Some Adaptive Enrichment Designs

Yonghong Gao, Ph. D.

Biomedical Advanced Research and Development Authority (BARDA),
Office of the Assistant Secretary for Preparedness and Response (ASPR),
US Department of Health and Human Services (HHS),
330 Independence Ave. SW, Washington, DC 20201

Abstract
Subgroup analyses are routinely conducted in analyzing data from clinical trials since variability in response across subgroups is frequently observed. In this paper we look at three two-stage adaptive designs where subgroup analyses are pre-planned as a formal component of the efficacy assessment. We apply those designs to a CRT trial data. We also compare the performance of the trial designs through simulation studies.

Key Words: adaptive, decision rule, subgroup, interim analysis, trial, data

1. Introduction

In many clinical trials variability in response across different subgroups is frequently observed, and sometimes data showed that the experimental treatments have different treatment effect for different subpopulations. For example, one study of a ventricular assist device (VAD) showed that female subjects treated with the investigational VAD were found to have much higher rate of stroke compared to male subjects. The differential findings among different stratum of subpopulation may be caused by random chance or true heterogeneity. The heterogeneity is desirable in that the results of the clinical trial can be generalized to a wide class of patients. However, the heterogeneity could introduce great challenges in trial design, data analysis and the interpretation of trial results, especially in regulatory setting. For example, routine subgroup analyses, usually conducted in a post hoc fashion, could identify one subgroup with a favorable treatment effect and the complementary subgroup with no or negative treatment effect. The decision of whether we should restrict the indication to the favorable subgroup would be a regulatory challenge.

For some medical product developers, the main interest of the clinical trial is to find any subgroup for which the experimental medical product works so that their product can be approved. Under the traditional fixed sample size trial, the sample size of the trial is usually estimated to provide just enough power to reject null hypothesis of no treatment effect on the overall population, and therefore the trial is usually underpowered to detect possible treatment effect on some of the subpopulations. In regulatory setting, when trial data fail to provide significant evidence to reject the null hypothesis for the overall population, and the subgroup analysis indicates clinically, but not statistically, significant treatment effect on some of the subpopulation, the usual approach trial sponsor takes is to design a new trial that is specifically targeting at the promising subpopulation so that the new trial is powered to reject the null hypothesis for that subpopulation. Essentially the first trial serves as an exploratory study for the second trial that is tailed to the specific subpopulation. This two-trial approach is not cost-efficient and can be very time

consuming for the industry to seek regulatory approval for their product. Adaptive design has been proposed in the literature to streamline trials to hopefully increase the chance of getting a success trial, in addition to shorten the trial time and decrease the cost. For example, Follmann (1997) presented a large class of enrichment designs to adaptively change subgroup proportions in trials. See Rosenblum and Van der Laan (2011) for a relatively comprehensive review of adaptive methodologies in the literature. In this paper, we first look at two adaptive designs, where assessment of the treatment effect of the experimental product on the overall population, on each of the non-overlapping subgroups are formally planned in the trial design stage. Then we present a new design which is a mild modification of the adaptive strategy proposed by Rosenblum and Van der Laan (2011). The operating characteristics, such as the type I error rate, the power and the average sample size, of the three trial designs aforementioned are compared through a simulation study. The three designs are applied to a real trial data set, in a post-hoc fashion, to show the possible benefit of the adaptive enrichment designs relative to the traditional fixed designs.

2. Three Two-stage Adaptive Enrichment Designs

We look at a typical two-arm randomized trial where subjects will be randomized at to receiving either the investigational treatment or the active control. The overall population consists of two non-overlapping subgroups, subgroup 1 and subgroup2, with p_1 as the proportion of subjects in subgroup 1. The designs we focus on are two-stage adaptive types where, based on stage I data, the planned interim analysis is conducted and decisions regarding enrollment plan for stage 2 are assessed. Let n_1 and n_2 be the planned sample sizes for stage I and stage II.

2.1 Design 1: proposed by Russek-Cohen and Simon

Russek-Cohen and Simon (1997) present a novel two-stage adaptive design that incorporates a test for a subpopulation by treatment interaction at interim analysis, which is used to determine the sample size, enrollment plan and hypothesis testing of the trial. Here we give a brief description of the design and data analysis strategy.

Suppose the hypothesis is about a normal distributed endpoint. The first stage of the study is planned in the usual way, assuming no important interaction between subgroups and treatment, and stage I sample size n_1 is calculated with the following formula:

$$n_1=4*(\sigma/\delta)^2(z_{1-\alpha/2} + z_{\beta})^2$$

where σ is the standard deviation of the endpoint, δ is the smallest difference in means clinically acceptable between means of the endpoint in the two treatment groups, α and β are the type I and type II error levels, and z_t is the t-th percentile of the standard normal distribution. At the end of stage I, a test for subgroup by treatment interaction is conducted. If no statistically significant interaction is found at a pre-specified significant level, then the trial will be terminated and the final sample size for the trial is n_1. Under this scenario, the two subgroups are deemed as poolable, and the assessment of the overall treatment effect is the goal of the study where traditional data analysis method can be applied. If the interim data indicate a statistically significant interaction, then the two subgroups should be assessed separately. The stage I sample size n_1 usually does not provide sufficient power to detect a treatment effect separately within each subgroup, so

there is need to collect additional data in a second stage that is the same size as the first stage. Suppose n_{11} is the stage I sample size of subgroup 1, if the stage I data for subgroup 1 didn't show significant treatment effect at a level adjusted for multiple looks, then the trial needs to enroll n_{11} subjects from subgroup 1 for stage II, and data analysis for subgroup 1 will be conducted in a two-stage group-sequential approach. Estimation and hypothesis testing of treatment effect for subgroup 2 will be conducted in a similar way.

2.2 Design 2: proposed by Rosenblum and Van der Lann

Rosenblum and Van der Laan (2011) proposed a general method for constructing two-stage randomized enrichment designs that allow changes to the population enrolled based on interim data using a pre-specified decision rule. The following is a brief description of Rosenblum and Van der Laan's (2011) general adaptive methodology.

Let H_{01} denote the null hypothesis for subgroup 1, that there is no treatment effect for subgroup 1. In an analogous manner, define the null hypothesis H_{02} corresponding to subgroup 2, and the null hypothesis H_{0a} corresponding to the overall population. At the end of stage I, three test statistics, $T_a^{(1)}$, $T_1^{(1)}$, and $T_2^{(1)}$, corresponding to the overall population, subgroup 1 and subgroup 2, respectively; are calculated. According to a pre-specified decision rule, enrollment plan for stage II subjects can be decided from the two possible choices: (i) continue enrolling n_2 subjects from both subpopulations in the same way as in stage I or (ii) enroll n_2 subjects only from the subgroup $s \in \{1, 2\}$, corresponding to the larger of the stage I estimates of the treatment effect. A test statistic $T_i^{(2)}$, based on stage II data only, can be calculated in a similar way when stage II data are available, where $i \in \{a, 1, 2\}$, depending on the actual enrollment plan for stage II. At the end of the trial, a final test statistic T is computed in the following way:

$$T = \sqrt{\frac{n_1}{n_1 + n_2}}\, T_a^{(1)} + \sqrt{\frac{n_2}{n_1 + n_2}}\, T_i^{(2)}$$

T is used for a possible rejection of one of the three null hypotheses $\{H_{0a}, H_{01}, H_{02}\}$. If the final test statistic T exceeds a threshold c, which turns out to be the usual critical value for fixed trial design, null hypothesis corresponding to either the enriched subgroup, or the overall population, will be rejected.

2.3 Design 3: proposed by Gao

One interesting feature of Rosenblum and Van der Laan's adaptive design is that the sample size under their design is fixed, regardless of the enrollment decisions for stage II. As re-estimation of sample size is one of the benefits of adaptive design, we propose a modified design of Rosenblum and Van der Laan's method, where the only modification is to re-calculate the sample size for stage II when interim data stipulate enrollment change for stage II. The sample size re-estimation can be conducted by the conditional power approach, so that appropriate power is provided for stage II hypothesis testing for the selected subgroup. So under Gao's design, the total sample size for the trial is random, which is different from Rosenblum and Van der Laan's design. The actual sample size for stage II n_2^*could be larger than the originally planned n_2, if the interim data show a weaker treatment effect, or could be smaller than n_2, if the interim data present a better than expected result for the selected subgroup. However, regardless the new sample size for stage II, the weights used to combine data from stage I and stage II are the pre-specified fixed weights: $n_1/(n_1 + n_2)$ and $n_2/(n_1 + n_2)$. This fixed weights

approach functions to control the type I error rate under the nominal level when the final sample size is data dependent.

3. MADIT- CRT Trial

The MADIT-CRT trial is Boston Scientific's Multicenter Automatic Defibrillator Implantation Trial – Cardiac Resynchronization Therapy study. The goal of this randomized study is to determine whether Cardiac Resynchronization Therapy Defibrillators (CRT-D) in high-risk heart failure (HF) patients will reduce the combined endpoint of all cause mortality or HF intervention when compared to implantable cardioverter defibrillator (ICD) therapy. The MADIT-CRT trial enrolled a total of 1820 patients from 110 centers in 14 countries. Among them 1089 were randomized into CRT-D arm and 731 in ICD arm. The primary endpoint is all-cause mortality or heart failure intervention, whichever occurs first. The following table showed the data from the overall population.

Table 1: Data from Overall Population

	Test Arm	Control Arm	Hazard Ratio (HR), 95% CI for HR
Subject number	1089	731	HR=0.62
Event number	208	208	
Event Rate	19.1%	28.4%	95% CI of HR: (0.50, 0.75)

The above data demonstrated that early CRT intervention reduces the relative risk of all-cause mortality or first heart failure event when compared to ICD therapy. Routine subgroup analyses for a wide range of different subgroups were conducted and a significant interaction between treatment and bundle branch block morphology was detected. Left Bundle Branch Block (LBBB) is a marker of an electrical conduction disorder in the heart and has been associated with a greater benefit in patients receiving CRT. For MADIT-CRT trial, there were 1281 and 539 patients in LBBB subgroup and no-LBBB subgroup. The following two tables displayed the primary endpoint results for the two non-overlapping subgroups.

Table 2: Data from LBBB subgroup

	Test Arm	Control Arm	Hazard Ratio (HR), 95% CI for HR
Subject number	761	520	HR=0.43
Event number	120	162	
Event Rate	15.8%	31.1%	95% CI of HR: (0.33, 0.56)

Table 3: Data from no-LBBB subgroup

	Test Arm	Control Arm	Hazard Ratio (HR), 95% CI for HR
Subject number	328	209	HR=1.32
Event number	81	41	
Event Rate	24.6%	19.6%	95% CI of HR: (0.85, 2.04)

The MADIT-CRT data indicated a quantitative interaction between treatment and the LBBB subgroup: the LBBB subgroup benefits greatly from CRT-D, but not no-LBBB subgroup. And it seemed that the observed statistically significant treatment effect on the overall population is largely driven by LBBB subgroup which constituted 70% of the enrolled patients in the trial. Due to this finding, Boston Scientific's CRT-D indication is limited to sub-population of MADIT-CRT patients with left bundle branch block morphology. An interesting intellectual exercise is to apply the three adaptive designs afore mentioned to the MADIT-CRT data and to see how those adaptive designs perform, compared against each other and against the fixed design.

4. Application of the Three Adaptive Designs to MADIT- CRT Trial Data

The MADIT-CRT data were randomly split into two halves, with the first half serving as stage I data. The following three tables showed the primary endpoint results for overall population (908 subjects), LBBB subgroup (658 subjects) and no-LBBB subgroup (250 subjects) based on the stage 1 data.

Table 4: Stage I Data from Overall Population

	Test Arm	Control Arm	Test Statistic
Subject number	555	355	$T_a^{(1)}$ =4.042
Event number	80	90	
Event Rate	14.4%	25.35%	

Table 5: Stage I Data from LBBB Subgroup

	Test Arm	Control Arm	Test Statistic
Subject number	402	256	$T_1^{(1)}$ =5.323
Event number	45	71	
Event Rate	11.19%	27.73%	

Table 6: Stage I Data from no-LBBB Subgroup

	Test Arm	Control Arm	Test Statistic
Subject number	153	97	$T_2^{(1)}$ = - 0.458
Event number	35	19	
Event Rate	22.87%	19.59%	

The interim data indicate sizable interaction between treatment and subgroup. A test statistic for testing interaction is calculated as the following:

$$Z_p = \frac{T_1^{(1)} - T_2^{(1)}}{\sqrt{2}} = 4.086$$

Z_p is asymptotically normal distributed under the null hypothesis of no interaction, and is significant at 10% significance level. Therefore the two subgroups are deemed as non-poolable based on stage I data. Under the three designs we considered here, the

estimation and testing of the treatment effect on the overall population is not considered when the two subgroups are not poolable.

4. 1 Application of Design 1 to MADIT- CRT Trial Data

The decision under design 1, after the non-poolability conclusion, is to assess the treatment effect for two subgroups separately.

For LBBB subgroup, test statistic is $T_1^{(1)}$=5.323 , which is larger than 2.797, the OB-boundary for a total of two looks, therefore we can reject H_{01} at interim and conclude that CRT-D demonstrated significant treatment effect for LBBB subgroups. So for LBBB subgroup, there is no need to collect more data, and the total sample size for LBBB subgroup is n_{11}=658.

The test statistic is $T_2^{(1)}$=-0.458 for non-LBBB subgroup, indicating possible treatment benefit of the control device, and therefore the stage I data fail to reject $H_{02.}$ Under design 1, the trial needs to move on to stage II for non-subgroup. Additional n_{12}=250 non-LBBB subjects should be enrolled for stage II. We randomly draw 250 non-LBBB subjects from the 287 available non-LBBB subjects left in the second half of the MADIT data, and the result on the primary endpoint is the following:

Table 7: Stage II Data from non-LBBB Subgroup, design 1

	Test Arm	Control Arm	Test Statistic
Subject number	153	97	$T_2^{(2)}$ =-0.3087
Event number	37	21	
Event Rate	24.18%	21.65%	

Non-LBBB Data from two stages are combined with the traditional group sequential approach:

$$T_2 = \sqrt{0.5}\,(T_2^{(1)} + T_2^{(2)}) = -0.5416,$$

and the OB-boundary of 1.977 for the second look, at significant level of 5%, is used to compare against T_2. Since T_2 is less than 1.977, the two-stage data failed to reject H_{02} for non-LBBB subgroup, based on a total of 500 non-LBBB subjects.

So under design 1, the trial conclusion would be that CRT-D demonstrates significant treatment effect for LBBB subgroup only, based on a total of 1160 subjects, 910 of them from stage I and 250 of them from stage II.

4. 2 Application of Design 2 to MADIT- CRT Trial Data

The decision under design 2, after the non-poolability conclusion, is to continue to enroll n_1 subjects from the better subgroup and to assess the treatment effect for that subgroup only. The stage I data indicate the LBBB as the better subgroup as $T_1^{(1)}$ is larger than $T_2^{(1)}$, so we need to enroll 910 LBBB subjects for stage II. Bootstrap technique is used

here to draw 910 LBBB subjects from the 623 available LBBB subjects left in the second half of MADIT dataset. The following table shows the results of the primary endpoint:

Table 8: Stage II Data from LBBB Subgroup, design 2

	Test Arm	Control Arm	Test Statistic
Subject number	520	390	$T_1^{(2)}$ =5.729
Event number	85	128	
Event Rate	16.35%	32.82%	

At the end of the stage II, testing of hypothesis H_{01} is conducted, based on data from both stage I and stage II, even though stage I data consist of non-LBBB subjects. The following formula calculates the overall test statistic for hypothesis $H_{01.}$

$$T_1 = \sqrt{0.5}\,(T_a^{(1)} + T_1^{(2)}) = \sqrt{0.5}(4.04 + 5.73) = 6.91,$$

Since the final test statistic T_1 is larger than the conventional critical value of 1.96, we can reject H_{01} at the 5% significance level and concluded that CRT-D provided more benefit than the control ICD for the LBBB subgroup. So under design 2, the trial conclusion would be that CRT-D demonstrates significant treatment effect for LBBB subgroup only, based on a total of 1820 subjects, 910 of them from stage I and 910 of them from stage II.

4. 3 Application of Design 3 to MADIT- CRT Trial Data

The decision under design 3 is to continue to enroll n_2^* subjects from the better subgroup and to assess the treatment effect for that subgroup only. The stage II sample size $n_2^*$is determined by the observed treatment effect at stage I for the better subgroup. The stage I data indicate the LBBB as the better subgroup, and the observed treatment effect for LBBB at stage I is $T_1^{(1)}$ =5.32. Using conditional power approach, the new sample size $n_2^*$is calculated as 365, in order to provide 80% conditional power for testing $H_{01.}$ Among the 623 available LBBB subjects left in the second half of MADIT dataset, 365 of them are randomly chosen and serve as the stage II data. The following table shows the results of the primary endpoint for stage II, under design 3.

Table 9: Stage II Data from LBBB Subgroup, design 3

	Test Arm	Control Arm	Test Statistic
Subject number	204	161	$T_1^{(2)}$ =2.979
Event number	31	46	
Event Rate	15.19%	28.57%	

The final test statistic for testing hypothesis H_{01} is the weighted average of the test statistics for the two stages. Note that the weights are the originally planned fixed weights, which are 0.5 and 0.5 for this case, even though the actual weights for the two stages are 0.71 and 0.29. The following formula calculates the overall test statistic for hypothesis $H_{01.}$

$$T_1 = \sqrt{0.5}\,(\mathrm{T}_{\mathrm{a}}^{(1)} + \mathrm{T}_1^{(2)}) = \sqrt{0.5}(4.04 + 2.98) = 4.96,$$

As in design 2, the final test statistic T_1 is larger than the conventional critical value of 1.96, therefore we can reject H_{01} at the 5% significance level and concluded that CRT-D provided more benefit than the control ICD for the LBBB subgroup. So under design 3, the trial conclusion would be that CRT-D demonstrates significant treatment effect for LBBB subgroup only, based on a total of 1275 subjects, 910 of them from stage I and 365 of them from stage II.

5. Simulation Study

From section 4 we see that the three designs lead to the same conclusion for MADIT-CRT trial data while requiring different sample sizes. To compare the performance of the three designs under broader circumstances, simulation studies are conducted under a range of assumptions of the treatment effects for the two subgroups. Suppose the primary endpoint is a continuous variable which has a normal distribution with known variance, and there are four balanced independent groups that corresponding to subgroup1-control, subgroup1-experimental, subgroup2-control, subgroup2-experimental. Let μ, μ_1 and μ_2 denote the true treatment effects for the overall population, subgroup1 and subgroup2, respectively, the following six scenarios are considered in the simulation study:

1) Poolable, complete null hypothesis: $\mu=\mu_1=\mu_2=0$,
2) Poolable, but treatment is effective: $\mu=\mu_1=\mu_2=0.2$,
3) Mild non-poolable, treatment is effective at different degree for two subgroups: $\mu=0.2$, $\mu_1=0.3$, $\mu_2=0.1$,
4) Mild non-poolable, treatment is effective for only one subgroup:$\mu=.2$, $\mu_1=0.4$ $\mu_2=0$,
5) Highly un-poolable, treatment is effective for only one subgroup and there is no treatment effect for the overall population: $\mu=0$, $\mu_1=0.2$ $\mu_2=-0.2$,
6) Highly un-poolable, treatment is effective for only one subgroup and there is treatment effect for the overall population: $\mu=0.2$, $\mu_1=0.5$ $\mu_2=-0.1$.

For our simulation study, significance level of 0.10 is used for poolability assessment at interim analysis, 5% significance level is used in the two-sided test of significant treatment effect for overall population and for the subgroups. For design 1, the two-look OB-Fleming group sequential boundaries are used. For design 3, 80% conditional power is used to calculate the new sample size for stage II if there is change in enrollment plan. Sample size of 140 for stage I is planned to provide adequate power for testing the overall treatment effect of 0.2, and 10,000 repetition is used in our simulation for all scenarios. For each of the six scenarios, we compute the following probabilities

1) P_p: declaring poolable for the two subgroups,
2) P_f: rejecting any null among the three null hypotheses,
3) P_0: rejecting null hypothesis for the overall population,
4) P_1: rejecting null hypothesis for subgroup 1,
5) P_2: rejecting null hypothesis for subgroup 2.

Note that P_f is the family-wise type I error rate of the trial under scenario 1, and family-wise power for scenarios 2, 3, 4, 5 and 6. P_0 is the type I error rate for the overall population testing under scenarios 1 and 5, and power under scenarios 2, 3, 4 and 6. P_1 is the type I error rate for the subgroup 1 testing under scenario 1, and power under scenarios 2, 3, 4, 5 and 6. In addition to those probabilities, average sample sizes of the

trials under the three designs are calculated also in our simulation study. For following tables present the simulation results.

Table 10: Simulation Result, Scenario 1

Design	P(poolable)	P_f	P_0	P_1	P_2	Avg. Sample size
Design 1	0.8978	0.0393	0.0219	0.0097	0.0077	154
Design 2	0.9011	0.0251	0.0228	0.0016	0.0007	280
Design 3	0.8954	0.0258	0.0223	0.0012	0.0023	309

Table 11: Simulation Result, Scenario 2

Design	P(poolable)	P_f	P_0	P_1	P_2	Avg. Sample size
Design 1	0.9015	0.6832	0.5887	0.0583	0.0622	148
Design 2	0.8995	0.9185	0.8265	0.0471	0.0449	280
Design 3	0.8981	0.9168	0.8203	0.0473	0.0492	278

Table 12: Simulation Result, Scenario 3

Design	P(poolable)	P_f	P_0	P_1	P_2	Avg. Sample size
Design 1	0.6752	0.7620	0.4399	0.3212	0.0217	169
Design 2	0.6835	0.9365	0.6240	0.3114	0.0011	280
Design 3	0.6788	0.9398	0.6203	0.3185	0.0010	275

Table 13: Simulation Result, Scenario 4

Design	P(poolable)	P_f	P_0	P_1	P_2	Avg. Sample size
Design 1	0.2405	0.9146	0.1554	0.7592	0.0078	202
Design 2	0.2380	0.9796	0.2182	0.7614	0.0000	280
Design 3	0.2372	0.9791	0.2180	0.7611	0.0000	252

Table 14: Simulation Result, Scenario 5

Design	P(poolable)	P_f	P_0	P_1	P_2	Avg. Sample size
Design 1	0.2288	0.5694	0.0056	0.5638	0.0000	239
Design 2	0.2370	0.3014	0.0067	0.2947	0.0000	280
Design 3	0.2355	0.5624	0.0063	0.5561	0.0000	425

Table 15: Simulation Result, Scenario 6

Design	P(poolable)	P_f	P_0	P_1	P_2	Avg. Sample size
Design 1	0.0283	0.9900	0.0184	0.9716	0.0000	213
Design 2	0.0285	0.9975	0.0261	0.9714	0.0000	280
Design 3	0.0292	0.9960	0.0261	0.9699	0.0000	217

6. Conclusion and Discussion

From tables 10 -15, we can see that the family-wise type I error rate is inflated under design 1 (see table 10) while both design 2 and design 3 maintain the type I error rate in the nominal level of 2.5%. We believe that is because only design 1 allows hypothesis testing on the two subgroups after declaring unpoolability of the two subgroups. Regarding the power of the design, design 2 and design 3 provide higher overall power than design 1 when the two subgroups are poolable or mild non-poolable. The price for the higher power is the larger sample size, as the average sample size for design 2 and design 3 are larger than that of the design 1 under scenarios 2, 3, and 4. When the two subgroups are highly unpoolable, corresponding to scenarios 5 and 6, design 1 and design 3 outperform design 2 in the overall power and in detecting success for the effective subgroup. Design 2 and design 3 have similar operating characteristics when the two subgroups are similar. However when the two subgroups are highly unpoolable, design 2 and design 3 have different operating characteristics, especially in terms of the average sample size. Design 3 requires smaller sample size than design 2 if one subgroup has large treatment effect without sacrificing power; and design 3 requires larger sample size than design 2 when the two subgroups are highly unpoolable and one subgroup has modest treatment effect. In conclusion, each design has its advantage and disadvantage, different trial scenario calls for different design. It is important to study the prior evidence of subgroup differences and choose the right design accordingly for the trial.

References

1. Cohen and Simon "Evaluating treatments when a gender by treatment interaction may exist", *Statistics in Medicine* (1997)
2. M. Rosenblum and M. Van Der Lann, "Optimizing Randomized Trial Designs to Distinguish which subpopulations benefit from treatment", *Biometrika* (2011), **98**, 4, *pp.* 845–860
3. Follmann, "Adaptively changing subgroup proportions in clinical trials", Statistica Sinica, 1997
4. Thall, Simon and Ellenberg, "Two-stage selection and testing designs for comparative clinical trials", Biometrika, 1988
5. Kieser, Bauer, and Lehmacher "inference on multiple endpoints in clinical trials with adaptive interim analyses", Biometrical Journal, 1999
6. http://www.bostonscientific.com/cardiac-rhythm-resources/clinical/madit-crt-trial.htmlMADIT trial