DJ 5/12/14

Adaptive Design To Determine The Safety And Effectiveness Of The Barricaid In The Treatment Of Back And Radicular Pain Subsequent To Primary Lumbar Disc Herniation

Statistical Considerations

May 6, 2014

Submitted to Intrinsic Therapeutics

Berry Consultants, LLC

The intention of this Statistical Analysis Plan (SAP) is to detail the planned analysis and reporting for the Intrinsic Therapeutics "Prospective Randomized, Multicenter Study to Demonstrate the Superiority of the Barricaid to Discectomy for Primary Lumbar Disc Herniation" Clinical Trial, Protocol Number EUBARD-CP-001. This SAP is supplemental to the clinical protocol, which provides further detail on the conduct of this study, the required protocol assessments and study timelines.

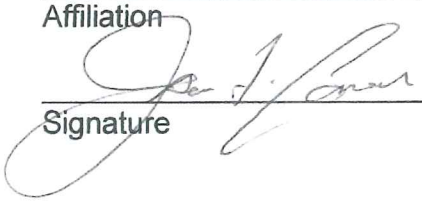
The undersigned hereby jointly declare that they have reviewed this statistical analysis plan and agree to its content. Furthermore, they confirm that the statistical analysis plan contains the information relevant for the evaluation of study data.

Statistical Analysis Plan Reviewed and Approved By:

Study Statistician: Jason Connor, Ph.D. (Name) 5/6/2014 (Date)

Director & Senior Statistical Scientist (Title)

Berry Consultants, LLC (Affiliation)

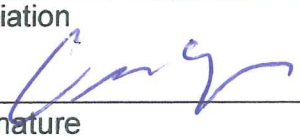


(Signature)

Sponsor Approval: Oscar Yeh (Name) 5/6/2014 (Date)

Vice President, Research (Title)

Intrinsic Therapeutics, Inc. (Affiliation)



(Signature)

Clinical Trial Objective: The objective of this trial is to show that the Barricaid is superior to control (i.e., limited discectomy as described by Spengler[1]) in the reduction of pain, dysfunction, and reherniations subsequent to discectomy for primary lumbar disc herniation.

Primary Outcome: The *treatment arm* (Barricaid) is compared to primary lumbar discectomy (control arm). There are two co-primary measures for this study. The first is a responder analysis at 24-months post-op where a subject is a responder if each of the following are satisfied:

a. A 15/100 point improvement in ODI compared to preoperative

b. A 20/100 point improvement in VAS Leg. This is based on the primary leg complaint. If both legs have a minimum of 40/100 at baseline the average leg score will be used.

c. A 75% or greater disc height maintenance when compared with preoperative disc height.

d. No reherniations at the index level (on either side)

e. No posterior migration of the Barricaid (Migration is defined as presence of AP or lateral motion of the anchor ≥ 2 mm relative to its initial position, AND/OR motion of the radio-opaque markers beyond the margin of the disc space, associated with extrusion of the occlusion component through the annulus.)

f. Maintenance or improvement in neurological score related to index level

g. No spontaneous fusions

h. No removal or revision of the Barricaid or supplemental fixation of any patient (at the index level).

The second, co-primary endpoint, is no reherniations at the index level (on either side).

The trial will be considered successful if superiority is demonstrated on each co-primary endpoint.

[1] Spengler, D. Lumbar Discectomy – Results with Limited Disc Excision and Selective Foraminotomy. Spine. Vol 7. No. 6, pgs 604 – 607, 1982.

Randomization: 1:1 for treatment to control. A randomized block design across all sites will be used.

Analysis Populations: The following subject groups or analysis populations will be used to complete the analysis of data:

Intent-to-treat patient population (ITT): The ITT patient population will include all patients randomized, where patients will be classified by the group in which they are randomized, regardless of the treatment received.

Modified Intent-to-treat patient population (mITT): The mITT patient population will include all patients randomized in whom the intended procedure is attempted. This includes all randomized control patients and all randomized treatment group patients for whom the investigational Barricaid device is opened and the delivery tool is inserted through the skin. Patients will be included in the treatment group to which they are randomized, regardless of the treatment received.

Generation 3 Population (G3): The Generation 3 population will include all mITT patients randomized to control and all mITT patients randomized to the treatment in whom an attempt at treatment with the Generation 3 device was made. Patients will be included in the treatment group to which they were randomized, regardless of the treatment received.

As-treated patient population (AT): The AT patient population will include all randomized mITT subjects that receive treatment, classified according to the treatment actually received (i.e. patients randomized to the treatment group who do not have the Barricaid device implanted will be classified as control subjects for this analysis).

Per protocol patient population (PP): The PP patient population will include all mITT subjects who received the assigned treatment and had no major inclusion/exclusion protocol deviations likely to impact primary and key secondary outcome assessments.

Screen failure population (SF): The SF patient population will include all subjects consented in whom the discectomy procedure is initiated but randomization did not occur.

Within each analysis population, all patients that are randomized to the treatment arm, but who do not have a Barricaid device successfully implanted will be considered failures of the composite endpoint.

The primary efficacy analysis will be conducted on the mITT patient population. Secondary efficacy analyses will be done on the ITT population, G3 population, AT, and PP populations. All safety data will be analyzed based on the mITT patient

population. Safety data on the SF population will be summarized separately.

Some patients will be blinded to the treatment they receive and some will not, depending upon approved protocol at the site where they are enrolled. Efficacy analyses will also be conducted, based on the mITT population, in each of these patient subgroups (blinded and unblinded).

Adaptive Sample Size: A Bayesian adaptive approach to sample size selection is used. A minimum total sample size of 400 and a maximum of 800 are considered. An interim look is made at the data when 400 subjects have been accrued. If trial success is highly likely accrual will be stopped. If accrual continues, another look is made after 50 additional subjects have been accrued. These 50-subject incremental looks continue (400, 450, 500, 550, 600, 650, 700, 750) until accrual is stopped or 800 subjects have been accrued. The details of the interim looks are presented in the "Statistical Analysis for Interim Looks" section.

Primary Statistical Analysis: Two primary analyses are conducted. This section describes the test being conducted for each analysis.

The test for superiority on the responder rate is as follows. The probability of a subject being a responder is labeled π_T and in the control group π_C. The hypothesis of superiority is

$$\pi_T > \pi_C.$$

The claim of superiority will be accepted if the posterior probability of superiority is larger than 0.95. That is, if,

$$\Pr\left(\pi_T > \pi_C \mid \text{Trial Results}\right) \geq 0.95.$$

This posterior probability is calculated using Beta(1,1) priors for each π.

The test for superiority on the reherniation rate is as follows. Let the probability a subject does not have a reherniation be θ_T and θ_C, for the treatment and control groups respectively. The hypothesis of superiority is

$$\theta_T > \theta_C.$$

The claim of superiority will be accepted if the posterior probability of superiority is larger than 0.95. That is, if,

$$\Pr\left(\theta_T > \theta_C \mid \text{Trial Results}\right) \geq 0.95.$$

This posterior probability is calculated using Beta(1,1) priors for each θ.

The cut-offs of 0.95 for each co-primary analysis control the type I error of the adaptive design. The operating characteristics of the design are presented in the Operating Characteristics section.

For each final endpoint 95% credible intervals (2.5^{th} to 97.5^{th} percentile) will be presented for the difference between the success rates, $\pi_T - \pi_C$, and the no-reherniation rate, $\theta_T - \theta_C$, in the two treatments. Additionally, a credible interval will be presented for each follow-up interval of 3 months, 6 months, 12 months, and 24 months.

Missing Data: The primary analyses will be done on the mITT population. Subjects with missing data will be included in this primary analysis using Bayesian multiple imputation. This approach will enable an analysis to be done based on the full mITT population. The details of this approach are provided in the statistical details section.

Statistical Analysis for Interim Looks: Interim looks are made at the data before accrual is stopped. The first looks conducted are *sample size looks*. At these looks the joint predictive probability of superiority for both co-primary endpoints is calculated. All interim results available are used to calculate the predictive probability of trial success for the currently accrued subjects. A decision is made whether to stop at the current sample size or to continue accrual. This sample size look is made at sample sizes of 400, 450, 500, 550, 600, 650, 700 and 750 assuming these sample sizes are reached. Let PP_n be the predictive probability of superiority (which implies both co-primary endpoints) for a trial with the current sample size of n. The following rules are used to guide the sample size selection.

Interim looks will be performed based on the mITT population. If $PP_n > S_n$ then the sample size is considered sufficient for superiority and accrual stops. If $PP_{800} < F_n$ then accrual is stopped for futility. The following "cut-offs" are used for this design:

Sample Size Look	F_n	S_n
400	0.01	0.95
450	0.01	0.90
500	0.025	0.90
550	0.05	0.90
600	0.05	0.90
650	0.05	0.90
700	0.05	0.90
750	0.05	0.90
800	--	--

Table 1: The cut-offs used for the efficacy design. The F_n are the futility cut-offs and the S_n are the sample size success cut-offs. The cut-offs are a function of the sample size look, n.

When the sample size has been determined then *early claim interim analyses* will be conducted. Early claim analyses will be based on the mITT population. These analyses take place when accrual is stopped, 6-months, 12-months, and 18-months following accrual being stopped (possibly a total of four early claim analyses). No early claim look will be made until at least 200 treatment group subjects have reached 24-months of follow-up. If the predictive probability of success (joint probability for each endpoint) in an early claim analysis is at least 0.99 then an immediate claim of superiority will be made.

Predictive Probabilities:

The two co-primary endpoints are correlated. The reherniations will be counted as a failure in each group. Because of this correlation the modeling of the two endpoints will incorporate this dependency. The primary analyses will analyze these endpoints separately, but the adaptive design incorporates the dependency.

Let θ_T and θ_C be the probability of a subject being a success on reherniation–that is 24-months with no reherniation, for the treatment and control groups, respectively. Let γ_T and γ_C be the conditional probability of a subject being a success on the responder analysis given they have no reherniation–that is a 24-month responder, given they had no reherniations. Since a reherniation failure is a failure on the responder analysis this makes the opposite condition, a success on a reherniation failure, trivial. The probability of a subject on therapy *t*, being a success on the responder analysis is $\theta_t{}^*\gamma_t$. The probability of being a success on the reherniation endpoint is θ_t.

At the time of the interim analyses there will be subjects who have completed 24 months, subjects who have 3-, 6-, and 12-month results, and subjects who don't have any interim results. Bayesian predictive distributions are employed for the *multiple imputation* of results for subjects who do not have complete results. Multiple imputation is used for both the reherniation and responder analysis. Each of these models are described below.

Reherniation

For subjects with interim data we only have to model those that are reherniation free. If a subject has a reherniation by an interim visit (month 3, 6, or 12) then that subject is known to be a failure on the reherniation and responder analyses. We model the likelihood that a subject that is reherniation-free at month 0, 3, 6, and 12 will be reherniation free at 24 months. Let the probability of a subject not having a reherniation by month *j–1* being reherniation free at month j (*j* = 1 is 3-month, *j*=2 is 6-month, *j*=3 is 12-month, and j=4 for 24-month) for treatment group t (*t* = 1 is *treatment* and t = 2 is *control*) be P_{tj}. The prior distribution for each of these probabilities is a beta(α_j, β_j) distribution. We use the following prior distributions:

Time □interval	*Beta*(α_j, β_j)
j = 1 (3-months)	9.74,0.26
j = 2 (6-months)	9.74,0.26
j = 3 (12-months)	9.74,0.26
j = 4 (24-months)	9.74,0.26

Table 2: The table entries are the a_{jr}, b_{jr} pair for each *j,r* pair.

These prior distributions are symmetric in the treatments to be conservative, and are based on the sponsor company beliefs and pilot data. The posterior mean of 0.974 for each treatment would result in a probability of 0.90 that the subject is reherniation free at 24 months ($0.974^4 = 0.90$). These prior distributions are not used in the final analysis of the treatment, but rather for sample size looks. The prior distribution used in the calculation of the probability of superiority or in the probability of early success is a beta(1,1).

Let R_{tj} be the number of subjects in treatment group t, that had a reherniation and attributed to visit j. If a reherniation occurs between visits it is attributed to the later visit. Let the number of subjects that have had their j-month visit and are reherniation free be N_{tj}. The posterior distribution for P_{tj} is a beta($\alpha_j + N_{tj}$, $\beta_j + R_{tj}$) distribution. In imputing the probability of a subject being reherniation free, conditional on being reherniation free at month j, is

Month j	Imputation probability of 24-month reherniation free
j=0 (Month 0)	$P_{t1}P_{t2}P_{t3}P_{t4}$
j=1 (Month 3)	$P_{t2}P_{t3}P_{t4}$
j=2 (Month 6)	$P_{t3}P_{t4}$
j=3 (Month 12)	P_{t4}

Table 3: The table entries are the imputation probabilities of reherniation free for a subject in treatment group *t* that has gone to month *j* without a reherniation.

Responder

For subjects with no reherniation, a beta-binomial distribution is used for modeling the transition from the j-month to 24-month value, separately within each treatment. Prediction is based on the subject level of success and failure—not on the eight components of success separately. Therefore, imputation of reherniation is done first, and those that are reherniation-free at 24-months have their responder outcome imputed. The responder outcome is then imputed conditional on no reherniation.

For subjects that are reherniation-free at 24-months, let S_{tjr} and F_{tjr} be the number of 24-month successes and failures, respectively, for treatment group t (t = 1 is *treatment* and t = 2 is *control*), at interim stage j (j=0 for no interim results, j = 1 is 3-month, j=2 is 6-month, and j=3 for 12-month), for interim result r (r = 1 is failure at j and r = 2 is success at j). That is, S_{112}, is the number of subjects who were a r=2 (success) at the j=3-month visit, in treatment group t=1, that ended as 24-month successes. Likewise, F_{112}, is the number of subjects who were a success (r =2) at the 3-month visit (j=3), in the treatment group (t=1), that ended as 24-month failures.

At an interim analysis time there will be 14 different beta distributions for the probability of a subject being a success (3 time periods with 2 interim results plus the no-interim data failures is 7, doubled for each treatment is 14). Label the probability of a subject being a success at their 24-month visit given their current status (t, j, and r) and no reherniations as Q_{tjr}. The beta posterior distribution at an interim analysis, for Q_{tjr}, is beta(a_{jr} + S_{tjr}, b_{jr} + F_{tjr}). We use the following prior distributions:

Time □interval	r = 1 (failure)	r = 2 (success)
j = 0 (0-months)	7,3	
j = 1 (3-months)	5,5	8,2
j = 2 (6-months)	2,8	8,2
j =3 (12-months)	2,8	9,1

Table 4: The table entries are the a_{jr}, b_{jr} pair for each j,r pair.

These prior distributions are symmetric in the treatments to be conservative, and are based on the sponsor company beliefs and pilot data. These prior distributions are not used in the confirmation of the treatment, but rather for shaping the sample size. The prior distribution used in the calculation of the probability of superiority or in the probability of early success is a beta(1,1).

Operating Characteristics

The design is simulated as described and assuming that the primary analysis will be based on an mITT population. These operating characteristics demonstrate the power of the design and demonstrate that the type I error is controlled at less than 0.05. For each simulation case values are assumed for the probability of reherniation-free, the probability of response conditional on no reherniations, and the individual visit outcomes. Eleven cases are simulated for the default operating characteristics. These scenarios are shown in Table 5. In each of the first three rows, which are each null hypotheses, the probability of a successful trial is less than 0.05 in each case. The design is efficient at learning that the trial is futile and these three scenarios result in small expected sample sizes (with approximately 200, 115, and 150 subjects saved from the maximum sample size of 800). The design has good power in cases 6, 7, and 10. Case 6, in which case the increase in response rate is 15% and the

reherniation-free rate is increased 6% is considered the hypothesis with the goal of maximizing power. In this scenario the power is 81.9% with a mean sample size of 747 subjects.

Case	Control		Treatment		Both	Only Resp	Only Reher	Early	Mean N	SD N
	no RH	Pr(R)	no RH	Pr(R)						
1	0.90	0.65	0.90	0.65	0.006	0.011	0.010	0.000	604.1	82.9
2	0.90	0.65	0.96	0.65	0.028	0.000	0.212	0.004	683.7	83.3
3	0.90	0.65	0.90	0.85	0.030	0.229	0.000	0.001	648.5	102.7
4	0.90	0.65	0.96	0.70	0.290	0.004	0.226	0.098	722.3	89.4
5	0.90	0.65	0.96	0.75	0.679	0.019	0.039	0.383	744.8	84.4
6	0.90	0.65	0.96	0.80	0.803	0.027	0.001	0.489	746.8	81.3
7	0.90	0.65	0.96	0.85	0.828	0.034	0.000	0.500	742.8	82.4
8	0.90	0.65	0.92	0.85	0.166	0.280	0.000	0.019	688.3	106.1
9	0.90	0.65	0.94	0.85	0.502	0.181	0.000	0.130	729.3	95.6
10	0.90	0.65	0.98	0.85	0.969	0.000	0.000	0.908	725.0	74.0
11	0.90	0.65	0.96	0.75	0.412	0.127	0.015	0.114	714.0	101.3

Table 5: The operating characteristics for the default cases. The true values assumed for the reherniation-free rate (no RH) and responder rate (Pr(R)) in each treatment group are reported in the first four columns. A set of simulation results are presented in the last five columns. The "Both", "Resp", and "Reher" columns are the probability of success on both co-primary endpoints, the responder analysis, and the reherniation analysis respectively. The probability of early success is reported in the Pr(E) column. The mean and standard deviation of the trial sample size is reported in the "Mean N" and "SD N" columns.

Additional operating characteristics have been provided in electronic files, described in the Additional Operating Characteristics" section. Additional simulations have been conducted for sensitivity. The details of these analyses are provided in the Sensitivity section. The details of the simulations are described in the following section.

Simulation Details

We use simulation to find the design's false-positive rate and power. Namely, we generate trial results assuming particular values of θ_T, θ_C, π_T and π_C and particular accrual rates. We follow the design of the trial exactly as described. In particular, after 400, 450, 500, 550, 600, 650, 700 and 750 patients have accrued we conduct a predictive analysis as described. If the predictive probability of eventual trial success is at least as large as the cut-off S_n then we stop accrual. Otherwise we continue to the next interim analysis (or stop at 800). When accrual is stopped the looks for an early claim of success are made. We then repeat this process, generating a total of 10,000 trials for each scenario.

In this section the details of the simulation are presented. Various assumptions are made to simulate the subjects. Many of these assumptions affect the simulations but have no impact on the results of the actual trial.

The accrual rate is assumed to be (per month): 5, 10, 15, 25, and then 25 every month until the study is finished. Subject results are simulated as follows.

A reherniation is simulated with probability $1-\theta$. If a reherniation is simulated to occur then it is assumed to be exponentially distributed between 0 and 24 months. If a reherniation occurs then a subject is assumed a responder failure at that point as well (as reherniation is a failure). If no reherniation occurs then a responder outcome is simulated as follows. A conditional probability of 24-month success, γ, is assumed. This value is selected so that $\theta\gamma$ is the desired probability of 24-month responder success. So, for example, if the probability of a reherniation is 10% by 24 months, and the probability of response is 0.65 then the probability of a responder success, conditional on no-reherniation is assumed to be 0.7222 (which results in a probability of response of 0.65).

The value of response or not is simulated for the 3, 6, and 12 visits as well. The following "default" transition probabilities are assumed:

At 3-months:
P(Success) = 0.70
P(Failure) = 0.30

At 6-months:
If 3-month FAILURE:
P(Success) = 0.20
P(Failure) = 0.80

If 3-month SUCCESS:
P(Success) = 0.90
P(Failure) = 0.10

At 12-months:
If 6-month FAILURE:
P(Success) = 0.20
P(Failure) = 0.80

If 6-month SUCCESS:
P(Success) = 0.90
P(Failure) = 0.10

At 24-months:
If 12-month FAILURE:

P(Success) = 0.10
P(Failure) = 0.90

If 12-month SUCCESS:
P(Success) = 0.90
P(Failure) = 0.10

Using these values the probability of success at 24-months is 0.6464. In order to create different probabilities of 24-month success we vary the above values. This is done by transforming the probabilities of success at each stage to create the desired 24-month probability of success. Each of the above probabilities of success at t has a corresponding value on the logistic scale—for a value of p, this is: $\gamma = \log(p/1-p)$. We use the following values of success at time t:

$$p = \frac{\exp(\gamma + \varepsilon)}{1 + \exp(\gamma + \varepsilon)}.$$

The value of ε is selected to give the appropriate 24-month success probabilities. The value of $\varepsilon = 0$ results in 0.6464.

In the simulations a missing rate for subjects is assumed. A constant exponential rate is assumed, with a 24-month probability of missing rate of 0.10. These subjects "missing" are then treated using the Bayesian multiple imputation, exactly the same as the primary analysis specifies. The power reported in this design accounts for the analysis of missing subjects and thus an addition "inflation" of the sample size is not needed.

Sensitivity:

In this section we present some scenarios in which the assumptions for simulating the data are altered.

Scenario A: In this scenario the rate of accrual reaches a maximum of 12.5 per month (half of the projected value).

Case	Control		Treatment		Both	Only Resp	Only Reher	Early	Mean N	SD N
	no RH	Pr(R)	no RH	Pr(R)						
1	0.90	0.65	0.90	0.65	0.006	0.011	0.008	0.001	495.0	110.5
2	0.90	0.65	0.96	0.65	0.033	0.000	0.099	0.010	545.5	131.1
3	0.90	0.65	0.90	0.85	0.041	0.178	0.000	0.004	596.2	139.2
4	0.90	0.65	0.96	0.70	0.337	0.005	0.170	0.156	650.3	145.8
5	0.90	0.65	0.96	0.75	0.779	0.025	0.043	0.494	684.6	129.4
6	0.90	0.65	0.96	0.80	0.909	0.037	0.001	0.616	671.5	129.9
7	0.90	0.65	0.96	0.85	0.912	0.045	0.000	0.609	665.5	133.8
8	0.90	0.65	0.92	0.85	0.201	0.298	0.000	0.041	671.4	139.8
9	0.90	0.65	0.94	0.85	0.585	0.202	0.000	0.212	710.2	126.0
10	0.90	0.65	0.98	0.85	0.997	0.001	0.000	0.931	555.5	114.5
11	0.90	0.65	0.96	0.75	0.520	0.162	0.013	0.197	695.0	136.1

Table 5-A: The operating characteristics for the default cases. The true values assumed for the reherniation-free rate (no RH) and responder rate (Pr(R)) in each treatment group are reported in the first four columns. A set of simulation results are presented in the last five columns. The "Both", "Resp", and "Reher" columns are the probability of success on both co-primary endpoints, the responder analysis, and the reherniation analysis respectively. The probability of early success is reported in the Pr(E) column. The mean and standard deviation of the trial sample size is reported in the "Mean N" and "SD N" columns.

Scenario B: In this scenario the rate reaches a maximum of 35 subjects per month.

Case	Control		Treatment		Both	Only Resp	Only Reher	Early	Mean N	SD N
	no RH	Pr(R)	no RH	Pr(R)						
1	0.90	0.65	0.90	0.65	0.006	0.018	0.023	0.000	664.0	109.0
2	0.90	0.65	0.96	0.65	0.038	0.000	0.770	0.003	771.1	72.6
3	0.90	0.65	0.90	0.85	0.037	0.436	0.000	0.001	698.1	109.3
4	0.90	0.65	0.96	0.70	0.370	0.006	0.467	0.084	775.5	68.4
5	0.90	0.65	0.96	0.75	0.794	0.028	0.061	0.317	779.2	64.0
6	0.90	0.65	0.96	0.80	0.863	0.044	0.000	0.406	782.0	60.4
7	0.90	0.65	0.96	0.85	0.882	0.045	0.000	0.417	785.8	54.2
8	0.90	0.65	0.92	0.85	0.200	0.453	0.000	0.011	734.4	99.6
9	0.90	0.65	0.94	0.85	0.554	0.257	0.000	0.097	764.9	79.3
10	0.90	0.65	0.98	0.85	0.983	0.001	0.000	0.872	796.6	27.4
11	0.90	0.65	0.96	0.75	0.499	0.201	0.021	0.077	753.3	89.0

Table 5-B: The operating characteristics for the default cases. The true values assumed for the reherniation-free rate (no RH) and responder rate (Pr(R)) in each treatment group are reported in the first four columns. A set of simulation results are presented in the last five columns. The "Both", "Resp", and "Reher" columns are the probability of success on both co-primary endpoints, the responder analysis, and the reherniation analysis respectively. The probability of early success is reported in the Pr(E) column. The mean and standard deviation of the trial sample size is reported in the "Mean N" and "SD N" columns.

Scenario C: In this scenario the control rate for each endpoint is increased.

Case	Control		Treatment		Both	Only Resp	Only Reher	Early	Mean N	SD N
	no RH	Pr(R)	no RH	Pr(R)						
1	0.92	0.75	0.92	0.75	0.007	0.010	0.009	0.000	606.7	82.9
2	0.92	0.75	0.92	0.75	0.030	0.000	0.207	0.003	682.3	84.5
3	0.92	0.75	0.92	0.90	0.029	0.219	0.000	0.002	645.2	102.4
4	0.92	0.75	0.97	0.80	0.336	0.008	0.197	0.107	725.4	90.5
5	0.92	0.75	0.97	0.85	0.701	0.035	0.014	0.354	745.7	85.5
6	0.92	0.75	0.97	0.90	0.775	0.049	0.000	0.398	746.7	82.8
7	0.92	0.75	0.97	0.95	0.800	0.052	0.000	0.405	746.2	82.1
8	0.92	0.75	0.94	0.90	0.190	0.268	0.000	0.022	690.0	107.1
9	0.92	0.75	0.96	0.75	0.587	0.132	0.000	0.184	735.3	92.9
10	0.92	0.75	0.98	0.90	0.909	0.007	0.000	0.689	751.6	71.3
11	0.92	0.75	0.95	0.85	0.341	0.198	0.002	0.064	707.9	104.7

Table 5-C: The operating characteristics for the default cases. The true values assumed for the reherniation-free rate (no RH) and responder rate (Pr(R)) in each treatment group are reported in the first four columns. A set of simulation results are presented in the last five columns. The "Both", "Resp", and "Reher" columns are the probability of success on both co-primary endpoints, the responder analysis, and the reherniation analysis respectively. The probability of early success is reported in the Pr(E) column. The mean and standard deviation of the trial sample size is reported in the "Mean N" and "SD N" columns.

Scenario D: In this scenario the control rates for each endpoint are decreased.

Case	Control		Treatment		Both	Only Resp	Only Reher	Early	Mean N	SD N
	no RH	Pr(R)	no RH	Pr(R)						
1	0.88	0.55	0.88	0.55	0.004	0.010	0.010	0.000	601.9	82.6
2	0.88	0.55	0.95	0.55	0.034	0.000	0.218	0.006	686.2	82.9
3	0.88	0.55	0.88	0.75	0.031	0.220	0.000	0.002	647.9	102.4
4	0.88	0.55	0.95	0.60	0.275	0.002	0.245	0.106	723.2	87.8
5	0.88	0.55	0.95	0.65	0.667	0.010	0.062	0.395	743.9	83.3
6	0.88	0.55	0.95	0.70	0.815	0.020	0.002	0.558	747.3	79.4
7	0.88	0.55	0.95	0.75	0.850	0.024	0.000	0.583	741.5	80.9
8	0.88	0.55	0.90	0.75	0.144	0.288	0.000	0.016	684.8	106.9
9	0.88	0.55	0.93	0.75	0.588	0.128	0.000	0.203	731.5	93.9
10	0.88	0.55	0.97	0.75	0.964	0.000	0.000	0.915	724.2	74.6
11	0.88	0.55	0.93	0.65	0.465	0.082	0.029	0.159	718.1	99.1

Table 5-D: The operating characteristics for the default cases. The true values assumed for the reherniation-free rate (no RH) and responder rate (Pr(R)) in each treatment group are reported in the first four columns. A set of simulation results are presented in the last five columns. The "Both", "Resp", and "Reher" columns are the probability of success on both co-primary endpoints, the responder analysis, and the reherniation analysis respectively. The probability of early success is reported in the Pr(E) column. The mean and standard deviation of the trial sample size is reported in the "Mean N" and "SD N" columns.

Scenario E: In this scenario, for the responder analysis, the 3-month result has the reported 24-month probability and there is no change in this value for the 6-, 12-, and 24-month results. This scenario simulates a case where the 3-month value is a perfect predictor of the 6-month result.

Case	Control		Treatment		Both	Only Resp	Only Reher	Early	Mean N	SD N
	no RH	Pr(R)	no RH	Pr(R)						
1	0.90	0.65	0.90	0.65	0.005	0.010	0.008	0.001	604.5	83.0
2	0.90	0.65	0.96	0.65	0.030	0.000	0.181	0.004	679.7	81.7
3	0.90	0.65	0.90	0.85	0.028	0.233	0.000	0.002	650.4	102.2
4	0.90	0.65	0.96	0.70	0.312	0.004	0.223	0.106	724.8	90.2
5	0.90	0.65	0.96	0.75	0.719	0.019	0.034	0.404	748.3	82.7
6	0.90	0.65	0.96	0.80	0.816	0.030	0.000	0.495	748.1	79.5
7	0.90	0.65	0.96	0.85	0.830	0.036	0.000	0.492	742.8	82.1
8	0.90	0.65	0.92	0.85	0.159	0.294	0.000	0.017	691.7	105.0
9	0.90	0.65	0.94	0.85	0.498	0.186	0.000	0.135	727.9	96.7
10	0.90	0.65	0.98	0.85	0.969	0.000	0.000	0.909	721.1	76.2
11	0.90	0.65	0.96	0.75	0.426	0.137	0.013	0.120	716.1	101.2

Table 5-E: The operating characteristics for the default cases. The true values assumed for the reherniation-free rate (no RH) and responder rate (Pr(R)) in each treatment group are reported in the first four columns. A set of simulation results are presented in the last five columns. The "Both", "Resp", and "Reher" columns are the probability of success on both co-primary endpoints, the responder analysis, and the reherniation analysis respectively. The probability of early success is reported in the Pr(E) column. The mean and standard deviation of the trial sample size is reported in the "Mean N" and "SD N" columns.

Scenario F: In this scenario, for the responder analysis, the 3-month, 6-month, and 12-month result each has the reported 24-month probability of success. The result from each time point is independent of each other time point.

Case	Control		Treatment		Both	Only Resp	Only Reher	Early	Mean N	SD N
	no RH	Pr(R)	no RH	Pr(R)						
1	0.90	0.65	0.90	0.65	0.004	0.008	0.009	0.000	603.5	81.1
2	0.90	0.65	0.96	0.65	0.022	0.000	0.207	0.004	679.9	83.1
3	0.90	0.65	0.90	0.85	0.031	0.236	0.000	0.001	651.8	103.1
4	0.90	0.65	0.96	0.70	0.257	0.003	0.232	0.088	717.2	90.4
5	0.90	0.65	0.96	0.75	0.661	0.015	0.038	0.364	740.9	85.4
6	0.90	0.65	0.96	0.80	0.797	0.027	0.000	0.479	746.9	81.2
7	0.90	0.65	0.96	0.85	0.832	0.032	0.000	0.493	743.8	81.9
8	0.90	0.65	0.92	0.85	0.159	0.293	0.000	0.016	690.7	105.7
9	0.90	0.65	0.94	0.85	0.493	0.182	0.000	0.131	727.0	96.5
10	0.90	0.65	0.98	0.85	0.969	0.000	0.000	0.908	724.1	75.3
11	0.90	0.65	0.96	0.75	0.411	0.119	0.012	0.116	711.6	101.3

Table 5-F: The operating characteristics for the default cases. The true values assumed for the reherniation-free rate (no RH) and responder rate (Pr(R)) in each treatment group are reported in the first four columns. A set of simulation results are presented in the last five columns. The "Both", "Resp", and "Reher" columns are the probability of success on both co-primary

endpoints, the responder analysis, and the reherniation analysis respectively. The probability of early success is reported in the Pr(E) column. The mean and standard deviation of the trial sample size is reported in the "Mean N" and "SD N" columns.

Scenario G: In this scenario the probability of missing is assumed to be 0.

Case	Control		Treatment		Both	Only Resp	Only Reher	Early	Mean N	SD N
	no RH	Pr(R)	no RH	Pr(R)						
1	0.90	0.65	0.90	0.65	0.007	0.010	0.007	0.002	491.0	108.3
2	0.90	0.65	0.96	0.65	0.040	0.000	0.079	0.022	535.3	126.5
3	0.90	0.65	0.90	0.85	0.047	0.155	0.000	0.011	590.6	137.9
4	0.90	0.65	0.96	0.70	0.363	0.002	0.136	0.254	640.0	145.2
5	0.90	0.65	0.96	0.75	0.819	0.016	0.026	0.674	676.1	128.4
6	0.90	0.65	0.96	0.80	0.932	0.020	0.001	0.784	657.0	130.3
7	0.90	0.65	0.96	0.85	0.938	0.023	0.000	0.780	655.2	133.0
8	0.90	0.65	0.92	0.85	0.230	0.244	0.000	0.085	664.8	138.4
9	0.90	0.65	0.94	0.85	0.632	0.158	0.000	0.362	705.3	125.1
10	0.90	0.65	0.98	0.85	0.998	0.000	0.000	0.979	538.4	106.0
11	0.90	0.65	0.96	0.75	0.566	0.122	0.010	0.325	688.9	136.2

Table 5-G: The operating characteristics for the default cases. The true values assumed for the reherniation-free rate (no RH) and responder rate (Pr(R)) in each treatment group are reported in the first four columns. A set of simulation results are presented in the last five columns. The "Both", "Resp", and "Reher" columns are the probability of success on both co-primary endpoints, the responder analysis, and the reherniation analysis respectively. The probability of early success is reported in the Pr(E) column. The mean and standard deviation of the trial sample size is reported in the "Mean N" and "SD N" columns.

Additional Operating Characteristics

For each simulation done the software output is included as a separate file. The software output reports additional simulation results for each simulated scenario in this report. Each scenario is reported in a file, with the file name XS.txt reporting the results, where X is the sensitivity label (0 is for default) and S is the scenario number (1, 2, …, 11).

Lost-to-Follow-Up: The primary analysis will be done on the full mITT group with imputation for missing data as described below.

Responder Analysis

For the responder primary analysis those subjects who withdraw or who are lost (LTFU) after randomization will be included in the analysis using Bayesian imputation. The same technique which is used for the predictive distribution in the interim analysis will be used to model those subjects LTFU. This analysis will be the primary analysis, but recognizing that there is no way, statistically, to handle these subjects without possibly introducing bias, the following additional steps will be taken.

a) Patients will be classified by reason for missingness. Patients will be classified

as never treated, but randomized (type A, only applicable for ITT sensitivity analysis), patients who drop from the study or do not have follow-up visits when their timing dictates they should (Type B), and patients who naturally have not reached their 24-month follow-up visit because of an early analysis (Type C). Results will be classified by treatment group for type of missingness and the interim results.

b) Covariates will be included to best aid in the Bayesian imputation. The following covariates will be addressed: Age, race, gender, BMI, smoking status, length of conservative care, nucleus removed, baseline neurological status, baseline ODI, baseline leg pain VAS, baseline back pain VAS, baseline disc height, T score, amount of bone removed during the procedure, type of annulus defect and geometry of defect and defect size.
c) Extreme cases will be presented in sensitivity analysis to capture the full range of possibilities.

We propose to use a logistic structure for the secondary multiple imputation. The following structure for the probability of 24-month success (p) will be used:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_T T + \sum_{j=1}^{6} \lambda_j \delta_j ,$$

where T is an indicator of the experimental treatment arm (T=0 is control and T=1 is treatment). There are six indicators for interim results. These are $\delta_1, \ldots, \delta_6$, corresponding to 3-month success, 6-month success, 12-month success, 3-month failure, 6-month failure, and 12-month failure, respectively. The intercept β_0, therefore, represents the log-odds for the probability of success for a subject with the control treatment and no interim data (this is the seventh possible state for a subject, which for identifiability does not have an indicator). Each of the parameters, β_0, β_T, $\lambda_1, \ldots, \lambda_6$, will have independent $N(0,10^2)$ prior distributions.

For the sensitivity analyses which includes covariates the following structure will be used, where Z_j represents the jth covariate of the k covariates for a subject (continuous covariates will be standardized):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_T T + \sum_{j=1}^{6} \lambda_j \delta_j + \sum_{j=1}^{k} \alpha_j Z_j .$$

The prior distributions for the coefficients of the covariates will be $N(0,10^2)$. The parameter representing the possible bias for subjects with missing data will be ϕ. This will be incorporated in the model as:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_T T + \sum_{j=1}^{6} \lambda_j \delta_j + \sum_{j=1}^{k} \alpha_j Z_j + \phi.$$

The value of ϕ will be varied from $-\infty$ to $+\infty$, to represent subjects with missing data are all failures to all successes, respectively. Interim values will also be selected to capture the sensitivity to the possible bias in the missing data. An additional sensitivity analysis will be to employ ϕ_0 and ϕ_1, which correspond to the possible bias, separately within each treatment (0 for control and 1 for the treatment). Extreme cases, including when ϕ_0 and ϕ_1 vary separately from $-\infty$ to $+\infty$, will be included.

Reherniation Analysis

For the reherniation primary analysis those subjects who withdraw or who are lost (LTFU) after randomization will be included in the analysis using Bayesian imputation. The same technique which is used for the predictive distribution in the interim analysis will be used to model those subjects LTFU. This analysis will be the primary analysis, but recognizing that there is no way, statistically, to handle these subjects without possibly introducing bias, the following additional steps will be taken.

a) Patients will be classified by reason for missingness. Patients will be classified as never treated, but randomized (type A, only applicable for ITT sensitivity analysis), patients who drop from the study or do not have follow-up visits when their timing dictates they should (Type B), and patients who naturally have not reached their 24-month follow-up visit because of an early analysis (Type C). Results will be classified by treatment group for type of missingness and the interim results.
b) Covariates will be included to best aid in the Bayesian imputation. The following covariates will be addressed: Age, race, gender, BMI, smoking status, length of conservative care, nucleus removed, baseline neurological status, baseline ODI, baseline leg pain VAS, baseline back pain VAS, baseline disc height, T score, amount of bone removed during the procedure, type of annulus defect and geometry of defect and defect size.
c) Extreme cases will be presented in sensitivity analysis to capture the full range of possibilities.

We propose to use a logistic structure for the secondary multiple imputation. The following structure for the probability of 24-month success (p) will be used:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_T T + \sum_{j=1}^{3} \lambda_j \delta_j,$$

where T is an indicator of the experimental treatment arm (T=0 is control and T=1 is treatment). There are three indicators for interim results. These are δ_1, δ_2, δ_3, corresponding to 3-month reherniation free, 6-month reherniation free, and 12-month reherniation free. The intercept β_0, therefore, represents the log-odds for the probability of reherniation free for a subject with the control treatment and no interim data (this is the fourth possible state for a subject, which for identifiability does not have an indicator). Each of the parameters, β_0, β_T, λ_1, λ_2, λ_3, will have independent $N(0,10^2)$ prior distributions.

For the sensitivity analyses which includes covariates the following structure will be used, where Z_j represents the jth covariate of the k covariates for a subject (continuous covariates will be standardized):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_T T + \sum_{j=1}^{3} \lambda_j \delta_j + \sum_{j=1}^{k} \alpha_j Z_j \, .$$

The prior distributions for the coefficients of the covariates will be $N(0,10^2)$. The parameter representing the possible bias for subjects with missing data will be ϕ. This will be incorporated in the model as:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_T T + \sum_{j=1}^{3} \lambda_j \delta_j + \sum_{j=1}^{k} \alpha_j Z_j + \phi \, .$$

The value of ϕ will be varied from $-\infty$ to $+\infty$, to represent subjects with missing data as all failures to all successes, respectively. Interim values will also be selected to capture the sensitivity to the possible bias in the missing data. An additional sensitivity analysis will be to employ ϕ_0 and ϕ_1, which correspond to the possible bias, separately within each treatment (0 for control and 1 for the treatment). Extreme cases, including when ϕ_0 and ϕ_1 vary separately from $-\infty$ to $+\infty$, will be included.

The imputed values of success and failure for each subject will be done using Markov chain Monte Carlo (MCMC) techniques. These results will be integrated over to find the posterior probability of superiority. Traditional collinearity methodology investigating the correlation between different coefficients in the Bayesian imputation sensitivity modeling will be carried out. The correlations between coefficients will be calculated. Additionally SAS (PROC REG with options VIF and TOL) will be used to calculate tolerances and variance inflation factors within frequentist fitted models.

Secondary Endpoints and Analyses: A gatekeeping strategy is used to test the secondary endpoints for a claim of superiority. A Bayesian test of superiority will be done on the following endpoints, in the following order:

1. VAS Back Pain Improvement. Rates of subject success in treatment and control will be compared at 24 months. A subject is a success if there is at least a 20 point improvement on the VAS back at 24 months relative to baseline.
2. Oswestry Improvement. Rates of subject success in treatment and control will be compared at 24 months. A subject is a success if there is at least a 15 point improvement in the ODI at the 24 month visit relative to baseline.
3. Reoperation. A subject will be deemed a success if they have not had a second operation at the index level by the 24 month visit.
4. Disc Height Maintenance. Rates of subject success in treatment and control will be compared at 24 months. A subject is considered a success if there is at least 75% of the pre-op disc height preserved at the 24-month visit.
5. VAS Leg Pain Improvement. Rates of subject success in treatment and control will be compared at 24 months. A subject is a success if there is at least a 20 point improvement on the VAS Leg Pain in the ipsilateral leg at 24 months relative to baseline.

If the posterior probability of superiority is at least 95% for a secondary analysis then superiority will be claimed. If superiority is claimed for VAS back pain improvement then a test for superiority will be done for Oswestry improvement (95% posterior probability). If superiority is found for Oswestry improvement then a test for superiority will be done for reoperation (95% posterior probability). If superiority is found for reoperation then a test for superiority will be done for disc height maintenance (95% posterior probability). If superiority is found for disc height maintenance then a test for superiority will be done for VAS leg pain improvement (95% posterior probability).

For each of the claims of superiority above the analysis is identical to the primary analysis of the study of responder analysis. The analysis will be done on the mITT group, with Bayesian imputation for the analysis (identical to the primary analysis of responder).

Additional secondary endpoints will be reported for those subjects in the mITT group with complete 24 month follow up data and presented without claim. These analyses will be described using 2-sided Bayesian 95% credible intervals. Secondary endpoints will be analyzed for superiority and non-inferiority. Non-inferiority will be assessed using delta = 10%. These additional endpoints include:

- Composite endpoint success. The rate of success in the primary composite endpoint will be compared at each follow-up, up to and including the 12 month follow-up.
- Neurological maintenance/improvement as defined as the rate of subjects that have maintained or improved the neurological status. Analysis will be performed at each follow-up relative to baseline.

- The rate of individual subject success in terms of improvement in VAS leg pain. A successful subject will have at least a 20 point improvement in ipsilateral leg pain relative to baseline. The rate of subject success will be analyzed at each follow-up.
- The rate of individual subject success in terms of improvement in Oswestry Disability Index. A successful subject will have at least a 15 point improvement relative to baseline. The rate of subject success will be analyzed at each follow-up.
- The rate of individual subject success in terms of disc height maintenance. A successful subject will maintain at least 75% of their pre-operative disc height. The rate of subject success will be analyzed at each follow-up.
- The rate of individual subject success in terms of improvement in VAS back pain. A successful subject will have at least a 20 point improvement relative to baseline. The rate of subject success will be analyzed at each follow-up.
- Reoperations at the index level. The total number of reoperations up to and including the 24 month follow-up will be analyzed. Separately, analyses will also be performed on the total number of reoperations performed specifically for recurrent disc herniation and those performed for symptoms unassociated with recurrent disc herniations.
- Rate of ipsilateral recurrent disc herniation (at original defect, i.e. same side, same level)
- Rate of secondary recurrent herniation (but not at the original defect, i.e. contra-lateral herniations at same level)
- % Improvement in VAS leg pain in the ipsilateral leg. Analysis will be performed at each follow-up relative to baseline.
- % improvement in VAS back pain. Analysis will be compared at each follow-up relative to baseline.
- % improvement in function as judged by the Oswestry Disability Index. Analysis will be performed at each follow-up relative to baseline.
- Mean VAS back pain. Analysis will be performed at each follow-up.
- Mean VAS leg pain in the ipsilateral leg. Analysis will be performed at each follow-up
- Mean Oswestry Disability Index. Analysis will be performed at each follow-up.
- Improvement in Quality of Life as judged by SF-36. The bodily pain and physical function scores will be independently analyzed at each follow-up relative to baseline.
- Adverse events. The total number of adverse events up to and including the 24 month follow-up will be compared between treatment and control, as well as the number of intra-operative and post-operative adverse events, as well as the adverse event rates by severity. Individual adverse event rates will also be compared.

- Post operative pain medication use. The rate of subjects using narcotic medication to manage back and/or leg pain will be analyzed at each follow-up.
- Economic cost as judged by direct medical expenses post-operatively by 1 year and 2 years.
- The rate of subjects returning to work without restriction will be analyzed at each follow-up time point.
- The time from surgery to return to work

Safety Analyses: For each safety analysis the outcome will be an event (possibly composite). Two-sided Bayesian 95% credible intervals will be used for this analysis.

Site Heterogeneity: Summaries of the demographics and baseline clinical characteristics by investigational site will be presented. Additionally summaries of these variables will be made by continent of the investigational site. A Breslow-Day test for homogeneity of odds ratios will be conducted on a per site basis, and a per continent basis, for each of the co-primary endpoints. Additionally a Breslow-Day test will be conducted for the number of AE's per site. Sites with less than five subjects will be combined to one mega-site. If the results of the test show evidence of a lack of heterogeneity then a hierarchical model will be created to model the success rate per site as the hierarchical component.

The hierarchical model for heterogeneity is specified as follows. Let s be an indicator of the investigational site and t the indicator of treatment (t = 1 is *treatment* and t = 2 is *control*). The probability of a responder is modeled as

$$\log\left(\frac{p}{1-p}\right) = \lambda_s + \theta_t .$$

The restriction is made that $\theta_2 = 0$. The prior distribution for θ_1 is $N(0,10^2)$. The distribution of study effects, λ, are modeled hierarchically. The model is

$$\lambda_s \sim N\left(\mu, \tau^2\right),$$

where the prior distributions for the hyper-parameters are

$$\mu \sim N\left(0,10^2\right)$$

and

$$\tau^2 \sim G^{-1}\left(0.01, 0.01\right).$$

This first hierarchical model incorporates additive heterogeneity across the study sites. Additionally a model with a study specific interaction with site will be presented. The model is

$$\log\left(\frac{p}{1-p}\right) = \lambda_s + \theta_{ts}.$$

The restriction is made that $\theta_{2s} = 0$. The hierarchical distribution of treatment effect distributions is

$$\theta_{1s} \sim N\left(\delta, \gamma^2\right)$$

with hyper-priors of

$$\delta \sim N\left(0, 10^2\right)$$

and

$$\gamma^2 \sim G^{-1}\left(0.01, 0.01\right).$$

Device Generation Poolability Analysis: At the completion of the trial, an analysis of poolability will be conducted to determine the validity of combining the data on the Generation 2 and Generation 3 versions of the device. Summaries of the demographics and baseline clinical characteristics by device generation will be presented. We will test whether the treatment effect for either endpoint varies by generation of device with a logistic regression model that includes terms for treatment, generation of the device and the interaction between the two. If, for either of the co-primary endpoints, the generation 2 and 3 treatment effects (i.e. the interaction term) are deemed significantly different at the 5% level, an analysis of the generation 3 alone (G3) population will be conducted and presented for the primary and key secondary endpoints. If no statistically significant interaction is found, the mITT pooled population will be utilized for determination of study success.