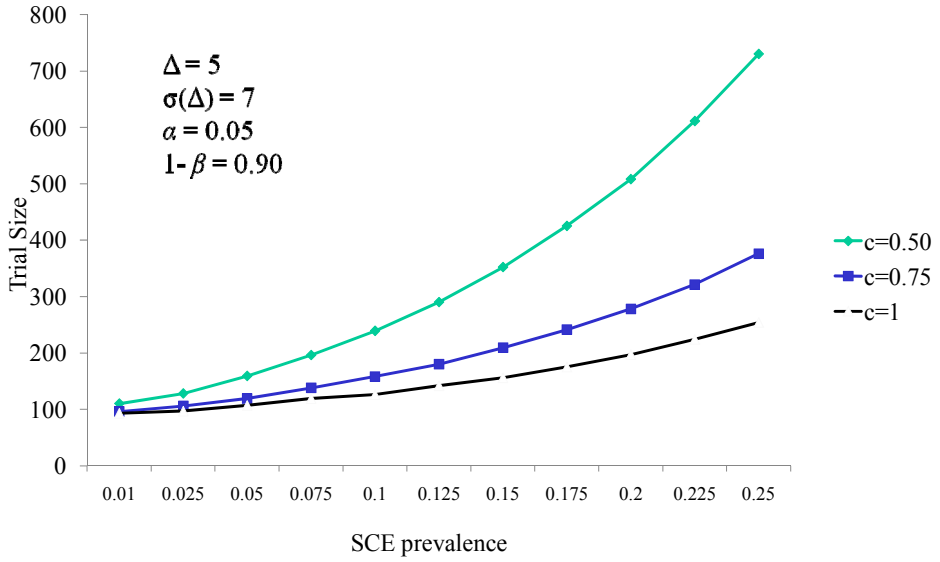

Figure 1. Continuous endpoint weight c influence on the relationship between SCE prevalence and sample size

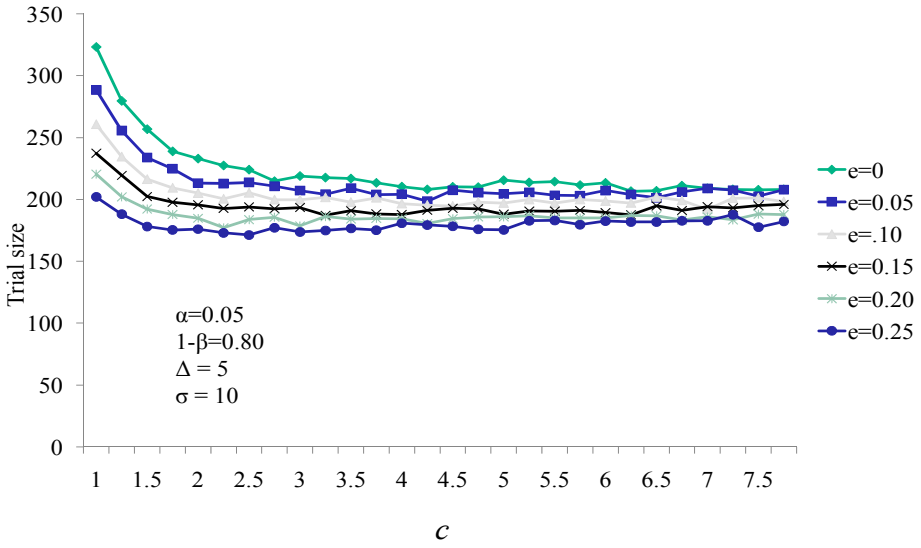

Figure 2. Influence of efficacy on the relationship between the continuous endpoint weight *c* and trial size

Figure 1 identifies the relationship between the trial size (total number of subjects in both the active and the placebo group) and the probability of a significant clinical event as a function of the effect of cell therapy on the significant clinical event rate as a function of *c*, the weight ascribed to the continuous endpoint measure in the analysis. In this circumstance we assume that the change in the response variable in the active group is five units greater than the change in the control group ($\Delta = 5$). We also assume the standard deviation of this change is 7 for each group (90% power and a type I error rate a two sided alpha of 0.05 is assumed for all analyses). In each of the curves in Figure 1, curves, the trial size is larger for larger probabilities of an SCE. Larger probabilities of an SCE increase the proportion of subjects who have no measure of the continuous endpoint that is obtained at the conclusion of the study, and larger sample sizes are required in order to main the power of the evaluation of the therapy's impact on the continuous measure.

We also note the sample size increases as the value of *c* decreases. The value of *c* is the relative weight in the scoring system. As *c* decreases the impact of a nonzero comparison between the active and control group measures has less weight than that of the comparison of SCE timings. This diminished weight for comparison generates the need for more continuous measure comparisons in the cohort, thereby increasing the sample size. Figure 2 demonstrates the same effect

of decreasing sample size for larger values of the continuous weighting function *c*. Here *e* represents the effect of the therapy on the SCE rate (represented as the percent reduction in the SCE control group rate experienced by the active group subjects). Sample size decreased as *e* increased. Note that for all values of efficacy evaluated, the sample size stabilized for values of *c* greater than 3.

The well established relationship between sample size and treatment effect (Δ) are demonstrated in Figure 3. In the paradigm of combining a continuous and a dichotomous endpoint, the sample size decreases as the effect size increases, and increases as the treatment standard deviation of the difference (σ_Δ) increases. Analogously, it is well accepted that when a dichotomous measure is used as a response variable in a clinical trial, the trial size increases as the prevalence of the dichotomous variable increases and decreases with increasing efficacy of treatment against that response variable. This is demonstrated in Figure 4. Thus, the score statistic is a function of the effect of the cell therapy on the continuous measure, as reflected by Δ and σ(Δ), and also by the efficacy of the therapy on the SCE rate as well, *e*. Figure 4 demonstrates that, while larger values of the probability of an SCE still produce larger trial sizes, the efficacy of the therapy on the SCE rate moderates this relationship.

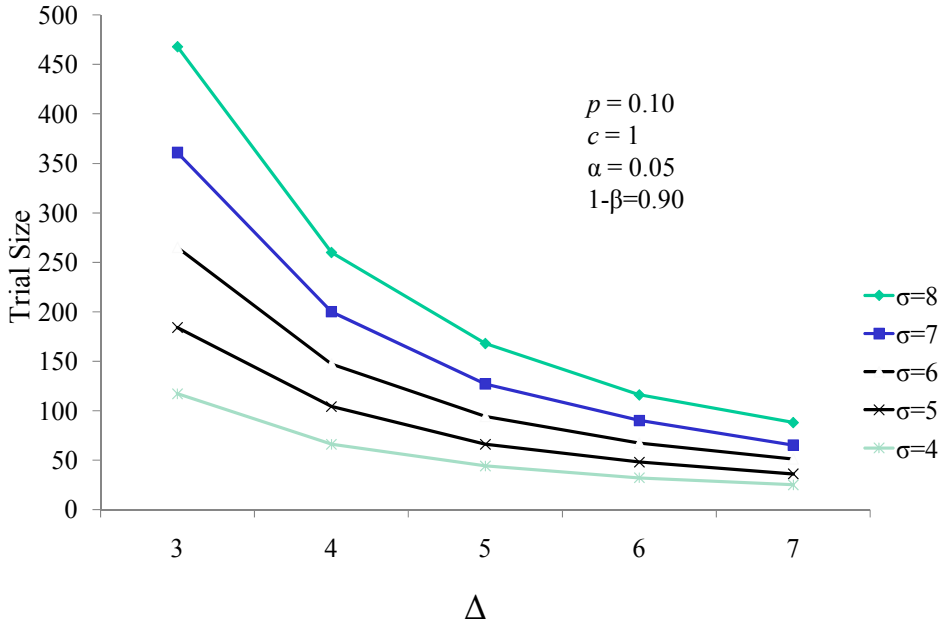

Figure 3. Trial size as a function of standard deviation (σ) and effect size (Δ) of the response variable.

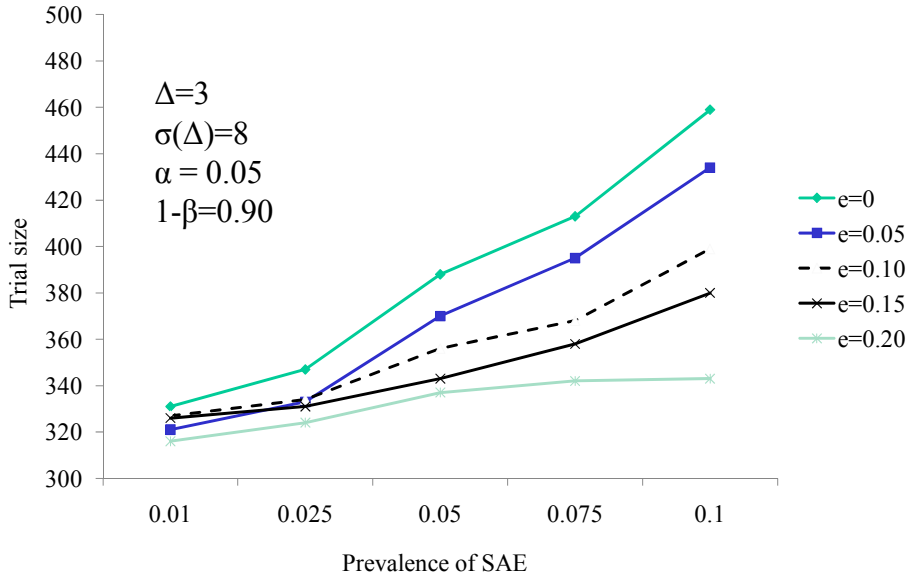

Figure 4. Trial size as a function of the SCE prevalence (*p*) and the intervention's efficacy on prevalence (*e*)

Case 2.

In this research scenario there are two continuous measures, each with weights *c* and *d*. As both *c* and *d* increase, the weight of each continuous endpoint increases, and the sample size decreases. However, the larger values of *c* and *d* have diminishing impact on the sample size.

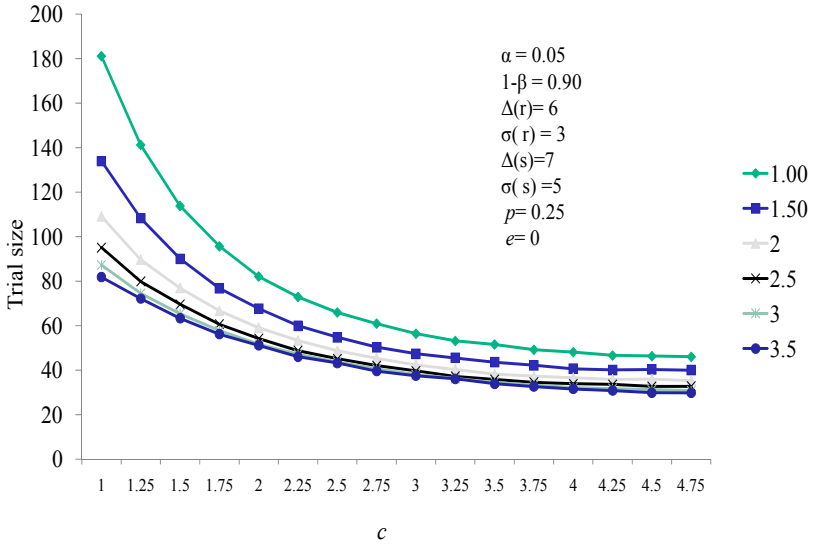

Figure 5. Relationship between sample sizes and weighting factors *c* and *d*.

Example. In a study in heart failure [Pfeffer MA, et. al. 1992] LVEF was expected to increase slightly in the control group, and anticipated to increase to a greater degree in the active group. However subjects with heart failure die or have clinical events precluding the assessment of this measure. In addition LVEF was measured in two ways. The premier measurement was through a radionuclide ventriculogram (RVG). RVG-LVEF measures were the most accurate; however, the requirement of radiation exposure limited the utility of this procedure. An alternative was to use echocardiography to obtain the LVEF. Echo-based LVEF, new at the time, was safe but less precise. The presence of two continuous measures with one preferred over the other, in addition to the occurrence of SCE's is a circumstance in which the proposed sample size computation was designed.

From a sample of data from this study (Table 2), we use the percent of significant clinical events, in concert with the data for expected changes in RVG-LVEF and echo muscle mass to compute $\mathbf{E}\left[\phi_{ij} \mid \mathrm{H}_a\right]$. The terms for the variance $\mathbf{E}\left[\phi_{ij}\phi_{i'j} \mid \mathrm{H}_0\right]$, $\mathbf{E}\left[\phi_{ij}\phi_{i'j} \mid \mathrm{H}_a\right]$, and $\mathbf{E}\left[\phi_{ij}\phi_{ij'} \mid \mathrm{H}_a\right]$ are computed from the eighteen terms in Table 1 each individual expression based on the data from Table 2 reflecting the expected changes in both continuous endpoint measures during the course of the study in both the control and active groups. Using equations (10) and (11) , an exact sample size of 540 (active plus control group subjects) was identified. The quantiles or probabilities of a bivariate (or general *k*-variate) nor-

mal distribution for under the null and alternative hypotheses are numerically available, e.g in the R library multcomp [Hothorn T, et. al., 2011], mvtnorm [Genz et. al., 2011] or as an add-in package in Excel. Using equation (15) the asymptotic solution of 536 is identified with little difference seen between the asymptotic and exact solutions.

Table 1. Values of $\varphi_{ij}\varphi_{ij'}$ used in computing $\boldsymbol{E}\left[\varphi_{ij}\varphi_{ij'}\right]$ for Case 1.

	$C_{x(i)}(+,R) < C_{Y(j)}(+,R)$	$C_{x(i)}(+,R) < C_{Y(j)}(-,R)$	$C_{x(i)}(+,R) > C_{Y(j)}(+,R)$	$C_{x(i)}(-,R) > C_{Y(j)}(+,R)$	$C_{x(i)}(-,R) < C_{Y(j)}(-,R)$	$C_{x(i)}(-,R) > C_{Y(j)}(-,R)$
$C_{x(i)}(+,R) < C_{Y(j)}(+,R)$	1	1	−1			
$C_{x(i)}(+,R) < C_{Y(j)}(-,R)$	1	1	−1			
$C_{x(i)}(+,R) > C_{Y(j)}(+,R)$	−1	−1	1			
$C_{x(i)}(-,R) > C_{Y(j)}(+,R)$				1	$-c$	c
$C_{x(i)}(-,R) < C_{Y(j)}(-,R)$				$-c$	c^2	$-c^2$
$C_{x(i)}(-,R) > C_{Y(j)}(-,R)$				c	$-c^2$	c^2

Table 2 LVEF Heart Failure Example

		Null		Alternative	
		Mean Change	SD	Mean Change	SD
RVG-LVEF	control	1.31	9.6	1.31	9.6
	active	1.31	10.6	6	10.6
Echo-LVEF	control	-1.5	7.3	-1.5	7.3
	active	-1.5	6.8	0	6.8
	SCE %		0.24		
	% missing		0.07		
	Type I error (two sided)		0.05		
	Power		0.8		
	c		1		
	d		1		
	k (active to control)		1		
	Sample size (asymptotic)		**536**		
	Sample size (exact)		**540**		

4. Discussion

This manuscript demonstrates a method to combine prospectively declared mortality measures with continuous endpoints that maintain the clinical hierarchy of the occurrence of events, using information from the continuous effect size. It is based on work involving a less complex function in a clinical trial with two dichotomous endpoints [Pfeffer MA et. al., 1992]. No imputation is required, and the difficulties with worse rank assignments to missing continuous endpoint data are avoided.

The score function used is very specific to the problems provided here and all mathematical derivations are tied to the score function. Although the general formulations for the mean and variance of W_e will be the same for any score function, these computations will reflect the score function used. As an example, there is a common index used in research known as a heart failure index. It is nonparametric composed of clinical assessments (medication changes) hospitalizations, and the occurrence of death. The U statistic procedure proposed here would be of value in the heart failure score scenario; however, in that case, the score function would be based on pair wise changing in heart function scores between active and control group patients.

The problem posed in this manuscript is distinct from the multiple endpoint scenario where investigators choose from among several different endpoint

measures. This latter dilemma has been central to clinical trial interpretation, and many important contributions to the literature have addressed this complex challenge.. Clinical trialists commonly face the issue of endpoint selection and cannot resolve it in the favor of one or the other. Clinical trials can have endpoints with no priority among their selection at all [Tilley BC, et al. (1996), National Institute of Neurological Disorders and Stroke rt- PA Stroke Study Group. (1995)]. O'Brien examined the role of a rank sum test in 34 endpoint setting [O'Brien PC. (1984)]. Lachin suggested the use of imputation, assigning a worst rank score to those subjects who are missing the continuous endpoint measure due to a mortal event [Lachin J. (1999)]. The use of Area Under the Curve (AUC) data has been particularly helpful in tumor models [Wu J. (2010)]. In addition several workers [Tan M., (2002)] proposed a heuristic test and a Bayesian procedure for the analysis of two small-sample parametric inference procedures for incomplete longitudinal data with truncation and informative censoring arising in cancer therapy development. Other authors have proposed alternative solutions [O'Brien PC. (1984), Tang DI, et al. (1989a), Tang DI, et al. (1989b), Tang DI, et al. (1993), Tang DI and Getter NL. (1999)].

Of particular use are the weighting values c and d. The investigator has complete control over the values of these weights but must choose them carefully. For example, in clinical trials in which the predominant response value is a dichotomous random variable, weights in the range $0 \leq d \leq c \leq 1$ are attractive. Since the dichotomous variable occurs so frequently (e.g., mortality) and is only replaced by the continuous measures in the cases where vital status information is not available, discounting the contribution of the continuous variables is appropriate. However, in studies, such as smaller cell therapy studies where the response variable is continuous, and relatively small numbers of subjects have SCE's, a greater weight for the continuous measure can be justified. In our cell therapy studies, the value of $c = 4$ is appropriate. We advocate selecting d such that $0 \leq d < c$ since the less precise measure should have less influence on the test statistic than the more precise one. However, these values must be chosen before any endpoint analysis takes place to avoid selections that are biased by the investigators observations of the values of the final response variables. The U statistic itself is close to normality for small samples [Mann, Whitney (1947)].

The analysis that we propose creates a new endpoint, and that new endpoint as defined by the score function is more complex. However that change does not overly complicate the interpretation of the result. For example in this case, the score function generates an analysis of either 1) an improvement in heart function or 2) longer survival without a death or heart attack. We believe that this new endpoint is understandable to a research, regulatory and clinical community already comfortable with complex endpoints e.g., fatal and nonfatal heart attacks, or fatal and nonfatal strokes. In addition, the fact that the score function never

makes cross modality comparisons, comparing only MRI to MRI changes, or echo to echo changes, helps to keep the interpretation clear.

Complications of the application of this procedure include the observation that the event rates of the significant clinical event and the standard deviation of the continuous measure differs from that assumed during the study's design phase. In addition, interim review of the statistics by Data Safety and Monitoring Boards introduces new complexities. Neither of these is assessed specifically in this manuscript.

Three clinical trials in the NHLBI sponsored Cardiovascular Cell Therapy Research Network (CCTRN) are currently underway in which we will assess the utility of this approach.

Appendix A. Computing $\boldsymbol{E}\left[\varphi_{i,j}\right]$

The notation from the previous section permits us to write the expected value of $\varphi_{i,j}$ under the hypothesis H_k, $k = 0$ for the null hypothesis, and $k = a$ for the alternative hypothesis. We assume throughout this manuscript that the time to an SCE and the continuous measure are independent. From equation (5), we may write

$$\begin{aligned}\boldsymbol{E}\left[\varphi_{ij} \mid H_k\right] &= \boldsymbol{P}\left[C_{x(i)}[+,R] < C_{y(j)}[+,R] \mid H_k\right] \\ &+ \boldsymbol{P}\left[C_{x(i)}[E,R] = C_{x(i)}[+,R] \cap C_{x(i)}[E,R] = C_{y(j)}[-,R] \mid H_k\right] \\ &- \boldsymbol{P}\left[C_{x(i)}[+,R] > C_{y(j)}[+,R] \mid H_k\right] - \boldsymbol{P}\left[C_{x(i)}[E,R] = C_{x(i)}[-,R] \cap C_{x(i)}[E,R] = C_{y(j)}[+,R] \mid H_k\right] \\ &+ c\boldsymbol{P}\left[C_{x(i)}[-,R] < C_{y(j)}[-,R] \mid H_k\right] - c\boldsymbol{P}\left[C_{x(i)}[-,R] > C_{y(j)}[-,R] \mid H_k\right].\end{aligned} \tag{17}$$

This computation is straightforward when the probability distributions of 1) the occurrence of SCE's and 2) the probability distribution of the continuous response variable *r* are known. For example, assume the time to an SCE follows an exponential distribution with parameter λ_x in the control group and λ_y in the active group. Also assume that the change in the continuous measure *r* follows a normal distribution with mean as before $\mu_{\Delta R}(x)$ and standard deviation $\sigma_{\Delta R}(x)$ in the control group, and analogously mean $\mu_{\Delta R}(y)$, and standard deviation $\sigma_{\Delta R}(y)$ in the active group. The first term on the right hand side of equation (17) is

$$\mathbf{P}\left[C_{x(i)}[+,R] < C_{Y(j)}[+,R]\right] = \int_0^T \lambda_y e^{-\lambda_y y} \int_0^y \lambda_x e^{-\lambda_x x} dx dy = 1 - e^{-\lambda_y T} - \frac{\lambda_y}{\lambda_x + \lambda_y}\left(1 - e^{-(\lambda_x + \lambda_y)T}\right)$$

As another example, the last term on the right hand side of equation (17) is

$$\mathbf{P}\left[C_{x(i)}[-,R] < C_{Y(j)}[-,R] \mid H_0\right] = e^{-(\lambda_x+\lambda_y)T}\Phi_Z\left[\frac{-(\mu_{\Delta R}(x)-\mu_{\Delta R}(y))}{\sqrt{\sigma^2_{\Delta R}(x)+\sigma^2_{\Delta R}(y)}}\right] \qquad (18)$$

Since the null hypothesis assumes no treatment effect, we let $\lambda_x = \lambda_y$, and $\mu_{\Delta R}(x) = \mu_{\Delta R}(y)$, to see that

$$\mathbf{E}\left[\varphi_{ij} \mid H_0\right] = 0. \qquad (19)$$

Under the alternative hypothesis H_a of a treatment effect, then either $\lambda_x \neq \lambda_y$, and/or $\mu_{\Delta R}(x) \neq \mu_{\Delta R}(y)$, permitting us to write,

$$\begin{aligned}\mathbf{E}\left[\varphi_{ij} \mid H_a\right] = {} & 1 - e^{-\lambda_y T} - \frac{\lambda_y}{\lambda_x+\lambda_y}\left(1-e^{-(\lambda_x+\lambda_y)T}\right) + e^{-\lambda_y T}\left(1-e^{-\lambda_x T}\right) \\ & - \left[1 - e^{-\lambda_x T} - \frac{\lambda_x}{\lambda_x+\lambda_y}\left(1-e^{-(\lambda_x+\lambda_y)T}\right) + e^{-\lambda_x T}\left(1-e^{-\lambda_y T}\right)\right] \\ & + ce^{-(\lambda_x+\lambda_y)T}\Phi_Z\left[\frac{-(\mu_{\Delta R}(x)-\mu_{\Delta R}(y))}{\sqrt{\sigma^2_{\Delta R_X}+\sigma^2_{\Delta R_Y}}}\right] - ce^{-(\lambda_x+\lambda_y)T}\left(1-\Phi_Z\left[\frac{-(\mu_{\Delta R}(x)-\mu_{\Delta R}(y))}{\sqrt{\sigma^2_{\Delta R_X}+\sigma^2_{\Delta R_Y}}}\right]\right)\end{aligned} \qquad (20)$$

Appendix B. Variance Computation

The computation of the variance of W_e, while somewhat more complicated than the mean, is executable. Assume n subjects in the control group and m subjects in the active group, then,

$$\mathbf{Var}\left[W_e\right] = \mathbf{Var}\left[\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right] = \frac{1}{(nm)^2}\mathbf{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right]. \qquad (21)$$

Further,

$$\mathbf{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right] = \mathbf{E}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right)^2\right] - \left(\mathbf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right]\right)^2. \qquad (22)$$

The last term on the right of the second line of (22) is easily evaluated.

$$\left(\mathbf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right]\right)^2 = \left((nm)^2\,\mathbf{E}\left[\varphi_{ij}\right]\right)^2 = (nm)^2\,\mathbf{E}^2\left[\varphi_{ij}\right].$$

where the expected value of $\mathbf{E}\left[\varphi_{ij}\right]$ has already been computed both under the null (19) and alternative (20) hypotheses.

To evaluate $\mathbf{E}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right)^2\right]$ from (22), we rewrite $\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right)^2$ as

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq j'}}^{m}\varphi_{ij}\varphi_{ij'} + \sum_{\substack{i=1\\ i\neq i'}}^{n}\sum_{j=1}^{m}\varphi_{ij}\varphi_{i'j} + \sum_{\substack{i=1\\ i\neq i'}}^{n}\sum_{\substack{j=1\\ j\neq j'}}^{m}\varphi_{ij}\varphi_{i'j'}$$

This helpful simplification is due to Gehan [Gehan EA. (1965)]. We may now pass the expectation argument through the preceding equation to find

$$\begin{aligned}\mathbf{E}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right)^2\right] &= \mathbf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}^2\right] + \mathbf{E}\left[\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i\neq i'}}^{n}\sum_{j=1}^{m}\varphi_{ij}\varphi_{i'j}\right] \\ &\quad + \mathbf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{\substack{j'=1\\ j\neq j'}}^{m}\varphi_{ij}\varphi_{ij'}\right] + \mathbf{E}\left[\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i\neq i'}}^{n}\sum_{j=1}^{m}\sum_{\substack{j'=1\\ j\neq j'}}^{m}\varphi_{ij}\varphi_{i'j'}\right]\end{aligned} \tag{23}$$

and evaluating term by term we see

$$\mathbf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}^{2}\right]=(nm)^{2}\,\mathbf{E}\left[\varphi_{ij}^{2}\right]$$

$$\mathbf{E}\left[\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n}\sum_{j=1}^{m}\varphi_{ij}\varphi_{i'j}\right]=nm(n-1)\,\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right]$$

$$\mathbf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{\substack{j=1\\ j\neq j'}}^{m}\varphi_{ij}\varphi_{ij'}\right]=nm(m-1)\,\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]$$

$$\mathbf{E}\left[\sum_{i=1}^{n}\sum_{\substack{i'=1\\ i'\neq i}}^{n}\sum_{j=1}^{m}\sum_{\substack{j=1\\ j\neq j'}}^{m}\varphi_{ij}\varphi_{i'j'}\right]=nm(n-1)(m-1)\,\mathbf{E}\left[\varphi_{ij}\varphi_{i'j'}\right]$$

$\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]$ is the expected value of the product of the scoring function between 1) the i^{th} control group subject and the j^{th} active group subject, and 2) the same i^{th} control group subject but a different j'^{th} active group subject where $j \neq j'$. $\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right]$ is an analogous computation involving the i^{th} and i'^{th} in the control group and the j^{th} subject in the active group. We may now rewrite (23) as

$$\begin{aligned}\mathbf{E}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right)^{2}\right]&=(nm)^{2}\,\mathbf{E}\left[\varphi_{ij}^{2}\right]+nm(n-1)\,\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right]\\&\quad+nm(m-1)\,\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]+nm(m-1)(n-1)^{2}\,\mathbf{E}^{2}\left[\varphi_{ij}\right]\end{aligned}\tag{24}$$

Thus

$$\begin{aligned}\mathbf{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij}\right]&=(nm)^{2}\,\mathbf{E}\left[\varphi_{ij}^{2}\right]+nm(n-1)\,\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right]+nm(m-1)\,\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]\\&\quad+nm(m-1)(n-1)^{2}\,\mathbf{E}^{2}\left[\varphi_{ij}\right]-(nm)^{2}\,\mathbf{E}^{2}\left[\varphi_{ij}\right]\end{aligned}$$

Rather than solve for two unknowns, m and n ,we let $n = kn$ and where k is known (for example, if there are twice as many subjects in the active group than in the control group then $k = 2$). We may then write

$$\mathbf{Var}\left[\sum_{i=1}^{kn}\sum_{j=1}^{n}\varphi_{ij}\right] = kn^2\mathbf{E}\left[\varphi_{ij}^2\right] + kn^2(n-1)\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right] + kn^2(kn-1)\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right] + kn^2(kn-1)(n-1)^2\mathbf{E}^2\left[\varphi_{ij}\right] - k^2n^4\mathbf{E}^2\left[\varphi_{ij}\right]$$

Further simplication reveals

$$\begin{aligned}\mathbf{Var}[W_e] &= \frac{1}{k^2n^4}\mathbf{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{kn}\varphi_{ij}\right] \\ &= \frac{n^3}{k^2n^4}\,k\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right] + k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right] - (k+1)\mathbf{E}^2\left[\varphi_{ij}\right]\right) \\ &\quad - \frac{n^2}{k^2n^4}\left(k\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right] + \mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right] - \mathbf{E}\left[\varphi_{ij}^2\right] - \mathbf{E}^2\left[\varphi_{ij}\right]\right)\right) \\ &= \frac{1}{k^2}\left[\frac{A}{n} - \frac{B}{n^2}\right]\end{aligned} \tag{25}$$

Since the variance can be computed under both the null and alternative hypothesis, we may write

$$\mathbf{Var}[W_e|H_0] = V_0 = \frac{1}{k^2}\left[\frac{A_0}{n} - \frac{B_0}{n^2}\right] \text{ and } \mathbf{Var}[W_e|H_a] = V_a = \frac{1}{k^2}\left[\frac{A_a}{n} - \frac{B_a}{n^2}\right]$$

where A_0 and B_0 are computed under the null hypothesis and A_a and B_a are computed under the alternative.

Appendix C. The Asymptotic approach for $\mathbf{Var}[W_e]$ **:**

Working from equation (25)

$$\begin{aligned}\mathbf{Var}[W_e] &= \frac{1}{k^2n^4}\mathbf{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{kn}\varphi_{ij}\right] \\ &= \frac{n^3}{kn^4}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right] + k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right] - (k+1)\mathbf{E}^2\left[\varphi_{ij}\right]\right) \\ &\quad - \frac{n^2}{kn^4}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right] + \mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right] - \mathbf{E}\left[\varphi_{ij}^2\right] - \mathbf{E}^2\left[\varphi_{ij}\right]\right)\end{aligned}$$

Ignoring terms on the order of n^{-2}, we have

$$\mathbf{Var}\left[W_e\right]=\frac{1}{kn}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right]+k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]-(k+1)\mathbf{E}^2\left[\varphi_{ij}\right]\right), \tag{26}$$

Equation (26) is the variance of W_e under the alternative hypothesis, $\mathbf{Var}\left[W_e \mid H_a\right]$. Under the null hypothesis, we assume $\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right]=\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]$, and $\mathbf{E}\left[\varphi_{ij}\right]=0$, producing $\mathbf{Var}\left[W_e \mid H_0\right]=\frac{k+1}{kn}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right]$. We therefore may write

$$\begin{aligned}\mathbf{Var}\left[W_e \mid H_0\right]&=\frac{k+1}{kn}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right]\\ \mathbf{Var}\left[W_e \mid H_a\right]&=\frac{1}{kn}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right]+k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]-(k+1)\mathbf{E}^2\left[\varphi_{ij}\right]\right)\end{aligned} \tag{27}$$

Now substituting equations for the $\mathbf{Var}\left[W_e \mid H_0\right]$ and $\mathbf{Var}\left[W_e \mid H_a\right]$ from (13) into (8) to compute the sample size of the trial, we write

$$\begin{aligned}&Z_\beta\sqrt{\frac{1}{kn}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_a\right]+k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_a\right]-(k+1)\mathbf{E}^2\left[\varphi_{ij} \mid H_a\right]\right)}\\ &=Z_{1-\alpha/2}\sqrt{\frac{k+1}{kn}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right]}-\mathbf{E}\left[W_e \mid H_a\right]\end{aligned} \tag{28}$$

Noting that the total number subjects in the study is *n* control group plus *kn* in the active group, we can write

$$N=(k+1)\left[\frac{\left(Z_{1-\alpha/2}\sqrt{\frac{k+1}{k}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right]}-Z_\beta\sqrt{\frac{1}{k}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_a\right]+k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_a\right]-(k+1)\mathbf{E}^2\left[\varphi_{ij} \mid H_a\right]\right)}\right)}{\mathbf{E}\left[W_e \mid H_a\right]}\right]^2 \tag{29}$$

Power can be expressed as

$$1-\beta = 1-\Phi_Z\left[\frac{Z_{1-\alpha/2}\sqrt{\frac{k+1}{kn}\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\,|\,H_0\right]} \;-\; \mathbf{E}\left[W_e\,|\,H_a\right]}{\sqrt{\frac{1}{kn}\left(\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\,|\,H_a\right] + k\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\,|\,H_a\right] - (k+1)\mathbf{E}^2\left[\varphi_{ij}\,|\,H_a\right]\right)}}\right] \quad (30)$$

Appendix D Structure for $\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]$

The structure for $\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]$ can be identified and tabulated (Table 1) revealing 18 terms, each of which is evaluated under the null and alternative hypothesis. For example one of the terms may be written as,

$$\mathbf{P}\left[C_{X(i)}(+,R) < C_{Y(j)}(+,R) \cap C_{Y(j')}(E,R) = C_{Y(j')}(-,R)\right] \;=\; \mathbf{P}\left[U < V < T < W\right]. \quad (31)$$

Here U follows an exponential distribution with parameter λ_x, T is the duration of the study, and V and W are i.i.d. exponentially distributed random variables with parameter λ_y. Note that the final expressions for the expectation are in terms of the parameters λ_x, λ_y, $\mu_{\Delta R}(x)$, $\mu_{\Delta R}(y)$, $\sigma^2_{\Delta R}(x)$, and $\sigma^2_{\Delta R}(y)$. These are available from the clinical scientists. Thus (31) may be written as

$$\begin{aligned}&\mathbf{P}\left[C_{X(i)}(+,R) < C_{Y(j)}(+,R) \cap C_{Y(j')}(E,R) = C_{Y(j')}(-,R)\right] \\ &= e^{-\lambda_y T}\left[\left(1-e^{-\lambda_y T}\right) - \frac{\lambda_y}{\lambda_x+\lambda_y}\left(1-e^{-(\lambda_x+\lambda_y)T}\right)\right].\end{aligned} \quad (32)$$

Which may be evaluated under the null hypothesis where $\lambda_x = \lambda_y$, or the alternative where $\lambda_x \neq \lambda_y$.

However, terms that involve comparison of the continuous response variable between three subjects must be handled differently. Consider the circumstance where the i^{th} subject in the control group's LVEF has increased by more than the j^{th} and the j'^{th} subject in the active group. Then one of the expressions required for $\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]$ is

$$\begin{aligned}&\mathbf{P}\left[C_{X(i)}(-,R) < C_{Y(j)}(-,R) \cap C_{X(i)}(-,R) < C_{Y(j')}(-,R)\right] \\ &= e^{-(\lambda_x+2\lambda_y)T}\mathbf{P}\left[\left(d_i < d_j\right) \cap \left(d_i < d_{j'}\right)\right].\end{aligned} \tag{33}$$

where d_i is the change in the response variable for the i^{th} subject in the control group over the duration of the study, and d_j and $d_{j'}$ are the response variable changes for the j^{th} and j'$^{\text{th}}$ subjects in the active group respectively. The expression $e^{-(\lambda_x+2\lambda_y)T}$ is the probability that all three subjects (one in the control group and two in the active group) have no SCE throughout the course of the trial and therefore have the continuous measure assessed at baseline and at the end of the study.

To compute $\mathbf{P}\left[\left(d_i < d_j\right) \cap \left(d_i < d_{j'}\right)\right]$, recall that for the one control group subject, d_i follows a $N\left(\mu_{\Delta R}(x), \sigma^2_{\Delta R}(x)\right)$, and for the two active group subjects, d_j, and $d_{j'}$ are identically distributed as $N\left(\mu_{\Delta R}(y), \sigma^2_{\Delta R}(y)\right)$. If we define the random variables U and V in the affine transformation

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} d_i \\ d_j \\ d_{j'} \end{pmatrix} = \begin{pmatrix} d_i - d_j \\ d_i - d_{j'} \end{pmatrix} \tag{34}$$

We may then write, $\mathbf{P}\left[\left(d_i < d_j\right) \cap \left(d_i < d_{j'}\right)\right] = \mathbf{P}\left[U < 0 \ \cap \ V < 0\right]$. Since the joint distribution of $r_i(x)$, $r_j(y)$, and $r_{j'}(y)$, is multivariate normal with mean vector $\underline{\mathbf{u}}$ and variance $\mathbf{\Sigma}$,

$$\underline{\boldsymbol{\mu}} = \begin{pmatrix} \mu_{\Delta R}(x) \\ \mu_{\Delta R}(y) \\ \mu_{\Delta R}(y) \end{pmatrix}; \qquad \Sigma = \begin{pmatrix} \sigma^2_{\Delta R}(x) & 0 & 0 \\ 0 & \sigma^2_{\Delta R}(y) & 0 \\ 0 & 0 & \sigma^2_{\Delta R}(y) \end{pmatrix} \tag{35}$$

we can apply the transformation of (34) we see that the joint distribution of U and V is bivariate normal with mean and variance

$$\begin{pmatrix} U \\ V \end{pmatrix} \text{ is } \mathbf{MVN_2}\left(\begin{pmatrix} \mu_{\Delta R}(x) - \mu_{\Delta R}(y) \\ \mu_{\Delta R}(x) - \mu_{\Delta R}(y) \end{pmatrix}, \begin{pmatrix} \sigma^2_{\Delta R}(x) + \sigma^2_{\Delta R}(y) & \sigma^2_{\Delta R}(x) \\ \sigma^2_{\Delta R}(x) & \sigma^2_{\Delta R}(x) + \sigma^2_{\Delta R}(y) \end{pmatrix} \right) \quad (36)$$

Thus, the desired probability $\mathbf{P}\left[U < 0 \cap V < 0\right]$ is simply the evaluation of this region over the bivariate normal distribution defined in (36), and the probability required by $\mathbf{P}\left[\left(d_i < d_j\right) \cap \left(d_i < d_{j'}\right)\right]$ is therefore available.

Each of the eighteen terms is computed similarly, and assembled in accordance with Table 1 to construct $\mathbf{E}\left[\varphi_{ij}\varphi_{ij'}\right]$ under the null and alternative hypotheses. An analogous table can be constructed for computing $\mathbf{E}\left[\varphi_{ij}\varphi_{i'j}\right]$. Then $\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right]$ and $\mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_0\right]$ can be substituted into equation (13) to compute the $\mathbf{Var}\left[W_e \mid H_0\right]$ and analogously, $\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_a\right]$ and $\mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_a\right]$ will be used to compute $\mathbf{Var}\left[W_e \mid H_a\right]$. Alternatively, $\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_0\right], \mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_0\right]$ and $\mathbf{E}\left[\varphi_{ij}\varphi_{ij'} \mid H_a\right], \mathbf{E}\left[\varphi_{ij}\varphi_{i'j} \mid H_a\right]$ can be substituted into equation (25) and (10) to compute the sample size for the exact computation, or equation (15) in the case of asymptotic solution. These two expectations can be used to compute the power in equations (12) and (16) for the exact and asymptotic solutions respectively.

References

1. Abdel-Latif A, et al.(2007) Adult bone marrow-derived cells for cardiac repair. Arch Intern Med. (2007). 167:989-997.
2. Assmus B, et al. (2002) Transplantation of progenitor cells and regeneration enhancement in acute myocardial infarction (TOPCARE-AMI). Circulation 2002;106:3009-3017.
3. Balsam LB, et al. (2004) Haematopoietic stem cells adopt mature haematopoietic fates in ischemic myocardium. Nature 2004;428:668-673.
4. Follmann D and Wu M. (1995) An approximate generalized linear model with random effects for informative missing data. *Biometrics*. 1995 Mar;51(1):151-68.
5. Ford ES, et al. (2007) Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. N Engl J Med 2007;356:2388-98.
6. Gehan EA. (1965) "A generalized Score test for comparing arbitrarily singly-censored samples". *Biometrika*, 52 203-223 (1965).

7. Genz A, Bretz F, Miwa T, I X, Leisch F, Scheipl F, Bornkamp B, Hothorn T (2011). Mvtnorm: multivariate normal and t distributions. R package version 0.9-96. URL http://CRAN.R-project.org/package=mvtnorm.
8. Grothues F, Smith GC, Moon JC, Bellenger NG, Collins P, Klein HU, Pennel DJ. (2002) Comparison of Interstudy Reproducibility of Cardiovascular Mangetic Resonance With Two-Dimensional Echocardiography in Normal Subjects and in Subjects With Heart Failure or Left Ventricular Hypertrophy. Am J Cardiol; 90: 29-34.
9. Hothorn T, Bretz F,Westfall P (2008). Simultaneous Inference in General Parametric Models. Biometrical Journal 50(3), 346-363.
10. Jackson KA, et al.(2001) Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. J Clin Invest 2001;107:1395-1402.
11. Janssens S, et al.(2006) Autologous bone marrow-derived stemcell transfer in subjects with ST-segment elevation myocardial infarction: doubleblind, randomised controlled trial. Lancet 2006;367:113-121.
12. Kocher AA, et al. (2001) Neovascularization of ischemic myocardium by human bone-marrow-derived angioblasts prevents cardiomyocyte apoptosis, reduces remodeling and improves cardiac function. Nat Med 2001;7:430-436.
13. Kowalski J and Tu XM. (2007) Modern Applied U-statistics. New York. Wiley. 2007.
14. Lachin J. (1999) Worst-rank score analysis with informatively missing observations in clinical trials. Controlled Clinical Trials 1999; 20:408– 422.
15. Levy D, et al. (2002) Long-term trends in the incidence of and survival with heart failure. N Engl J Med 2002;347:1397-402.
16. Lunde K, et al.(2006) Intracoronary injection of mononuclear bone marrow cells in acute wall infarction. NEJM 2006;355:1199-1209.
17. Mann B. and Whitney DR (1947): On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18:50-60.
18. Moyé LA, et al. (1992) Analysis of a Clinical Trial Involving a Combined Mortality and Adherence Dependent Interval Censored Endpoint. S*tatistics in Medicine*. Vol. 11, 1705-1717 (1992).
19. Moyé LA for the SAVE Cooperative Group (1991) Rationale and Design of a Trial to Assess Subject Survival and Ventricular Enlargement after Myocardial Infarction. *Am J Cardiol* 1991;68:70D-79D.
20. Murry CE, et al. (2004) Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts. Nature 2004;428:664-668.

21. National Institute of Neurological Disorders and Stroke rt- PA Stroke Study Group. (1995). Tissue plasminogen activator for acute ischemic stroke. New England Journal of Medicine 333, 1581-1587.
22. O'Brien PC. (1984) Procedures for Comparing Samples with Multiple Endpoints Biometrics,1984;40; 1079-1087.
23. Orlic D, et al. (2001) Bone marrow cells regenerate infracted myocardium. Nature 2001;410:701-705.
24. Penicka M, et al. (2007) Research Correspondence: Intracornonary injection of autologous bone marrow-derived mononuclear cells in subjects with large anterior acute myocardial infarction: A prematurely terminated randomized study. Journal of the American College of Cardiology. 49:2373-2374.
25. Pfeffer MA, et al. (1992) Effect of Captopril on mortality and morbidity in subjects with left ventricular dysfunction after myocardial infarction - results of the Survival and Ventricular Enlargement Trial. *N Eng J Med* 327(10):669-677, Sep 3, 1992.
26. Roger VL, et al. (2004) Trends in heart failure incidence and survival in a community-based population. JAMA 2004;292:344-50.
27. Rosamond W, et al. (2007) Heart disease and stroke statistics--2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2007;115:e69-171.
28. Schächinger V, et al. (2006) Intracoronary bone marrow-derived progenitor cells in acute myocardial infarction. NEJM 2006;355:1210-1221.
29. Strauer BE, et al. (2002) Repair of infarcted myocardium by autologous intracoronary mononuclear bone marrow cell transplantation in humans. Circulation 2002;106:1913-1918.
30. Tan M, Fang HB, Tian GL, Houghton PJ. Small sample inference for incomplete longitudinal data with truncation and censoring in tumor xenograft models. Biometrics 2002; 58:612–620.
31. Tang DI, et al. (1989a) An approximate likelihood ratio test for a normal mean vector with nonnegative components with application to clinical trials. Biometrika 76, 577-583.
32. Tang DI, et al. (1989b) Design of group sequential clinical trials with multiple endpoints. Journal of the American Statistical Association 84, 776-779.
33. Tang DI, et al. (1993) On the design and analysis of randomized clinical trials with multiple endpoints. Biometrics 49, 23-30.
34. Tang DI and Getter NL. (1999) Closed Testing Procedures for Group Sequential Clinical Trials with Multiple Endpoints Biometrics, 55; 1188-1192.

DOI: 10.2202/1557-4679.1286

35. Tilley BC, et al. (1996) Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA stroke trial. Stroke 27, 2136- 2142.
36. Traverse JH, Henry TD, Vaughan D, Ellis SG, Pepine CJ, Willerson JT, Zhao DXM, Piller LB, Penn MS, Byrne BJ, Perin EC, Gee AP, Hatzopoulous AK, McKenna DH, Forder JR, Taylor DA, Cogle CR, Olson RE, Jorgenson BC, Sayre SL, Vojvodic RW, Gordon DJ, Skarlatos SI, Moyé LA, Simari RD for the Cardiovascular Cell Therapy Research Network. Rationale and Design for TIME: A Phase-II, Randomized, Double-Blind, Placebo-Controlled Pilot Trial Evaluating the Safety and Effect of Timing of Administration of Bone Marrow Mononuclear Cells Following Acute Myocardial Infarction. American Heart Journal:158:356-63.
37. Traverse JH, Henry TD, Vaughan D, Ellis SG, Pepine CJ, Willerson JT, Zhao DXM, Piller LB, Penn MS, Byrne BJ, Perin EC, Gee AP, Hatzopoulous AK, McKenna DH, Forder JR, Taylor DA, Cogle CR, Olson RE, Jorgenson BC, Sayre SL, Vojvodic RW, Gordon DJ, Skarlatos SI, Moyé LA, Simari RD for the Cardiovascular Cell Therapy Research Network. Rationale and Design for TIME: A Phase-II, Randomized, Double-Blind, Placebo-Controlled Pilot Trial Evaluating the Safety and Effect of Timing of Administration of Bone Marrow Mononuclear Cells Following Acute Myocardial Infarction. American Heart Journal:158:356-63.
38. Willerson JT, Perin EC, Ellis SG, Pepine CJ, Henry TD, Zhao DX, Lai D, Penn MS, Byrne BJ, Silva G, Gee A, Traverse JH, Hatzopoulos AK, Forder JR, Martin D, Kronenberg M, Taylor DA, Cogle CR, Baraniuk S, Westbrook L, Sayre SL, Vojvodic RW, Gordon DJ, Skarlatos SI, Moyé LA, Simari RD; Cardiovascular Cell Therapy Research Network (CCTRN).Intramyocardial injection of autologous bone marrow mononuclear cells for subjects with chronic ischemic heart disease and left ventricular dysfunction (First Mononuclear Cells injected in the US [FOCUS]): Rationale and design.Am Heart J. 2010 Aug;160(2):215-23.
39. Wittkowski KM, et al. (2004) Combining several ordinal measures in clinical studies. *Statistics in Medicine*. Vol. 23, 1579-1592 (2004).
40. Wollert KC, et al.(2004) Intracoronary autologous bone-marrow cell transfer after myocardial infarction: the BOOST randomized controlled clinical trial. Lancet 2004;364:141-148.
41. Wu JR, Houghton PJ (2010) Interval approach to assessing antitumor activity for tumor xenograft studies. *Pharmaceutical Statistics* 9;1:46-54.
42. Yoon YS, et al. (2005) Clonally expanded novel multipotent stem cells from human bone marrow regenerate myocardium after myocardial infarction. JCI 2005;115:326-338.