

April 13, 2004

In-depth Statistical Review for Expedited PMA (P040006) Charité Artificial Disc, DePuySpine, Inc. (dated Feb. 23, 2004)

I. Background

The Charité Artificial Disc (Charité) is indicated for spinal arthroplasty in skeletally mature patients with degenerative disc disease (DDD) at one level from L4-S1. DDD is defined as discogenic back pain with degeneration of the disc confirmed by patient history and radiographic studies. As an alternative to spinal fusion permitting near physiological segment movement, the device is a weight-bearing modular implant consisting of two cobalt-chromium alloy endplates and one ultra-high molecular weight polyethylene sliding core. The bi-convex core articulates between the two concave endplates. Charité has been used outside the United States since 1987 and has not been withdrawn from the market for any reason. More than 7,000 patients have been implanted with the disc worldwide including the U.S. IDE and Continued Access patients (see below).

In order to evaluate the effectiveness and safety of the device, a pivotal clinical study was conducted under an approved IDE G990303 (Phyllis Silverman was the statistical reviewer). Since Charité is a new therapeutic option for treatment of DDD to preserve function in the lumbar vertebral region, this PMA was granted expedited review status, which was confirmed by the Agency at a pre-PMA meeting on Nov. 14, 2003 (Yuan Who Chen was the statistician at the meeting). During this pre-PMA meeting, the sponsor presented the initial study results showing that even as early as 6 weeks, a statistically significantly higher proportion of Charité patients already had achieved 25% improvement in the Oswestry Disability Index (ODI) score compared to BAK group. The sponsor further explained that the data appears to show consistently greater improvement in the Charité group at earlier time points, while the data at 24 months support non-inferiority claim compared to BAKTM Interbody Fusion System (BAK).

This memorandum reviews the results of the pivotal study, which is a prospective, randomized multi-center study comparing Charité and BAK.

II. Summary of the Clinical Study

1. Study Objectives

The objective of the study is to evaluate the safety and effectiveness of Charité compared with that of BAK for the treatment of single-level DDD in patients without prior fusion or other spinal surgery, except prior discectomy, laminotomy/ectomy (without accompanying facetotomy), or nucleolysis at the same level to be treated.

2. Study Design

This investigative study was a multi-center, prospective, randomized, controlled clinical study. The first five subjects enrolled at each center were to receive Charité as training cases. Subsequent subjects were randomized in a 2:1 ratio of Charité recipients to BAK recipients. Both devices were to be implanted via an anterior approach to ensure comparability. Blocking techniques were used to ensure balance between the treatment groups at each center. Neither the subject nor investigator could be blinded to the treatment. All radiographic data was to be evaluated independently by a core laboratory. Functional (neurological) status was to be evaluated by a blinded evaluator (MD, RN or a PA) at follow-up times after 3 months (i.e., 6, 12 and 24 months).

Study Hypothesis:

The statistical hypothesis is one of equivalence to the control (BAK), where equivalence is defined such that the success rate of Charité (p_e) is no worse than that of BAK (p_s) by an acceptable margin ($d = 15\%$), using the definition of individual patient success set forth on pages 24-25 of Vol. 2 (see page 3 of this review: primary effectiveness endpoint):

$$H_o: p_s - p_e > d$$

$$H_a: p_s - p_e = d$$

With the assumptions of 70% success rates in both Charité and BAK groups and a 10% drop-out rate, the study requires 291 subjects (194 Charité, 97 BAK) to achieve 80% power at a one-sided 0.05 significance level. Taking into account the five training subjects for Charité at each center, a total of 366 subjects (269 Charité, 97 BAK) were to be enrolled at 15 sites.

Patient population:

Patients of 18-60 years old who had single-level DDD (L4/L5 or L5/S1) but no previous thoracic or lumbar fusion and have met the other inclusion/exclusion criteria as listed on pages 16-18 of Vol. 2 were to be enrolled in the study.

Treatment duration and follow-up:

Each patient was to remain in the study for 24 months post-implantation. The study duration was to comprise the pre-treatment, intra-operative, and immediate post-operative periods, followed by evaluations 6 weeks, 3 months, 6 months, 12 months and 24 months.

Effectiveness endpoints:

The primary effectiveness endpoint is individual success outcome, a composite endpoint of measures for both effectiveness and safety as defined by all the following: (1) improvement of at least 25% in the Oswestry Disability Index (ODI) at 24 months compared with the score at baseline; (2) no device failures requiring revision, re-operation or removal; (3) absence of major complications, defined as major vessel injury, neurological damage, or nerve root injury; (4) maintenance or improvement in neurological status at 24 months with no permanent neurological deficits compared to baseline status.

The secondary endpoints include: (1) pain relief, defined as improvement of at least 20 mm on a 100 mm visual analog scale (VAS) at 24 months compared to baseline; (2) improvement of quality of life defined as improvement of 15% in the overall score using the short-form-36 questionnaire (SF-36) at 24 months compared to baseline; (3) disc height measured by standard lateral radiograph (only changes of more than 3 mm will be considered clinically significant); (4) displacement or migration of the device (changes of > 3mm significant); (5) no significant radiolucency for Charité at 24 months compared with post-operative radiolucency; (6) components of the primary endpoint (e.g., ODI score and neurological score).

Safety endpoints:

The primary safety endpoint is the incidence of all device related adverse events (AEs) or complications throughout the course of the study, which may include implant breakage, component degradation, implant displacement, pain, spinal instability, injury to kidneys or ureters, vessel damage/bleeding, deterioration in neurological status, infection, etc. (see pages 162-167 of Vol. 14 for the full list).

Analysis Plan:

There was no statistical analysis plan (SAP) proposed in the original IDE protocol (five versions, Vol. 14: Pages 9-180). Only after the Agency's request during the pre-PMA meeting on Nov. 14, 2003, the sponsor's consultant statisticians (Kathie Drouin and George DeMuth) from Stattech Services, LLC provided the SAP via email on Nov. 25, 2003. The following analysis plan was based on this SAP, dated March 27, 2003, but appeared to be originated on Oct. 18, 2001 and finalized on Oct. 27, 2003. **Please note that it is not clear to me when this SAP was proposed and whether the statisticians at Stattech Services, LLC were blinded to the data access before developing the SAP.**

Analysis Populations: The Intent-to-Treat (ITT) population consists of all patients who were randomized to the study. These subjects were to be included in the group to which they were assigned regardless of the treatment they actually received. The ITT population is the basis of the effectiveness analysis. Patients who were randomized and actually received treatment will constitute the population for safety analysis. These patients will be included in the treatment group based on the actual treatment received.

Primary effectiveness analysis: The success outcome (i.e., the composite endpoint) of subject was to be summarized by treatment groups using counts and percentages. Blackwelder's test with the delta of 15% was to be used for comparing the difference in success rates between the two groups. A one-sided 95% confidence interval for the difference was also to be constructed to assess the non-inferiority of Charité to BAK. In addition, a Cochran-Mantel-Haenszel (CMH) test stratified by center was to be included.

Secondary effectiveness analyses: Each component of the primary effectiveness endpoint was to be tested individually as a secondary endpoint. Pearson's Chi-square test was to be used for ODI and neurological status. Kaplan-Meier estimates of time to device failure and time to major complication were to be calculated. The effects of some covariates including age category (greater or less than 45 years), gender, device configuration and site were to be examined using PROC CATMOD with response logit. Other covariates of interest

including presence of osteoporosis, use of Hormone Replacement Therapy (HRT) could be added to the model as needed.

Secondary endpoints such as pain VAS Score, total Oswestry Score, total quality of life score (SF-36), displacement or migration of the device and their respective change from baseline scores were to be summarized at each time point by treatment group. For continuous variables, a Student's paired t-test was to be used for within-treatment comparison in change from baseline and a Student's non-paired t-test was to be conducted to test for differences between treatments at 24 months. Categorical variables such as displacement or migration of the device with change of more than 3mm, change in disc height of at least 3 mm, absence of major complications and absence of device revisions, re-operations or removal, pain and quality of life improvement status at 24 months were to be summarized by counts and percentages. Chi-square test or Fisher's exact test and CMH adjusted for study site were to be used for the comparison between treatment groups.

Safety analysis: All AEs were to be summarized under each treatment group by tabulating the numbers and percentages of subjects reporting each event by decreasing incidence. Counts and percentages of subjects with a particular AE by severity and treatment group and by relationship to device by treatment group were to be tabulated. Unanticipated adverse device effects (UADEs) were to be tabulated or listed depending on the number of subjects with UADEs. AEs that occurred within 48 hours of surgery were to be summarized in an attempt to identify procedure related events. In addition, duration of hospital stay, presence of radiolucency, formation of longitudinal ossification from endplates, migration of prosthesis more than 3 mm, decrease in disc space more than 3 mm and pseudoarthrosis were to be summarized.

3. Study Results

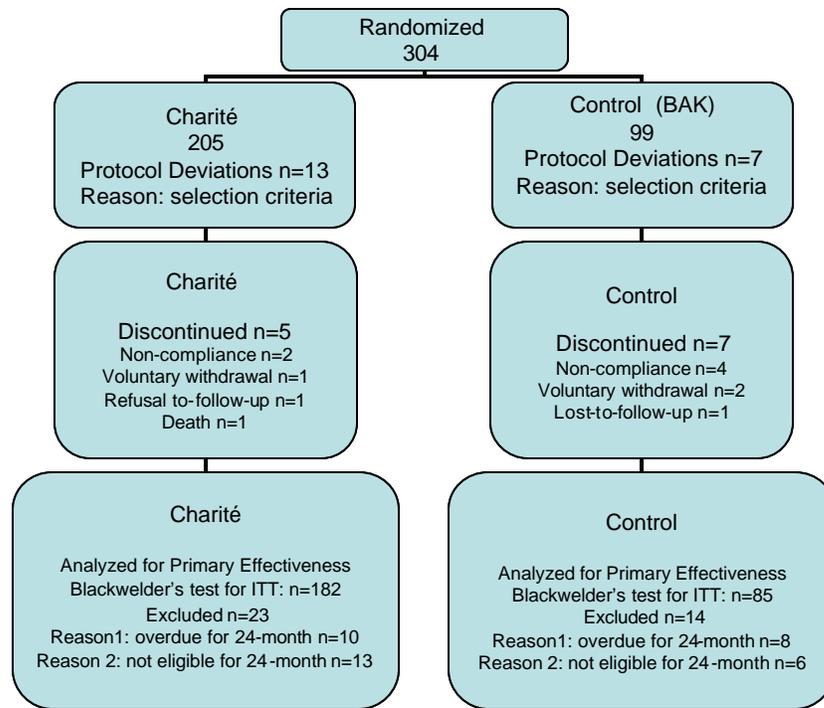
Patient disposition

A total of 304 patients were randomized at 14 sites from May 16, 2000 to April 24, 2002 (Table 1). Enrollment among the 14 sites varied from 7 to 77 subjects with 7 sites of 17 or more subjects. As shown in Figure 1, there was a higher rate of early discontinuations (less than 8 weeks in study) in BAK patients (n=7, 7%) compared to Charité patients (n=5, 2%). All discontinued patients were included in the ITT analysis for the primary effectiveness analysis, but a total of 23 (11%) Charité patients and 14 (14%) BAK were excluded from the ITT analysis because those excluded patients were overdue for the 24 month follow-up or have not reached the 24-month follow-up. There was a comparable proportion of protocol deviations (BAK n=7, 7.1% vs. Charité n=13, 6.3%) with the violation of inclusion/exclusion criteria as the only reason. No patient with protocol deviations was excluded from the ITT analysis for the primary effectiveness analysis.

Table 1: Subject Enrollment

Characteristic	Total
Number of randomized subjects	304
Charité	205
BAK	99
Number of training subjects	71
Number of continuing access subjects	71
Total number of subjects enrolled	446

Figure 1: Patient Accountability



Ref. Table 11, Vol.2, page 31 and Table 19, Vol. 2, page 39

Patient demographics and baseline characteristics:

Patient demographics were summarized (Tables 2a) for the “ITT population” (i.e., all the randomized patients excluding those overdue or non-eligible patients (Charité 23 vs. BAK 14).

Table 2a: Demographics

	Charité n=182 Excluded n=23	BAK n=85 Excluded n=14	P-value*
Gender			0.15
Female	99 (54%)	38 (45%)	
Male	83 (46%)	47 (55%)	
Race			0.43
Caucasian	169 (93%)	75 (88%)	
African American	6 (3%)	4 (5%)	
Min – Max	7 (4%)	6 (7%)	
Age (years)			0.64
Mean (SD)	39.5 (8.1)	40.1 (9.43)	
Median	40.0	41.0	
Min – Max	19 – 60	20 – 60	
Age Categories			0.07
> 45 years	41 (23%)	28 (33%)	
= 45 years	141 (77%)	57 (67%)	
Height (cm)			0.38
Mean (SD)	172.4 (9.50)	173.5 (10.06)	
Median	171.5	172.7	
Min – Max	152 - 201	155 - 196	
Weight (kg)			0.01
Mean (SD)	77.3 (15.98)	82.5 (16.74)	
Median	77.1	81.2	
Min – Max	46 - 120	54 - 122	
BMI			0.01
Mean (SD)	25.9 (4.19)	27.4 (4.77)	
Median	25.8	27.4	
Min – Max	17 – 39	18 – 40	
Level of Disc Disease			0.57
L4/L5	53 (29%)	28 (33%)	
L5/S1	129 (71%)	57 (67%)	
* Fisher’s exact test for categorical variables and t-test for continuous variables			
Ref. Appendix 1 –Table 9.1			

Overall, the demographics of the “ITT population” were comparable between the two treatment groups. There were more women in the Charité group (54%) than in the BAK group (45%). Also, the Charité group had more patients younger than 45 years old (Charité 77% vs. BAK 67%). The mean of Body Mass Index (BMI) was higher in BAK group (27.4) compared to the Charité group (25.9), which was consistent with the findings that the mean weight of BAK patients (82.5 kg) was larger than that of the Charité patients (77.3 kg) while the mean heights between the two groups were very similar (Charité 172.4 cm vs. BAK 173.5 cm).

Table 2b: Baseline Evaluation

	Charité n=182 Excluded n=23	BAK n=85 Excluded n=14	P-value*
Medical History			
<i>Gait</i>			0.31
Normal	131 (72%)	56 (66%)	
Abnormal	51 (28%)	29 (34%)	
<i>Normal Activity Level before back injury</i>			
			0.23
Active	167 (92%)	73 (86%)	
Moderate	13 (7%)	10 (12%)	
Light	1 (1%)	2 (2%)	
Minimal	1 (1%)	0	
<i>Pre-Operative Activity</i>			
			0.02
Active	9 (5%)	0	
Moderate	25 (14%)	5 (6%)	
Light	48 (26%)	23 (27%)	
Minimal	100 (55%)	57 (67%)	
Concomitant Disease (>3%)			
Hypertension	15 (8%)	12 (14%)	
Asthma	11 (6%)	6 (7%)	
Hepatitis	5 (3%)	6 (7%)	
Osteoarthritis	7 (4%)	3 (4%)	
Anemia	6 (3%)	4 (5%)	
Peptic Ulcer	6 (3%)	3 (4%)	
Cancer	2 (1%)	3 (4%)	
Other	80 (44%)	33 (39%)	
* Fisher's exact test for categorical variables and t-test for continuous variables			
Ref. Appendix 1–Table 11.1, Table 15.1			

With respect to the baseline characteristics (i.e., medical history and concomitant diseases, Table 2b) of the “ITT population”, the two treatment groups were also comparable except that there were much more active subjects just prior to surgery in the Charité group than in the control group (Charité 9, 5% vs. BAK 0, 0%). In addition, the “ITT subjects” were similar between the two groups with regard to menopausal status, prior therapy, risk to bone healing and surgical procedure related characteristics such as total surgery time and blood loss (Appendix 1, Tables 12.1-16.1).

The demographics and baseline characteristics of all the randomized patients (i.e., Charité 205 vs. BAK 99) were similar to the “ITT population”(Appendix 1, Table 9.2, Vol. 3, page 25), although all the randomized patients tend to be more balanced between the two groups with respect to age categories, weight, BMI and pre-operative activity level compared to the “ITT population”.

Sponsor's Statistical Evaluations and Results:

?? Effectiveness

- *Primary endpoint:*
 - a. *Unadjusted analysis (Blackwelder's Test):* Overall success at 24-month for the "TTT population" was 63% (114/182) for the Charité subjects and 53% (45/85) for the BAK subjects ($p < 0.0001$, Appendix- Table 19). **Please note that all discontinued patients were treated as failures** (Charité 5, 3% vs. BAK 7, 8%) **and that patients who were either overdue for the 24-month follow-up** (Charité 10, 5% vs. BAK 8, 8%) **or had not reached the 24-month visit** (Charité 13, 6% vs. BAK 6, 6%) **were excluded from the "TTT population"**.
 - b. *Sensitivity analysis:* The sponsor performed a number of sensitivity analyses for the primary effectiveness endpoint. All sensitivity analyses supported the non-inferiority hypothesis. These included considering the impact of ongoing subjects and looking at alternative imputations for non-completers (Appendix- Table 21a, Vol. 3, pages 66-67). The first analyses consider last observation carried forward (LOCF) results for non-completers and LOCF with discontinuations as failures. An LOCF analysis was also provided that included the overdue subjects. LOCF was performed for all randomized subjects and all randomized subjects with discontinuations as failures. Across all of the analyses, the overall success rate in the Charité group ranged from 63% to 68% and the rate in the BAK group ranges from 48% to 54%. An analysis was also performed ignoring the neurological component of success (Appendix-Table 21b). This analysis benefited the Charité group.
 - c. *Repeated measures analysis:* The sponsor fitted a repeated measures model that included all randomized subjects using their available data (Appendix 1 - Table 22). For this analysis, the estimated response rates for the Charité group and the BAK group were 64% and 55%, respectively. A contrast used to test the hypothesis that the estimated response rates between the treatment groups were the same at all visits (Appendix 1- Table 22) was statistically significant ($p = 0.0082$). Two additional analyses suggested that the Charité group had a more favorable time course than the BAK group with non-inferiority established at the 24-month visit (Appendix 1, Table 24). A sustained response was defined as a response at 24 months and the time of the response was the first time point where success was observed and continued through 24 months. The response rates for the Charité group were 44%, 51%, and 63% at 6 months, 12 months, and 24 months, respectively. Similarly, the response rates for the BAK group were 35%, 41%, and 53% at 6 months, 12 months, and 24 months, respectively. Earlier time points were not considered because neurological evaluations were not completed at earlier follow-up times. The trend favored the Charité but was not statistically significant ($p = 0.1217$). In

summary, analyses of response over time are clearly supportive of overall non-inferiority.

- d. *Subgroup and covariate analyses:* The sponsor investigated the potential impact of a number of covariates on the primary effectiveness endpoint. These covariates were established prior to study completion or selected post-hoc based on observed treatment group differences in the distribution of the covariate. A univariate marginal probability model was considered (using PROC CATMOD) for each covariate, treatment, and the treatment by covariate interaction. In addition, the difference in the response rates between the two groups was evaluated and the difference was compared to the non-inferiority margin of 15% used in the Blackwelder's test. In all cases, the non-claim was supported regardless of the covariate considered.

As shown in Appendix 1-Tables 23.1 and 23.2, a number of covariate presented non-statistically significant effects at the 0.05 level for the covariate main effect or interaction. These non-significant covariates included Age (<45, >45), Baseline Oswestry, Gender, Operative Level, use of HRT, and use of Pain Medication at any time.

The following covariates were associated with the outcome either as a main effect or in the interaction term. "Body mass index" was a significant main effect (p=0.0441) but no treatment interaction was indicated with a trend toward the best results in the second quartile of subjects. "Current activity level" indicated a significant interaction term (p=0.0351), but not main effect. The association for "current activity level" seemed to present a trend towards better results in active subjects for the Charité group and the opposite in the BAK group. However, the Charité subjects had observed response rates as good as the BAK group for all levels. "Osteopenia" showed a significant interaction with treatment (p=0.0077) with Charité subjects in the subgroup performing better than BAK subjects. "Study site" had a significant main effect (p=0.0238) but did not show significant interaction effect. Equipment configuration was included in the covariate table but was not relevant to the BAK group. Configuration demonstrated a significant trend (p=0.0258) with Oblique endplates subjects appearing better than parallel only endplate subjects.

In summary, the covariate analyses remain supportive of the overall non-inferiority claim.

- o *Secondary endpoints:*
 - a. *Components of the primary endpoint:* As shown in Table 20, Vol. 2, page 41, the sponsor performed Fisher's Exact tests for each component of the primary outcome, indicating that there was no statistically significant difference in any of the components of success at 24 months among the "ITT" population. For all the randomized patients, similar results were obtained (Appendix 20.2, Vol. 3 Page 65). Student t-tests of ODI percent change from baseline suggested that the Charité group experienced significantly greater

improvement at 6 weeks (unadjusted $p=0.0485$), 3 months (unadjusted $p=0.0087$) and 6 months (unadjusted $p=0.0126$) followed by a period when the BAK group closed the gap between the treatments (Table 21, Vol. 2, Page 44). The proportion of patients with 25% improvement in ODI from baseline was significantly higher in the Charité group at 6 weeks compared to the BAK group (Fisher's exact tests, unadjusted $p=0.0269$), 3 months (unadjusted $p=0.0091$) and 6 months (unadjusted $p=0.0130$, Table 22, Vol.2 page 46). With respect to the neurological status, the occurrence of neurological deterioration at 6, 12 and 24 months was not significantly different between the two groups (both groups: $<10\%$, Table 23, Vol. 2, Page 47).

- b. *VAS pain score*: The changes from baseline in the VAS pain scores were statistically significant (paired t-tests, unadjusted $p<0.001$) at all time points (Appendix 1-Tables 30.1, Vol. 3, Pages 108-109). Furthermore, the mean change from baseline was statistically significantly better for the Charité Artificial Disc group at 6 months (unpaired t-test, unadjusted $p=0.0174$) with the BAK group narrowing the difference over time. Between 68% and 75% of the Charité subjects had at least a 20 mm improvement from baseline in their VAS pain scores at all time points as compared to 60% to 70% of subjects in the BAK group (Appendix 1-Table 31.1, Vol. 3, Pages 115-119). In the "ITT population", differences between treatment groups in the proportion of subjects with at least a 20 mm improvement were borderline statistically significant at 6 months (Fisher's exact test, unadjusted $p=0.0501$). At 24 months, 75% of Charité subjects and 70% of BAK subjects had at least a 20 mm improvement from baseline in their VAS pain scores. Similar results were obtained from the analyses for all randomized subjects (Appendix 1-Tables 30.2 and 31.2).
- c. *Quality of Life assessments (SF-36)*: For the "ITT population" (Appendix 1-Tables 32.1 and 33.1), there is a statistically significant improvement after the 3 month visit compared to baseline for the SF36-PCS (Physical Composite Score) and SF36-MCS (Mental Component Score) for both treatment groups. Even though both the Charité group and the BAK group started at the similar mean PCS scores (31.1 and 31.8, respectively), at each postoperative time point after 3 months, the Charité group was statistically significantly better than the BAK group in the physical health measure (t-tests, unadjusted p-values <0.05). At 24 months, the proportion of Charité subjects who had a 15% improvement from baseline in the SF-36 PCS was 73% (**99/136**), compared with 66% (**41/62**) for the BAK subjects (Chi-square test, unadjusted $p=0.3392$). Similar results were obtained for all randomized patients (Appendix 1-Tables 32.2 and 33.2).
- d. *Other secondary endpoints*: There was a very low incidence of disc space height loss ($=3\text{mm}$) in both groups (1%, Table 26, Vol. 2, page 50). The Charité patients showed a near-physiological mean range of motion (4.9, 6.0, 7.0 and 7.4 degrees at 3, 6, 12 and 24 months, respectively). Three patients in

the Charité group had device migration > 3 mm at 24 months whereas there was no BAK patients had device migration > 3 mm. Radiolucency was found for 1 patient at 12 months and 2 patients at 24 months in the Charité group.

?? Safety

The overall rate of subjects experiencing at least one AE during the 24-month study was similar between the two groups (Charité 156/205=76.1% vs. BAK 77/99=77.8%). A total of 10 Charité subjects (10/205=4.9%) and 8 BAK patients (8/99=8.1%) had device failures.

However, the Charité group had a higher rate of device-related AEs (15/205=**7.3%**) compared to the BAK group (4/99=**4.0%**). There was a total of **25** device-related AEs among the **205** Charité patients, whereas only a total of **5** device-related AEs among the **99** BAK patients. Please note that 24 infections (24/205=11.7%) in the Charité group and 6 infections (6/99=6.1%) in the BAK group were considered as **not device related** AEs. In addition, the Charité group had a higher rate of severe/life-threatening AEs (30/205=15%) compared to the BAK group (9/99=9%).

The sponsor did not perform any statistical tests to compare the safety profile between the two groups.

III. Comments

1. The sponsor did not pre-specify the principal features of any statistical analysis in the original IDE protocol (five versions dated from Feb 23, 2000 to Oct. 26, 2001, Vol.14, pages 9-180). Only after our request during the pre-PMA meeting on Nov. 14, 2003, the sponsor's consultant statisticians (i.e., Kathie Drouin and George DeMuth from Stattech Services, LLC) provided a statistical analysis plan (SAP) via email on Nov. 25, 2003. This SAP appeared to be finalized on Oct. 27, 2003 by which most trial data were probably available (patients were randomized from May 16, 2000 to April 24, 2002). Since unblinded development of SAP will generally introduce bias, the sponsor should clarify when the SAP was finalized and whether the SAP developers were blinded to the treatment assignment if a preliminary review of the data was conducted to modify the SAP.
2. The study was designed to randomize a total of 291 subjects (Charité 194, BAK 97) without any planned interim analysis. However, the data analysis for this current PMA was conducted before all the randomized patients completed the study (there were 13 remaining patients in the Charité group and 6 in the BAK group). The sponsor should clearly state that no interim analysis was performed to reach the decision of the early submission.
3. The sponsor performed sensitivity analyses for the primary endpoint (Appendix-Table 21a, Vol. 3, pages 66-67) to evaluate the impact of discontinuation (Charité 5, 2.5% vs. BAK 7, 8%), lost-to-follow-up (i.e., overdue: Charité 10, 5% vs. BAK 8, 8%) and early-stop (i.e., not-due: Charité 13, 6% vs. BAK 6, 6%). Last Observation Carry Forward (LOCF) was used to handle the missing data at 24-month for the overdue and not-due patients. Those discontinued patients were treated as all failures or LOCF. In order to assess the validity of LOCF, the sponsor should provide the following information:
 - a. summary information for the numbers of LOCF by missing type (i.e., discontinuation, overdue and not-due) and success/failure.
 - b. when was the last observation (12-month, 6-month) for each LOCF;
 - c. what was the percentage of patients succeeding at an earlier follow-up (12-month, 6-month) continued to succeed at the 24-month. Please note that time points earlier than 6 month cannot be considered for LOCF because neurological evaluations were not completed at earlier follow-up intervals (Vol. 2, page 40).

With regard to the first two items (a. and b.), I created the following Tables 3 and 4 based on the sponsor's Tables 20.1a, 20.1b and 21a (Appendix-1, Vol. 3, pages 61-67). Please ask the sponsor to verify or fill in the numbers and the last follow-up time points of LOCF as shown in Tables 3 and 4. Alternatively, the sponsor should provide a line listing of those patients (discontinued, overdue and not-yet-due) by treatment group who did not complete the study.

Table 3: Sensitivity analysis with overall LOCF success for discontinued, overdue or not-due

Analysis Set		Primary Endpoint	Charité	BAK	Sponsor's Blackwelder P-value	Upper Bound 1-sided 95% CI of $P_{BAK} - P_{charité}$
ITT	Completers	Failure	63 (36%)	33 (42%)	0.0005	0.042
		Success	114 (64%)	45 (58%)		
	Discontinued	Failure	4	6		
		Success	1	1		
Overdue		Failure	4	6		
		Success	6	2		
Not-due		Failure	4	3		
		Success	9	3		
All Randomized		Failure	75 (37%)	48 (48%)	<0.0001	-0.020
		Success	130 (63%)	51 (52%)		

Ref: Appendix Table 21a, Vol 3 Pages 66-67

Table 4: Components of the primary endpoint

Oswestry Disability Index

Analysis Set		25% improvement	Charité	BAK	Fisher's Exact Test P-value	Upper Bound 1-sided 95% CI of $P_{BAK} - P_{charité}$
ITT	Completers	No	50 (28%)	29 (37%)	0.1860	
		Yes	127 (72%)	49 (63%)		
	Discontinued	No	5	7		
		Yes	0	0		
Overdue		No	= 9?	8		
		Yes	= 1?	0		
Not-due		No	=12?	6		
		Yes	= 1?	0		
All Randomized		No	77 (38%)	50 (51%)	0.0354	
		Yes	128 (62%)	49 (49%)		

Device Failure

Analysis Set		Device Failure	Charité	BAK	Fisher's Exact Test P-value	Upper Bound 1-sided 95% CI of $P_{BAK} - P_{charité}$
ITT	Completers	Failure	7 (4%)	7 (9%)	0.1350	
		Success	170 (96%)	71 (91%)		
	Discontinued	Failure	1	1	0.1632	
		Success	4	6		
Overdue		Failure	= 1?	= 1?		
		Success	= 9?	= 7?		
Not-due		Failure	=1?	=1?		
		Success	= 12?	= 5?		
All Randomized		Failure	9 (4%)	9 (9%)	0.1219	
		Success	196 (96%)	90 (91%)		

Major Complications

Analysis Set		Major Complications	Charité	BAK	Fisher's Exact Test P-value	Upper Bound 1-sided 95% CI of $P_{BAK} - P_{charité}$
ITT	Completers	Failure	2 (1%)	1 (1%)	1.0000	
		Success	175 (99%)	77 (99%)		
	Discontinued	Failure	0	0		
		Success	5	7		
Overdue		Failure	0	0		
		Success	10	8		
Not-due		Failure	0	0		
		Success	13	6		
All Randomized		Failure	2 (1%)	1 (1%)	1.0000	
		Success	203 (99%)	98 (99%)		

Neurological Deterioration

Analysis Set		Neurological Deterioration	Charité	BAK	Fisher's Exact Test P-value	Upper Bound 1-sided 95% CI of $P_{BAK} - P_{charité}$
ITT	Completers	Failure	17 (10%)	4 (5%)	0.3239	
		Success	160 (90%)	74 (95%)		
	Discontinued	Failure	5	7		
		Success	0	0		
Overdue		Failure	?	?		
		Success	?	?		
Not-due		Failure	?	?		
		Success	?	?		
All Randomized		Failure	43 (21%)	24 (24%)	0.5559	
		Success	162 (79%)	75 (76%)		

"?": to be filled in or verified by the sponsor

Ref: Appendix Tables 20.1a, 20.1b and 20.2, Vol 3 Pages 61-65

Table 5: FDA's sensitivity analyses for the primary effectiveness endpoint				
1. Sponsor's "ITT subjects" =All randomized subjects - Overdue/Not-due subjects				
a. Discontinued Charite = 5 failure BAK= 7 failure				
	Success	Total	Suc/Total	UpB 90% CI of $P_{BAK}-P_{charité}$
Charité	114	182	62.6%	0.0099
BAK	45	85	52.9%	
b. Discontinued Charite = 5 failure BAK= 7 success				
	Success	Total	Suc/Total	UpB 90% CI of $P_{BAK}-P_{charité}$
Charité	114	182	62.6%	0.0905
BAK	52	85	61.2%	
2. True ITT subjects: All randomized subjects				
a. Discontinued Charite = 5 fail BAK= 7 suc Overdue/not-due Charite= 23 fail BAK= 14 suc				
	Success	Total	Suc/Total	UpB 90% CI of $P_{BAK}-P_{charité}$
Charité	114	205	55.6%	0.2072
BAK	66	99	66.7%	
b. Discontinued Charite = 5 fail BAK= 7suc Overdue/not-due Charite= 23 fail BAK= 14 fail				
	Success	Total	Suc/Total	UpB 90% CI of $P_{BAK}-P_{charité}$
Charité	114	205	55.6%	0.0695
BAK	52	99	52.5%	
c. Discontinued Charite = 5 fail BAK= 7 suc Overdue/not-due Charite= 12 suc, 11 fail; BAK= 14 suc				
	Success	Total	Suc/Total	UpB 90% CI of $P_{BAK}-P_{charité}$
Charité	126	205	61.5%	0.1479
BAK	66	99	66.7%	
d. Discontinued Charite = 5 fail BAK= 7 suc Overdue/not-due Charite= 23 fail BAK= 8 suc, 6 fail				
	Success	Total	Suc/Total	UpB 90% CI of $P_{BAK}-P_{charité}$
Charité	114	205	55.6%	0.1489
BAK	60	99	60.6%	

Note: UpB 90% CI of $P_{BAK}-P_{charité}$: Upper Bound of 90% Confidence Interval of $P_{BAK}-P_{charité}$

Assuming different scenarios in favor of the BAK group and against the Charité group, I performed sensitivity analyses. As shown in Table 5, for all randomized subjects (i.e., the true ITT population), under the worst case scenario (Case 2a), that is to treat all the missing data of the Charité subjects as failures (28/205=14%) but successes for all the BAK missing data (21/99=21%), the upper bound of the two-sided 90% confidence interval of the difference ($P_{BAK}-P_{Charité}$) exceeds the non-inferiority margin of 15%. Therefore, the non-inferiority claim cannot be made for the Charité disc. As we move from Case 2a to 2c by assuming a success rate of 43% (12/28) for the missing Charité subjects, or from Case 2a to 2d by assuming a success rate of 71% (15/21) for the missing BAK subjects, the non-

inferiority claim can still be made. Please note that the scenario under each of these two assumptions is in favor of the BAK group considering the success rates of 65% (115/177) for the Charité completers and 59% (46/78) for the BAK completers (Table 19, Vol. 2, page 39).

4. To evaluate the impact of covariates on the primary success outcome, the sponsor conducted subgroup and covariate analyses (Appendix 1 –Tables 23.1, 23.2). According to the SAS programs (eff_covar.sas and eff_covar_locf.sas) provided by the sponsor on March 22, 2004, the sponsor used PROC CATMOD to fit a separate model of marginal probability of success with each individual covariate, treatment and treatment by the covariate interaction. We suggest to the sponsor to perform covariate-adjusted analysis by fitting a single model which simultaneously includes all the covariates of interest. Please note that there was an imbalance of covariates (e.g., BMI and pre-operative activity level, Tables 2a and 2b) between the two treatment groups and adjustments are necessary. To avoid loss of information, continuous covariates (e.g., age, BMI, baseline Oswestry score, etc.) should be entered into the model as continuous variables as opposed to categorical variables. I would suggest that the sponsor should try GEE (Generalized Estimating Equations) method using PROC GENMOD due to its robustness against missing values and misspecification of correlations among the repeated measurements. Taking the raw, continuous ODI score as a response variable, GEE method can also be used to compare the rates of ODI score improvement between the two groups after controlling the confounding effects of those important covariates.
5. The Intent-to-Treat (ITT) population should consist of all patients who were randomized to the study as stated in the SAP (Stattech Services, LLC). However, the sponsor’s actual “ITT analysis” for the effectiveness endpoints (Tables 19 and 20, Vol. 2, pages 39-41) excluded those patients who were either overdue for the 24-month follow-up (Charité 10, 5% vs. BAK 8, 8%) or had not reached the 24-month visit (Charité 13, 6% vs. BAK 6, 6%). Such exclusion of randomized patients from the analysis will likely lead to strong bias. Therefore, the results from the sponsor’s “ITT analysis” should be interpreted with caution and the results of the sensitivity analyses (see my comment 3) should be taken into consideration. The sponsor should call the analyses of all randomized patients as the ITT analyses.
6. The sponsor performed a repeated measures analysis to evaluate the success rates over the follow-up period from 6 to 24 months (Appendix 1-Table 22, Vol. 3, Pages 70-72). The 90% confidence interval of the difference in success rates between the two groups (“90% CI for Difference”) reported on page 71 of the Vol. 3 (-1.0, 20.1) is not consistent with the p-value (<0.0001) of Blackwelder’s test. My own calculation for the upper bound of the 90% CI for the difference is 0.004. In addition, all the standard errors and the estimated success rates at 12-month for the Charité group did not match those in the electronic version of the table provided by the sponsor on March 22, 2004.
7. Although the sponsor’s summary of safety endpoints appeared to be adequate, there was no single statistical analysis conducted for comparing the safety profiles between the two

groups. The sponsor should perform appropriate statistical analyses to compare the incidence of device-related adverse events between the two groups.

8. For all Blackwelder's tests (e.g., Table 4), the sponsor should also provide the 95% confidence intervals as well as the P-values.
9. Please ask the sponsor to correct the following mistake: In Table 19 "Comparison of Success Rates for Efficacy at 24 Months – ITT" on page 39 of Vol. 2, the numbers/percentages of overall success for Charité and BAK completers should be 114 (65%) and 45 (58%), respectively (see Appendix 1-Table 19, Vol. 3, page 59).

IV. Summary

Overall, I think that the study was well conducted and adequately reported. However, there are three important issues which cast doubt on the sponsor's primary "ITT" analysis result (success rates: Charité 114/182=63% vs. BAK 45/85=53%, Blackwelder's test: $p < 0.0001$) to support the non-inferiority claim:

- a. A higher percentage of patients in the BAK group (7/99=7.1%) than in the Charité group (5/205=2.5%) were discontinued and treated as failures. This treatment of discontinued patients is in favor of the Charité group;
- b. Those patients who were either overdue or not-yet-due (Charité 23/205=11.2% vs. BAK: 14/99=14.1%) were excluded from the sponsor's ITT analysis and such exclusion may introduce patient selection bias;
- c. There were some covariate imbalances between the two groups (e.g., age, BMI and pre-operative activity level, etc. see Tables 2a and 2b) indicating that patients in the BAK group could have been worse to start.

To address the first two, the sponsor performed sensitivity analyses using LOCF (Tables 3 and 4). With regard to the last one, the sponsor performed subgroup analyses by fitting a separate model with each individual covariate, treatment and treatment by the covariate interaction.

In order to assess the validity of LOCF, the sponsor should provide the details of LOCF for handling the missing data (see comment 3). To correctly adjust covariates, the sponsor should fit a single model which simultaneously includes all the covariates of interest (GEE method recommended, see comment 4). The other seven deficiencies as listed in above should also be addressed.