Clinical
Study Report

# Report of United States Clinical Study Results
# BRYAN® Cervical Disc System [1]

## EXECUTIVE SUMMARY

A prospective, randomized, controlled IDE clinical study was conducted to evaluate the safety and effectiveness of the BRYAN® Cervical Disc in support of a PMA application for the device. Patients requiring single-level surgery in the treatment of cervical degenerative disc disease participated in the study. The control group was comprised of patients having an anterior plated fusion procedure with structural allograft, considered the current standard of care.

The primary endpoint of the study was a composite variable termed "overall success", which included key safety and effectiveness considerations. The study had a non-inferiority design based on the primary endpoint. Statistical superiority was also examined where appropriate. A total of 463 patients participated, with 242 receiving the BRYAN (investigational) device and 221 having the control treatment. The results and conclusions in this report were based on an interim analysis as pre-defined in the protocol.

The investigational device was found to be at least as safe as the control treatment. The rate of investigational device patients having at least one adverse event was similar to the control group rate. Investigational patients had lower rates of serious adverse events that were classified as implant- or implant/surgical procedure-associated. Of particular note, the investigational group had statistically lower rates of second surgical procedures related to supplemental fixations. Investigational patients also had similar rates of revisions and removals to control patients. The investigational group neurological success rate was similar to that of the control group.

In terms of primary dataset effectiveness measures, NDI score improvement following surgery was dramatic for both treatment groups, exceeding 28 points at 12 and 24 months. The mean improvements in NDI scores for the investigational group were consistently higher than those of the control group at all postoperative periods. A comparison of the NDI success rates showed that the investigational group was superior to the control treatment. Neck and arm pain results also showed non-inferiority of the investigational treatment to the control.

Radiographically, mean angular motion values for investigational patients increased slightly after surgery, thus indicating that the device maintains motion at the treated level. This is one of the intended functions of the device. In terms of adjacent level motion, the two treatments showed similar performance following surgery. Functional spinal unit (disc) height was maintained postoperatively in a very high percentage of patients in both treatment groups.

Investigational patients returned to work more quickly than control patients. The median time to return to work for investigational patients was 48 days, which was 13 days shorter than the time for control patients. This difference was statistically significant.

At 24 months following surgery, the primary dataset overall success rate for the investigational group was 80.6%, approximately 10 percentage points higher than the 70.7% rate for the control group. A similar difference of nearly 11 percentage points was also seen at 12 months. The overall success rates for the investigational group were found to be not only statistically non-inferior to the control group rates but also superior.

Therefore, the clinical study objective was met, thus indicating that the BRYAN® Cervical Disc System is as safe and effective as the current standard of care, fusion, for treating cervical disc disease.

---

## I. Introduction

In May of 2000, Spinal Dynamics Corporation filed an application for an Investigational Device Exemption [redacted] with the FDA to study the use of the BRYAN® Cervical Disc in patients with symptomatic cervical disc disease. [redacted] Medtronic Sofamor Danek purchased Spinal Dynamics in June of 2002 and began managing the study about one year later.

Clinical study surgeries were performed during a period from May 28, 2002, to October 8, 2004. A total of 463 study surgeries were completed, consisting of 242 investigational patient surgeries and 221 control patient surgeries. The patients are currently being evaluated at the prescribed postoperative time periods.

This report details the results of the clinical study. The conclusions are based on interim analyses which were pre-defined in the protocol.

# II. Clinical Study Description

## A. Clinical Study Goals and Design

The goal of the IDE clinical study of the BRYAN® Cervical Disc was to evaluate the safety and effectiveness of the anterior cervical spinal use of the device in the treatment of patients with symptomatic degenerative disc disease. The assessments of safety and effectiveness of the BRYAN Cervical Disc were through direct clinical data comparisons between data collected from patients implanted with the BRYAN device and an equivalent group of patients who received a surgical fusion utilizing structural allograft bone with plate stabilization (ATLANTIS™ Anterior Cervical Plate System). The investigational and control treatments were randomized in a 1:1 manner.

Per the protocol, patient evaluations were to occur preoperatively (within 2 months of surgery), at surgery, and postoperatively at 6 weeks (±2 weeks), 3 months (±2 weeks), 6 months (±1 month), 12 months (±2 months), and 24 months (±2 months). Patients in the investigational arm will be followed biennially until the last subject enrolled in the study has been seen for his/her 24-month evaluation.

The effectiveness of the BRYAN device was based primarily on a patient having Neck Disability Index (NDI) pain/disability improvement. In addition, neck pain, arm pain, patient gait, general health status, patient satisfaction, and radiographic parameters were evaluated. Safety was based primarily on the nature and frequency of adverse events and second surgeries. The maintenance or improvement in neurological status following surgery was also a safety measurement.

The primary endpoint for the clinical investigation was a composite variable termed "overall success" (at 24 months). The overall success variable was comprised of NDI and neurological results. Success for these factors, as well as the patient not having a serious implant- or implant/surgical procedure-associated adverse event or having a second surgery classified as a "failure", determined whether the patient was an overall success. Investigational treatment success was based on the 24-month overall success rate being statistically non-inferior to the control group rate.

For additional information pertaining to the clinical study design, please refer to the current protocol, case report forms, patient

informed consent template, and Statistical Considerations provided in **Attachment A**.

## B. Investigational Plan Changes

During the clinical trial, a number of IDE supplements were submitted that accommodated the course of the clinical study. For example, supplements were submitted that changed the sample size and/or amended the statistical plan. Perhaps the most relevant of these supplements was one that allowed the enrollment of an additional 20 investigational patients to accommodate randomization of the originally allotted number of control patients. See Section II.C.4 of this report for additional information. This supplement, which was approved by FDA on July 23, 2004, brought the approved study enrollment to a total of 470 patients (245 investigational, 225 control), whereas the initial full approval in 2003 was for 450 patients.

Another supplement dated February 7, 2003, removed exclusion criteria pertaining to Worker's Compensation and lumbar-treated patients. In addition, a supplement dated December 13, 2002, added an inclusion criterion requiring a minimum score of 30 points on the Neck Disability Index questionnaire. On July 25, 2003, Medtronic Sofamor Danek informed FDA that it would be assuming responsibility as the study sponsor due to the purchase and closure of the Spinal Dynamics facility. None of these supplements were believed to have any negative effects on the scientific soundness of the clinical trial, and all were approved by the FDA.

The FDA also granted approval of a supplement to allow continued access to the investigational device while the postoperative follow-up of the original IDE study patients occurred and the PMA application was being prepared and processed.

Currently, 29 patients have had surgeries in the continued access program, and information concerning them can be found in **Section IV.C**, "Other Relevant Clinical Data", in this module. None of these patients have been evaluated at their 24-month postoperative visit.

## C. Patient Population

### 1. Patient Accountability

The accountability of patients in the investigational and control groups at the different clinical study periods is provided in **Tables 1**, **2**, and **3**. These tables also provide patient evaluation distributions as a function of time within each study period. A total of 242 patients received the investigational device, and the control group had a total of 221 patients. The date of database closure for analyses was June 5, 2006.

**Table 1** presents patient accountability on the basis of having received any information on an individual at the prescribed time periods. The table also provides a time course distribution of the information at each study period. The composite follow-up rate for the two treatment groups was approximately 90% at 24 months. The 24-month follow-up rate for the investigational group was 95.2%, compared to a control group rate of 85.4%.

**Table 2** is very similar to Table 1 except it is a more conservative presentation of patients who had overall success outcomes. At 24 months, the composite rate for the two treatment groups and the percentages of patients having overall success outcomes were identical to those in Table 1.

As discussed later in this report (Section IV.G.4), we do not believe that the observed difference in follow-up rates between the two groups had a material impact on the study results or conclusions.

**Table 3** further examines the overall success information as a function of windowing. It also shows when "out of window" patients were observed. At 24 months following surgery, over 91% of the overall success results arose from "in window" visits. Analyses presented later in this report (Section IV.G.2) show that similar study results or conclusions are obtained with "out of window" patients excluded as compared to those with them included.

### 2. Patient Demographics

Demographic information pertaining to the investigational and control treatment groups is presented in **Table 4**.

Statistical comparisons were made to determine whether the two treatment groups had different patient population characteristics. The two treatment groups were very similar demographically, and there were no statistically significant differences ($p<0.05$) for any of the variables.

### 3. Preoperative Evaluations of Clinical Endpoints

**Table 5** summarizes the preoperative status of several effectiveness endpoints for the treatment groups. These analyses focused on NDI, SF-36 component summaries, and neck and arm pain. Except for the Mental Component Summary (MCS) of the SF-36, there were no statistically significant differences ($p<0.05$) between the treatment groups for any of the examined variables.

The MCS finding is of limited importance in this study since it is not part of the primary outcome variable, overall success, nor is the MCS a particularly revealing aspect of cervical treatment success since it is not specific and is influenced by many other factors. Finally, the difference in mean values for the two treatments is approximately two points, which is not considered clinically relevant.

### 4. Consented Patients Who Declined Participation Prior to Surgery

One hundred seventeen (117) patients were randomized but declined participation in the study prior to receiving the assigned treatment. Thirty-seven (37) of these patients would have received the investigational treatment, while 80 were potential control patients. A summary of the reasons is provided in the following table:

| Reasons for Declination Prior to Surgery | | |
|---|---|---|
| | Investigational | Control |
| Insurance Denied | 7 | 1 |
| Condition Improved | 7 | 11 |
| Dissatisfied with Randomization | 0 | 32 |
| Inclusion/Exclusion Criteria Not Met | 3 | 5 |
| Decided not to participate | 6 | 7 |
| Combination* | 0 | 2 |
| Other** | 11 | 18 |
| Unknown | 3 | 4 |
| Total | 37 | 80 |

* Combination of Condition Improved and Dissatisfaction with Randomization

** "Other" includes the following reasons which were cited by fewer than 5 total patients each: needed multi-level surgery (4), various types of scheduling conflicts (3), not medically cleared for surgery (2), did not receive protocol-specified treatment (2), could not wait on insurance verification (2), no time off work (2), and worker's comp would not authorize (2). Reasons reported by only one patient each included discovery of lung lesion, fear of risks, did not stop smoking, financial hardship, legal counsel advised against surgery, wanted autograft, did not have insurance, low back pain more severe than neck pain, went with second opinion, surgery not necessary, surgery not authorized, and had surgery in insurance network.

The aforementioned demographic and baseline status data were collected for these patients, and statistical comparisons were made to compare these patients to those who did receive study treatments. These data are presented in **Table 7** parts a and b (the patients who withdrew prior to receiving a study surgery are noted as non-study patients in these tables). The non-study patients appear similar to those who underwent a study surgery.

There were only three comparisons made between study and non-study patients for which statistical differences were found ($p < 0.05$). For investigational patients, the non-study patient cohort had a higher frequency of tobacco users. For control patients, a higher rate of alcohol usage and a higher mean SF-36 PCS score was noted in non-study patients.

None of the statistical differences described above are believed to impact the ultimate clinical outcomes of the study. The occurrence of differences was infrequent and could be by chance due to the large number of comparisons. More importantly, the demographic characteristics and baseline status of the investigational and control patients who did participate in the study were very similar, as detailed in previous sections of this report.

### D. Investigator Information

Sixty-five (65) investigators and co-investigators from 30 sites[2] performed surgeries in this clinical study. No single investigational site contributed more than 12% of the total study patients. Please refer to **Attachment B** for a listing of the investigators. This attachment lists all investigators and co-investigators who signed the Investigator Agreement, regardless of whether they performed any surgeries. Also, please see **Section IV.B** later in this module for financial disclosure forms for the investigators; this section contains information only on those investigators who performed study surgeries.

## III. Statistical Methodology

### A. Clinical Trial Objectives and Hypotheses

The primary objective of the clinical trial was to determine if the proportion of patients having favorable overall success outcomes at 24 months after surgery (the primary endpoint) in the investigational treatment group was statistically non-inferior to the overall success rate in the control treatment group. Secondary objectives were also developed for the clinical trial. These objectives were used to determine whether the investigational group demonstrated superior overall success results as compared to the control group. These objectives were also used to determine if the success rates for the individual effectiveness and radiographic variables, such as NDI and FSU height, as well as neurological status, were statistically non-inferior for the investigational treatment group, as compared to the control treatment group. In addition, if non-inferiority was established, analyses were proposed to determine if the investigational group had superior outcomes when compared to the control group for those individual variables. As FDA recommended, a fixed value of 0.10 was used as the non-inferiority margin for assessing all of the non-inferiority hypotheses.

For adverse events, additional surgical procedures/interventions, and surgery and hospital information, only superiority hypotheses were proposed, and statistical comparisons were only done for reference purposes because of the large number of categories of individual adverse events and additional surgeries/interventions.

## B. Analysis Datasets

Three different analysis datasets (primary, per-protocol, and missing-equals-failure datasets) were defined. The primary dataset consisted of all the patients who received study devices[3] and completed surgical procedures. In the case where a patient received the other study treatment – that is, a patient was randomized as control but actually received the investigational treatment or vice versa – the patient was grouped according to the actual treatment that the patient received (there were twelve patients in this study who were randomized to the investigational group but received the control treatment and one patient who was

randomized to the control group but received the investigational treatment). Primary statistical comparisons were based on the observed data, and missing data due to lost-to-follow-ups were not imputed. For patients who had additional surgical procedures/interventions that were classified as "failures", they were deemed as failures for overall success, the primary endpoint. For other individual variables, the last observations taken before the additional surgical procedures/interventions were carried forward for all future evaluation periods.

The per-protocol dataset was a subset of patients who were included in the primary analysis dataset. Patients who had major protocol deviations, i.e., those who did not meet the inclusion/exclusion criteria, those who received wrong study treatments (patients who were randomized as control but actually received the investigational treatment or vice versa), or those who had other major protocol deviations that could potentially affect clinical outcomes, were excluded from this dataset. A list of those patients and a brief description of their major protocol deviations are provided in **Attachment C**. Additional surgical procedures/interventions and missing values due to lost-to-follow-ups were handled in the same manner as in the primary dataset. The per-protocol dataset was constructed only for the primary endpoint (overall success) and its component variables. Statistical comparisons using this dataset should be considered as a secondary analysis.

To assess the effects of lost-to-follow-ups and missing observations (including deaths) on study outcomes, a "missing-equals-failure" dataset was constructed for the primary endpoint overall success and its component variables. In this dataset, all missing responses in the patients who received study devices and completed surgical procedures, regardless of the reason, were assumed to be failures. Success rates were computed and presented for each treatment group, but no formal statistical comparisons were performed with this dataset. Results of this type of analyses are largely dependent on the follow-up rates. It would bias against the control if the control group has a relatively lower follow-up rate as in this study. We presented this analysis only because FDA/CDRH has traditionally requested it.

## C. Statistical Methods and Computations

Bayesian statistical methods were used to determine non-inferiority and superiority of the investigational device to the control for success rates in overall success, individual effectiveness and radiographic variables, and neurological status.

patients are past the 12 month evaluations, but not all of them

Bayesian methods were also used for comparing surgery and safety data between the investigational group and the control group.

4

## D. Bayesian Interim Analysis

The Statistical Considerations in the investigational plan pre-defined that the data would first be analyzed after a total of approximately 300 patients (investigational and control combined) had follow-up visits at 24 months. At that time point, all the patients would be expected to have reached the 12-month evaluation period. The data would also be summarized when the entire cohort of patients had reached the 24-month time point. Thus, one interim analysis and one final analysis were planned.

This PMA application is primarily based on the pre-defined interim analysis criteria, using the first 300 patients (ordered by the surgery date) who had valid outcomes in overall success at 24 months – the primary endpoint. The data collected at or before 12-month visits from all of the patients were also included and presented. These data have been monitored in an appropriate manner, cleaned, and verified.

Because of the time required for data monitoring and cleaning, more than 300 patients have had 24-month visits as of the cutoff date, June 5, 2006, and have valid outcomes for overall success at 24 months. We labeled the whole cohort of data as "all currently available 24-month data" and presented them, along with the data collected at or before 12-month visits from all the patients and Bayesian statistical analyses, in Section IV.G.8 of this report. The presentation is intended to show the robustness of the study conclusions.

Because of the cumulative nature of the information on adverse events and additional surgical procedures/interventions, the most updated data from all the patients were analyzed and presented in this report.

# IV. Results

## A. Surgery Information

**Table 7** provides summaries of information related to the surgical procedures and postoperative hospitalizations of patients. The results of the statistical analyses comparing the investigational and control groups are provided in **Attachment D**. The mean operative times for the investigational and the control treatment groups were 2.2 hours and 1.4 hours, respectively. The mean operative time for the investigational group was approximately 48 minutes longer, and this difference was statistically different based on Bayesian analyses. This statistical difference is likely due to the newness of the surgical technique and to overpowering of the large sample sizes.

Investigational patients were found to have a statistically higher blood loss than control group patients (91.5 ml versus 59.6 ml). We believe the blood loss findings can possibly be explained, in part, in a similar manner to the operative times, i.e., the novelty of the surgical procedure and less streamlined instruments. The approximate 30 mL difference in mean values is not considered clinically surprising since blood loss is typically an estimated value. The median and maximum blood losses were similar for the two treatment groups.

The mean hospital stays for investigational and control patients were statistically similar (1.1 days vs. 1.0 days, respectively). Both treatment groups had a median hospital stay of 1 day.

Even though statistical analyses were not performed, it is evident that the treatment level distributions for the two study groups were similar. Over 92% of the patients in both treatment groups had procedures at either C5-C6 or C6-C7. Consistent with the standard of care, orthoses, such as soft or hard collars, were permitted under the protocol. As expected, more patients in the control group received postoperative external orthoses. Approximately 74% of the investigational patients did not use them, as compared to 32% for control patients.

In summary, investigational device patients had statistically longer operative times and higher blood losses as compared to control patients. These findings are believed to be attributed to the newness of the investigational device surgical technique and the use of first generation instruments. The large sample sizes which power this study also play a major role in these statistical differences. Despite the statistical findings, the actual mean differences are considered to be of marginal clinical relevance. These findings should be considered when interpreting the clinical outcomes. Should the investigational group be found to be statistically non-inferior to the control group, one can surmise that either the operative findings were non-influential or that the investigational treatment overcame them. The latter consideration would be a potential indicator of still better clinical outcomes in the future as surgeons become more familiar with the product [redacted] The hospital stays were statistically similar for the two treatments. Also, most of the procedures involved the lower cervical spine.

## B. Safety Measurements

### 1. Adverse Events

The safety of the investigational device was evaluated based on the nature and frequency of adverse events, as compared to those occurring in the control group. Adverse events vary in severity. Some may resolve without any subsequent treatment, some may require nonoperative medical intervention, and others may result in another surgical procedure.

Adverse events were categorized by their nature. There are 20 categories of adverse events, such as neurological, infection, dysphonia/dysphagia, etc. If the underlying cause of the adverse event was known, it was categorized accordingly. If the underlying cause was unknown, the adverse event was categorized according to the symptoms. For example, if a patient had neck and/or arm pain secondary to a fall, the event was categorized as “Trauma”. On the other hand, if the cause of the neck and/or arm pain was not known, the event was categorized as “Neck and/or Arm Pain”.

Adverse events were classified according to their severity utilizing World Health Organization (WHO) criteria. If the adverse event was graded as a “3” or “4”, it was considered “serious”; otherwise, it was considered “non-serious”.

Typically, adverse events that resulted in an emergency room visit or a hospitalization were regarded as "serious". Adverse events were also classified according to their potential relationship to the implant and/or surgical procedure. Many of the adverse events may have no connection to the study treatment.

Detailed narratives of the reported adverse events and the classification and grading of them are provided in **Attachment E** for both the investigational and control patients. Please note that patients who had an event that caused them to be considered a second surgery failure are identified with specific language in the adverse event narrative.

**Table 8** provides a time course summary of operative and postoperative adverse events reported for investigational and control patients as a function of adverse event category. The total number of occurrences per category and the number of patients involved are also provided.[5] Statistical comparisons of the occurrence rates were made using Bayesian methods (**Attachment F**). The rates were based on dividing the number of patients having at least one occurrence of a particular adverse event type by the total number of patients in that treatment group.

A total of 202 (83.5%) investigational patients had at least one adverse event. Similarly, the number of patients in the control group with any adverse event was 174 (78.7%). These rates were not statistically different.

The investigational device group had statistically lower adverse event rates as compared to the control group for non-union (0.0% vs. 2.3%) and pending non-union (0.0% vs. 2.3%) categories. Such events were not possible for investigational patients. These were the only two categories in which a statistical difference was noted.

As shown in **Table 9**, the number of patients having serious adverse events, i.e., those with a WHO grade of 3 or 4, in the investigational group was very similar to that that found

[5] The discussions and analyses regarding adverse events are based on those events occurring from surgery to 24 months postoperative. However, the tables do provide information regarding adverse events, if any, occurring after 24 months. Information regarding adverse events that occurred preoperatively (after the consent was signed), as well as further details about events after 24 months, is provided along with the comprehensive narratives of adverse events in **Attachment E**.

in the control group (26.4% vs. 24.9%). The treatment group rates for the various categories were fairly similar. There were only six categories in which the rates differed by one percentage point or more. For these, the investigational group had lower rates of grade 3 or 4 adverse events classified as neck and/or arm pain, non-unions, and respiratory. Similarly, the control group had lower incidences of adverse events classified as other, trauma, and urogenital. For the latter finding, there were five investigational patients who had grade 3 urogenital adverse events as compared to one in the control group. In the investigational patients, these five events were due to hematuria, pelvic pain from a benign dermoid tumor, ureteropelvic junction stone, bladder stones, and dysmenorrhea. The one urogenital event in a control patient was due to kidney stones. None of these were considered related to the treatments.

**Table 10** summarizes the adverse events that are classified as implant-associated or implant/surgical procedure-associated.[6] The number of patients with these types of adverse events was lower in the investigational group than in the control group (2.9% vs. 5.4%). Upon closer examination, these event rates were fairly similar for both treatment groups in most of the specific categories. The main exceptions were the non-union and pending non-union categories, where the control group rates were 2.3% and 2.3%, respectively, as compared to 0.0% rates for investigational patients.

**Table 11** summarizes those adverse events that were both serious and classified as implant- or implant/surgical procedure-related. Four such events occurred in investigational patients (1.7%). One of these was a malpositioned implant at the time of surgery, two were neck and/or arm pain events (one at 6 weeks and one at 3 months postoperative), and one was a trauma (work injury) at 6 months postoperative.

Seven events classified as serious and implant- or implant/surgical procedure-related occurred in control patients (3.2%). One of these was neck and/or arm pain at 6 weeks, and one was a spinal event at 24 months. The

[6] Typically, if a second surgery occurred at the involved level, the adverse event was classified as implant/surgical procedure-associated. Some other, similar adverse events may be classified as undetermined because they did not result in a second surgery.

other five events, which were all due to non-unions, occurred at 3 months, 6 months (2 events), 12 months, and 24 months postoperative.

There were no unanticipated adverse device effects (UADE) reported in this study.

In terms of specific adverse event categories, there were six categories in which the investigational and/or control group adverse rates, regardless of severity and causality, were greater than or equal to 10%. The following is a discussion of these particular findings. In addition, discussions are provided for two other categories of adverse events (cancer and deaths) due to the nature of the events.

**Neurological**

A total of 60 neurological events occurred in 48 patients in the investigational group (19.8%). The most commonly reported event among investigational patients was numbness (26 events). Of these 26 events, 15 involved the upper extremities (arms, hands and fingers). There were two reports of numbness in the C6 distribution. There was one report each of general numbness, numbness in the C7 distribution, numbness in the C8 distribution, and numbness in the face and extremities. In addition, there were five events that occurred in the lower extremities, including numbness in the feet, numbness that affected the thigh, and nonspecific lower extremity numbness.

The next most frequently reported neurological events in investigational patients involved paresthesia, tingling, numbness and tingling, numbness and pain, neuropathy, and radiculopathy. There were two events of paresthesia affecting the arm and one event of nonspecific paresthesia. There was one report each of tingling affecting the hands, the feet, the fingers, left upper extremity, and the left shoulder and fingertips. There were two events of numbness and tingling affecting the fingers, and one report each of numbness and tingling affecting the hands and numbness and tingling affecting the toes and knees. There were four events of numbness and pain affecting the neck, and one event of numbness in the neck accompanied by intrascapular pain.

There were two events of neuropathy affecting the upper extremities and two events of neuropathy affecting the lower

extremities. There were four events of nonspecific radiculopathy.

There were two events of weakness reported. One event of weakness affected the left upper extremities (one event), and the other affected the left side (one event). There were two reports of dysesthesia. One event of dysesthesias affected the fourth and fifth digits of the left hand, and the other event was a nonspecific report of dysesthesias. Additionally, there were two instances of tremors (one event was nonspecific, and the other event affected the right hand).

There were three events that only occurred once in the investigational group. These events included: tingling in the finger accompanied by arm pain, sensory changes, and a burning sensation in the hand.

A total of 50[7] neurological events occurred in 46 patients in the control group (20.8%). The most frequently reported event was numbness (20 events). Of these 20 events, there were 18 events involving the upper extremities (arms, shoulder, elbow, hands, and fingers). There was one instance involving numbness of the leg and one report of general numbness.

The next most frequently reported neurological event involved numbness accompanied by tingling, pain, and/or burning (11 events). There were five events involving numbness associated with tingling. These events involved the arm (three events) and/or finger (one event) with one report of nonspecific numbness and tingling. There were four events involving numbness associated with pain. These four events included two reports of numbness in the thumb accompanied by neck pain, one report of hand numbness accompanied by arm pain, and one report of numbness accompanied by arm pain and tingling. There were two events of numbness and burning. One of these events involved upper extremity numbness concomitant with burning in the spine. The remaining event involved numbness in the arms with burning in the neck.

[7]

The next most commonly reported events involved radiculopathy (two events), myelopathy (two events), dysteshesia (two events), tingling (two events), and weakness (three events). The two events involving radiculopathy were in the C5-6 distribution and C6 distribution, respectively. There were two reports of myelopathy, both of which were present preoperatively, but persisted despite treatment. There were two reports of dysesthesia. One event was in the right arm. The remaining dysesthesia event was in the arm and was accompanied by parascapular pain. There were two reports of tingling. One of these events was tingling of the scalp, and the other was tingling of the left index finger and thumb. There were three events involving weakness. These three events involved the triceps, fingers, and left arm/hand, respectively.

Finally, there were eight events that occurred only once. These events included the following: abnormal reflexes, ulnar neuropathy, cold sensation in feet, tremors in the left hand, ulnar entrapment, "rubbery legs", "hollowness in arms", and spinal cord compression.

As can be observed from the above descriptions, the nature and frequency of neurological adverse events were very similar for the two treatment groups. Bayesian statistical analyses showed no difference in the rates for the two groups.

### Other Pain

A total of 56 events classified as "other pain" occurred in 49 patients in the investigational group (20.2%). The most frequently occurring categories were back pain (21 events) and headaches (13 events). In addition, there were five reports of back and leg pain, four reports of non-cardiac chest pain, two reports of foot pain, and two reports of thoracic pain. Each of the following events occurred once: edema, plantar fasciitis, hallux valgus, and hand/knee swelling. Additionally, each of the following types of pain was reported once: SI joint, trapezius, axial joint, leg, and hip.

A total of 47 events classified as "other pain" occurred in 44 patients in the control group (19.9%). Again, the most frequently reported events were back pain (16 events) and

headaches (10 events). In addition, there were three reports of leg pain, two reports of knee pain, two reports of non-cardiac chest pain, one report of headaches accompanied by neck and shoulder pain, and one report of chronic arthritis in the cervical region. Each of the following types of pain was reported once: foot, trapezius, hip, diffuse spine, back/leg, trapezius/back, thoracic/back, and hip/feet/hand. Finally, there was one occurrence each of degenerative hip disease, paraparesis, osteoarthritis, and inflammatory arthritis.

The rates of adverse events classified as "other pain" were not different for the two treatment groups.

### Trauma

A total of 42 trauma events occurred in 34 patients in the investigational group (14.0%). The most frequently reported events were motor vehicle accidents (15 events) and falls (14 events). Also reported were five work-related injuries, two sporting injuries, and two lacerations. Additionally, there were four events that occurred only once and were reported as trauma. These included the following: finger slammed in the door, broken foot, broken wrist, and sprained wrist.

In the control group, there were 27 trauma events noted in 22 patients (10.0%). Similarly, the most frequently reported were falls (11 events) and motor vehicle accidents (six events). Also reported were two head injuries and two ankle injuries. Additionally, there were six events that occurred only once and were reported as trauma. These included the following events: work injury, bumped elbow, object falling on patient, finger laceration, running into a door, and swallowing a hot object.

The rates of adverse events categorized as trauma were not statistically different for the two treatment groups.

### Neck and/or Arm Pain

A total of 140 events occurred in 115 patients in the investigational device group (47.5%). The events included the following: 34 neck pain; 18 shoulder pain; 16 arm pain; seven neck and arm pain; seven neck and shoulder pain; seven neck spasms; four arm and shoulder pain; four rotator cuff events; four trapezius pain; three neck and scapular pain; three neck and trapezius pain; three scapular pain; two

neck and scapular pain with neck spasms; two neck and shoulder pain with spasms; two neck pain with headache; two neck pain with spasms; two rotator cuff tendonitis; two tightness at incision site; two trapezius spasm; and two shoulder, scapular, and arm pain.

Additionally, there were 14 events that occurred only once in the investigational group and were reported as neck and/or arm pain. These included the following: arm pain with muscle spasm; arthritis in neck; clavicle pain; neck pain and muscle spasm in upper back with burning sensation; neck locks up in cold weather; neck pain and arm numbness; neck swelling in left side; neck, arm, and shoulder pain with hand numbness; shoulder and elbow pain; shoulder and trapezius pain; shoulder bursitis; tendonitis of the elbow; wrist and thumb pain; and wrist pain.

By comparison, a total of 128 events occurred in 96 patients[8] (43.4%) in the control group. The events included the following: 36 neck pain; 12 neck and arm pain; 11 shoulder pain; eight arm pain; eight trapezius pain; five scapular pain; five neck and shoulder pain; five neck spasms; four rotator cuff events; three neck pain with headache; three neck pain with muscle spasms; three neck, shoulder, and arm pain; two arm and shoulder pain; two elbow pain; and two neck, trapezius, and arm pain.

Additionally, there were 19 events that occurred only once in the control group and were reported as neck and/or arm pain. These were as follows: arm and trapezius pain; arm pain with tingling and heaviness; deltoid pain; muscle spasm on back of shoulder; neck discomfort; muscle spasm with neck, shoulder, and elbow pain; neck and arm pain with burning in thumbs; neck and scapular spasms; neck pain with spasm and arm pain; neck, shoulder and trapezius pain; neck, shoulder, arm, and head pain; shoulder capsular tear; shoulder pain with muscle spasms; shoulder tendonitis; tightness in throat; trapezius spasms; tricep cramping; wrist and elbow pain; and wrist injury.

The reporting of neck and/or arm pain as an adverse event in a study such as this is questionable since these events

8

are measured effectiveness endpoints. However, these were recorded in the spirit of providing a complete picture of the study treatments. Cumulatively, over 45% of the patients in both treatment groups had at least one neck and/or arm pain complaint that was reported as an adverse event. One should keep in mind that those events were reported over the course of 24 months.

The rates of neck and/or arm pain adverse events for the two treatment groups were not statistically different.

### Dysphagia/Dysphonia

A total of 28 dysphagia/dysphonia events occurred in 26 patients in the investigational group (10.7%). There were 15 reports of dysphagia and 13 reports of dysphonia. A total of 20 dysphagia/dysphonia events occurred in 19 patients in the control group (8.6%). There were 16 reports of dysphagia and four reports of dysphonia. The rates of dysphagia/dysphonia adverse events for the two treatment groups were not statistically different.

### Other Adverse Events

Some adverse events occurred infrequently and did not fit a particular relevant category. These adverse events were combined into an "Other" category. A total of 84 events occurred in 59 investigational patients (24.4%), and 47 events occurred in 39 control patients (17.6%). These rates were not statistically different for the two treatment groups. The nature of these events for both treatment groups is presented in **Attachment G**.

### Cancer

The incidence rate of cancer in the investigational group was 0.8% (two patients). Neither cancer was deemed to be related to the study treatment.

The first investigational patient was diagnosed approximately seven months following the study surgery with a metastatic, neuroendocrine carcinoma with a primary carcinoid of the duodenum. The patient had a positive history of kidney stones and initially presented to the emergency room with complaints of flank pain. During the

work-up for the flank pain, an abdominal mass was discovered, which was eventually diagnosed as cancer. The patient has since undergone two separate hospitalizations for exploratory surgery and treatment. At the 24-month study evaluation, the patient was reported to be recovering and is under the routine care of his family physician. It is noted that the patient has a positive family history of cancer.

The second investigational patient was diagnosed with a 2 mm papillary carcinoma of the thyroid approximately 24 months following the study surgery. Prior to enrollment in the study, the patient was known to have a cystic mass on the thyroid. Four weeks following the study surgery, the patient complained of difficulty swallowing, and the patient was eventually referred for possible surgery to remove the cystic mass. The entire thyroid was removed after the discovery of the carcinoma. A full body scan was performed four months following the removal of the thyroid, and no evidence of metastatic thyroid carcinoma was found.

There were no incidents of cancer in the control group.

### Deaths

There were no deaths in the investigational device group.

In the control group, there was one death (0.5%). Approximately 17.5 months postoperatively, the patient died as a result of injuries sustained in a motor vehicle crash. The death was not reported to the clinical site until approximately four months after the event. No further details of the injuries sustained were made available.

## 2. Radiologist Findings

The radiographs from the clinical study were evaluated by independent reviewers. These individual reviewers were asked to report any observations related to bent, fractured, separated, or migrated investigational implants, as well as bent or fractured control implants. A summary of their findings is presented in **Table 12.**[9]

[9] Please note that this discussion includes those for whom such an observation was noted by **any** of the (up to three) individual reviewers. All of this information is captured in the data listings; however, **Table 12** lists the observations of Reviewer 1, Reviewer 2, and the adjudicated value.

In the investigational group[10], one patient ( ) was noted by a single reviewer to have a bent, fractured shell, as well as separated implant components at the surgery/discharge timepoint. Another patient ( ) was noted by a single reviewer to have a migrated implant at the 3-month evaluation. Neither patient nor had an additional surgical procedure, and both were overall successes at 24 months. Finally, a single reviewer noted a third patient to have separated implant components at the 3- and 6-month evaluations. Patient did have a second surgical procedure classified as "other", which was an adjacent-level fusion at the 12-month time period. However, there was no second surgical procedure at the treated level in this patient, and this patient was an overall success at 24 months.

In the control group, bending and fracture of plates and/or screws was noted in two patients at the surgery/discharge timepoint. None of these observed events resulted in an apparent second surgical procedure. Also in the control group, three patients (nos. and were reported to have both bent plates and/or screws as well as bent and/or fractured grafts. This was noted in patient at 3 and 6 months; in patient at 6, 12, and 24 months; and in patient at 3, 6, and 12 months. Finally, bent and/or fractured grafts were reported in the following four patients: at 12 months, at 6 months, at 3 months, and at 12 and 24 months.

Several of the control patients for whom these observations were noted also had second surgical procedures. Patient who is considered a second surgery failure, underwent a posterior fusion procedure classified as a supplemental fixation at 4 months postoperative. Patient who is also considered a second surgery failure, underwent a posterior fusion procedure classified as a supplemental fixation at approximately seven months postoperative, and had a facet neurotomy classified as "other" at fourteen months postoperative. Patient QA04 underwent four additional surgical procedures, two of which were unrelated to the spine (heart catheterization and percutaneous intervention, and scrotal exploration and

[10] It should be noted that the radiographic observations made for patients concerning the BRYAN® implants were made by one reviewer only. The two other reviewers did not make similar observations. The sponsor received the applicable radiographs and could not discern the findings reported by the individual reviewer. Copies of these radiographs can be made available to the Agency upon request.

removal of mass). At approximately 31 months postoperative, this patient also had a posterior fusion classified as a supplemental fixation at the treated level and, at the same time, had a posterior hemilaminectomy and foraminotomy with medial facetectomy two levels above the treated level. Finally, patient [redacted] underwent two additional surgical procedures, but neither of these was in the spine (ankle reconstruction and sinus surgery for chronic sinusitis that existed prior to study enrollment).

### 3. Secondary Surgical Procedures

Some of the adverse events led to surgical interventions subsequent to the clinical study surgery. These additional surgical interventions were classified as revisions, removals, supplemental fixations, reoperations, or "other". A revision is a procedure that adjusts or in any way modifies the original implant configuration. A removal is a procedure that removes one or more components of the original implant configuration without replacement with the same type of device. Removals are further classified into elective and non-elective subgroups. Elective removals encompass those due to patient and surgeon preference, whereas non-elective removals arise from a real medical need. Supplemental fixation is a procedure in which additional spinal devices not approved as part of the protocol are placed. This may include the use of bone growth stimulators (either internal or external). A reoperation is any surgical procedure at the treated spinal level that does not remove, modify, or add any components. Some surgical procedures do not fit into the previously mentioned categories and may not even involve the cervical spine; these are classified as "other". **Table 13** summarizes the secondary surgical interventions in the investigational and control groups, and **Attachment H** provides the case histories for second surgery patients in both treatment groups. The statistical analyses of the rates of secondary surgical procedures between the investigational and the control groups are provided in **Attachment I**.

There was one reported revision procedure (0.4%) in the investigational group and none in the control group. The revision of the investigational patient was due to a malpositioned implant after wound closure at surgery, and this patient was revised by repositioning the implant. There was no statistical difference in revision rates.

There were no supplemental fixations performed on investigational patients, as compared to seven procedures on six (2.7%) control patients. These rates were statistically different. The supplemental fixation procedures were all related to suspected non-unions in the control patients. Non-unions were not a consideration for investigational patients, because they did not receive a fusion procedure. Two of the seven reported supplemental fixations were attributed to the use of bone growth stimulators.

Implant removals occurred in both treatment groups. The investigational group removal rate was statistically similar to that of the control group (1.2% vs. 0.9%). There were three removals in the investigational group. Two of them were due to residual pain, and the third was secondary to trauma. Fusion procedures followed these removals. Both of the control implant removals were non-elective and followed non-unions.

For the three investigational removal procedures, histological analyses were performed in two of the cases, and the final reports are provided in **Attachment J**.[11] No information is available for the third patient since the explanted device was not available for analysis. The histological analyses found tissue responses consistent with those typically seen in proximity to metallic and polymeric implants, with some reports of metallic and polymeric particles. Photomicrographs documenting the histological analyses are available on request. Overall, the condition of the devices was good. One explanted device was observed to have some abrasive wear on the anterior aspect of the shells, but this was attributed to the device having been implanted in an anterior closed-shell alignment rather than with the shells parallel.

Investigational patients experienced higher rates of reoperations and surgical procedures classified as "other" (0.8% vs. 0.4% and 17.8% vs. 15.4%, respectively) when compared to the control patients. In neither comparison was the difference in rates found to be statistically different.

---

1

If a study patient had a revision, removal, or supplemental fixation procedure, he/she was then classified as a second surgery "failure". These events are considered in the calculations of "overall success" rate for the study. Cumulatively, the investigational group had five second surgery "failures", as compared to six for the control group. Two of the "failures" in the control group occurred after 24 months postoperative.

### 4. Neurological Status

The neurological status of the patients participating in the clinical trial was assessed preoperatively and postoperatively at every follow-up visit. The neurological status assessment tool examined motor function, sensory, and reflexes. An algorithm was developed to transform the scores for each parameter into an overall classification representing a maintenance or improvement in neurological status at a given postoperative time as compared to their preoperative neurological status. The values were totaled for each neurological subsection, i.e., motor, sensory, and reflexes, and then expressed as a percent of the maximum possible score for that subsection.

After determining the percentage scores, the postoperative subsection scores were then compared to the preoperative scores and a successful outcome was declared if the postoperative score was greater than or equal to the preoperative score, i.e. maintenance or improvement in condition. Overall neurological success was based on demonstrating maintenance or improvement, i.e., success, in all three neurological parameters.

**Table 14** shows the distributions of patients in the two treatment groups having a maintenance or improvement in neurological condition following surgery for the three indicators.

The overall neurological maintenance or improvement rates at all postoperative time periods were high, exceeding 90% for both treatment groups.

**Attachment K** contains the Bayesian analyses comparing the overall neurological investigational success rate to that of the control group. These analyses yielded a posterior probability of non-inferiority of essentially 100%, thus

indicating the neurological outcomes for the investigational group were non-inferior to those of the control group.

### 5. Summary

In summary, the investigational device was found to be at least as safe as the control treatment. The rate of investigational patients having at least one adverse event was very similar to the control group rate. This was also true for grade 3 or 4 adverse events. The rates of adverse events that were classified as implant- or implant/surgical procedure-associated, both serious and non-serious, were lower for investigational patients. Investigational patients had similar rates of revisions and removals to control patients. Of particular note, the investigational group had statistically lower rates of second surgical procedures related to supplemental fixations. These findings resulted in a lower second surgery failure rate for investigational patients. The investigational group's neurological success rate was statistically non-inferior to that of the control group.

## C. Effectiveness Measurements

The effectiveness variables represent those measurements that describe the clinical outcomes of the study patients. These include indicators of pain relief, general health status, and doctor and patient perceptions of outcomes.

The results of statistical analyses of the effectiveness measurements between the investigational treatment group and the control treatment group are provided in **Attachment K**.

### 1. Pain/Disability

The Neck Disability Index (NDI)[12] was used to measure the effects of neck pain on a patient's ability to manage everyday life (i.e., a combined measure of pain and disability). The NDI questionnaire is based on a patient's response to ten questions, which focus on pain intensity, personal care, lifting, reading, headaches, concentration, work, driving, sleeping, and recreation. The responses to each question range from zero to five. A lower numeric score represents a better pain and disability status regarding that variable. A total NDI score can be determined by adding the scores of the individual questions and dividing

[12] Vernon H, Mior S. The Neck Disability Index: A study of reliability and validity. *J Manipulative Physiol Ther*. 1991; 14(7): 409-415.

that total by the maximum possible total score (50 if all questions are answered). This yields a percentage. Therefore, NDI scores are in a range of 0% to 100%, with a lower percentage indicating less pain and disability. The NDI questionnaire was administered preoperatively as well as at each postoperative visit.

The mean NDI scores for the investigational and control patients for the different clinical study periods are provided in **Table 15** and in the graph below. At all postoperative time periods for both treatment groups, the mean overall NDI scores improved when compared to the preoperative scores, and these improvements were highly statistically significant ($p<0.001$). The mean improvements in NDI scores for the investigational group at 12 and 24 months postoperative were 36.0 and 32.1, respectively. These values are greater than the mean improvement scores of 31.4 and 28.7 for the control group.

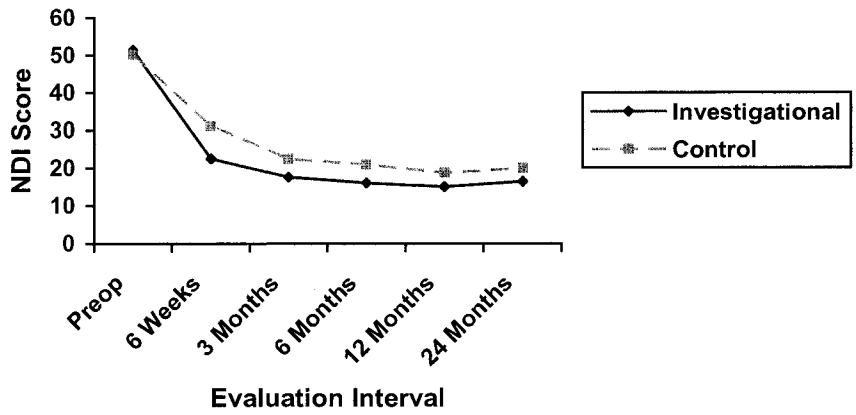


**Table 16** shows the distributions of patients demonstrating successful NDI outcomes. The NDI success criterion is a function of the preoperative NDI score. A 15-point or greater NDI score improvement following surgery was required to be deemed a successful outcome. The table indicates that the investigational group rates at all postoperative periods were greater than the corresponding control group rates. At 12 months following surgery, the investigational group rate exceeded the control group rate by nearly 11 points. Similarly at 24 months, the investigational rate led that of the control by over eight points.

Bayesian statistical analyses showed that the posterior probability of non-inferiority of the investigational group to

the control is essentially 100%. The posterior probability of superiority was 98%. These results indicate that the NDI success rate for the investigational group was not only non-inferior, but also statistically superior, to the rate for the control group.

## 2. Neck Pain

Numerical rating scales[13] were used to specifically evaluate neck pain intensity and duration. The scales for each parameter ranged from 0 to 10, with a lower score representing a better condition. A composite neck pain score was derived by summing the numeral rating scores from the intensity and duration scales, dividing by 20, and multiplying by 100%. Thus, the composite score could range from 0% to 100%.

A summary of neck pain scores is provided in **Table 17**. Like the NDI findings, the mean neck pain scores at all postoperative time periods were less than the preoperative mean values for both treatment groups, thus indicating significant status improvements following surgery. In addition, the mean improvement scores for the investigational group were found to be greater than those of the control group at all postoperative periods. In fact, at 24 months postoperative, investigational patients enjoyed a five-point edge over control patients (51.5 vs. 46.3).

Neck pain success was determined by comparing the postoperative composite neck pain score to the preoperative score on a patient basis. Success was based on the patient having no worsening in neck pain score following surgery. The distributions of patients with successful outcomes are provided in **Table 18**. At 12 and 24 months postoperative, the investigational group had higher neck pain success rates. The investigational group rates were 96.1% and 95.6%, respectively. The control group rates were 95.4% and 92.9%, respectively.

Bayesian statistical analyses showed that the posterior probability of non-inferiority of the investigational device to the control at 24 months is essentially 100%, i.e., statistically non-inferior.

[13] McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires.* New York: Oxford University Press, 1996.

### 3. Arm Pain

Arm pain was assessed in a similar manner to neck pain using numerical rating scales for pain intensity and duration. A summary of arm pain scores is provided in **Table 19**. The mean arm pain scores for each treatment group were similar, and there were statistically significant improvements in condition following surgery. At 24 months, the mean improvement in arm pain score was approximately 50 points in both treatment groups.

Arm pain success was determined in a similar manner to neck pain success. Success was based on the patient having no worsening in arm pain score following surgery. The distributions of patients with successful outcomes are provided in **Table 20**. At 12 months postoperative, the arm pain success rate for the investigational device group was 96.6%, as compared to a 92.4% rate for the control group. At 24 months, the rates were 94.3% and 89.3%, respectively.

Bayesian statistical analyses showed that the posterior probability of non-inferiority of the investigational group to the control was essentially 100%, thus demonstrating non-inferiority. Statistical superiority for the investigational group was approached, with a posterior probability value of 93.8%.

### 4. Quality of Life

The Medical Outcomes Study 36-Item Short Form Health Survey (SF-36) was used to assess general health status of all study patients. The SF-36 is a self-administered test completed by the patient prior to surgery and at all postoperative visits. The SF-36 scale measures specific health concepts related to physical functioning and limitations, social functioning, and health perceptions. The questionnaire contains 36 questions that pertain to eight subscales of health status. These eight subscales are physical function, role-physical, pain index, general health perception, vitality, social function, role-emotional, and mental health. These eight SF-36 scales can be summarized into two measures pertaining to physical health and mental health. The physical component summary (PCS) is based primarily on the physical functioning, role-physical, bodily pain, and general health scales of the SF-36 survey. The mental component summary (MCS) is comprised primarily of the vitality, social functioning, role-

emotional, and mental health scales. **Table 21** presents the mean scores of the eight SF-36 scales, as well as the PCS and MCS, for the different study periods. Higher scores represent higher levels of health.

In terms of the mean PCS and MCS results, all mean postoperative scores were higher than preoperative scores for both treatment groups. The mean improvement in PCS scores from preoperative to 12 and 24 months following surgery for the investigational group (15.7 and 14.4) compared well to those values for the control group (13.9 and 14.5, respectively). The mean improvements in MCS scores from preoperative to 12 and 24 months postoperative for the investigational patients (9.9 and 8.1) were also comparable, if not higher, to those values for the control group (6.9 and 7.3).

**Table 22** presents the proportions of patients who demonstrated maintenance or improvement in SF-36 results postoperatively as compared to the preoperative condition. At 12 months following surgery, the PCS and MCS success rates for the investigational group were higher than those of the control (93.1% vs. 88.2% and 77.3% vs. 72.8%, respectively). However, at 24 months postoperative, the findings changed. For both the PCS and MCS, the control group success rates bettered the investigational rates (90.6% vs. 85.5% and 72.5% vs. 69.8%, respectively). These 24-month findings are intriguing since the mean improvement scores for both variables indicated better outcomes for investigational patients. There are no obvious reasons for these reversals. The Bayesian statistical analyses did not establish non-inferiority of the investigational group at 24 months for the PCS and MCS results. For PCS, the minimum delta for non-inferiority was 10.6%, which is very close to the utilized non-inferiority criterion of 10.0%. The minimum delta for MCS non-inferiority was 13.5%.

These SF-36 findings are of less clinical relevance considering the success rates are not consistent with the mean improvement scores, which are arguably a better measure of outcomes.

Later in this report, a presentation of all currently available data will be made. When these analyses are performed using the larger dataset, investigational group non-inferiority was demonstrated for both PCS and MCS.

SF-36 scores are general health measures and are not specific to cervical surgery. In this study, neither PCS nor MCS are components of the primary outcome variable, overall success. Therefore, these findings are believed to be of limited importance in determining the effectiveness of the investigational device.

### 5. Patient Global Assessment

At each postoperative time period, patients were asked to evaluate their overall impression of their study treatment effectiveness as a function of pain. The seven possible answers ranged from "completely recovered" to "vastly worsened". The results to this question are provided in **Table 23**. At 12 and 24 months following surgery, respectively, 91.0% and 92.4% of the investigational patients indicated that they had either "completely recovered" or were "much improved". These rates were noticeably higher than the 81.7% and 86.4% rates for the control group at 12 and 24 months, respectively.

### 6. Investigator Global Assessment

At each postoperative visit, the doctors were asked to provide their perceptions of the patients' conditions. The responses could be "excellent", "good", "fair", or "poor". The results to this question are provided in **Table 24**. At 12 months following surgery, 93.6% of the doctors responded that investigational patients were in "excellent" or "good" condition. This rate is higher than the 89.8% value for the control group. At 24 months postoperative, 93.8% of the investigational device and 89.3% of the control responses were either "excellent" or "good". These findings show that a substantial majority of patients in both treatment groups were progressing well clinically in the overall opinions of the doctors.

## D. Radiographic Measurements

For this clinical study, the radiographs were evaluated by trained reviewers at SYNARC, Inc., San Francisco, California, under the direction of Harry K. Genant, M.D. Please refer to **Attachment L** for information pertaining to the reviewers and their training.[14]

[14] After the early radiographic data for approximately 30 patients were received, inter- and intra-reader variation was noted. Therefore, additional training and validation activities were conducted, and the radiographs for these patients were re-read.

Financial disclosure forms for these individuals are provided in **Section IV.B** along with those for the clinical investigators.

### 1. Functional Spinal Unit (FSU) Height/Implant Subsidence

Measurements pertaining to the functional spinal unit (FSU) height and implant subsidence were made to evaluate whether the disc space had been maintained during the postoperative course.

The FSU height was determined from lateral neutral radiographs of the treated spinal area and was expressed in millimeters. The anterior FSU height was obtained by measuring from the anterior-most point of the endplate on the superior ventral cortical margin of the cephalic vertebral body to the anterior-most point on the inferior ventral cortical margin of the caudal vertebral body of the treated segment. The posterior FSU height was determined similarly from the posterior aspect. By comparing the magnification-corrected measurements over time, one could determine if the FSU height had changed. FSU height was considered to be maintained or improved, i.e., success, if either the anterior or posterior postoperative measurement was no more than 2 mm less than the 3-month postoperative measurement.

Subsidence was assessed by measuring the distance, in millimeters, through the vertebral midline from the apex of the superior metallic shell to the outermost margin of the superior cortical endplate of the vertebra cephalad to the implant disc space. The same measurement was then repeated from the inferior metallic shell to the inferior cortical endplate of the vertebra caudad to the target disc space. A successful outcome was defined as no more than a 2-mm decrease from the 3-month measurement. Overall subsidence success required successful outcomes for both the superior and inferior onservations.

Further, FSU and subsidence information was combined at each timepoint beginning at the 6-month evaluation. FSU/subsidence success was based on a patient not having a surgical intervention related to a failure finding for either FSU or subsidence.

Measurements were performed by two radiographic reviewers. If their determinations of success differed, a third reviewer was used to break the tie.

The rates of FSU and subsidence success at 6, 12 and 24 months following surgery are presented in **Table 25**. The FSU, subsidence, and FSU/subsidence success rates were high, exceeding 99%, for the investigational treatment group at the three postoperative periods. These success rates for the control group were likewise high, but at 6 and 12 months, there were a few failure reports in the FSU and subsidence categories. At 6, 12, and 24 months, both treatment groups experienced 100% success rates for FSU/subsidence.

Bayesian analyses comparing the investigational FSU success rate to that for the control group demonstrated a posterior probability of non-inferiority value of essentially 100%, thereby demonstrating statistical non-inferiority (**Attachment K**).

### 7. AP Implant Migration

Anteroposterior position of the prosthesis and cervical plate was measured at each of the postoperative radiographic timepoints, and success was defined as no anterior or posterior migration of the device >3.5 mm. The distributions of patients with successful outcomes are provided in **Table 26**. At 12 and 24 months postoperative, the AP implant migration position success rates for the investigational device group were 99.4% and 100.0%, respectively, which were higher than the 93.8% and 86.8% rates seen in the control group.

### 8. Treated Level Measurements

#### a. Investigational Group

Angular motion in the sagittal plane was measured at each study period by comparing lateral flexion and extension radiographs, and the results are given in

**Table 27**. Two independent radiographic reviewers made the measurements. The mean angular motion value prior to surgery was 6.4°. This level of motion was maintained following the implantation of the investigational device. At both 12 and 24 months postoperative, the mean angular motion values were 7.8° and 7.7°, respectively. The findings are quite revealing since this is one of the primary purposes of using the investigational device instead of fusing the segment – to maintain the level of motion.

Translational motion (**Table 28**) was also measured throughout the course of the study by comparing lateral flexion and extension radiographs. Again, the postoperative values approximated the preoperative determinations. The mean values at every study period were very low, at less than 0.4 mm.

Radiographic success for the investigational group was based on 1) the existence of flexion/extension angular motion >4°, 2) no evidence of bridging trabecular bone forming a continuous connection between vertebral bodies, and 3) no radiolucency >50% of the convex surface of either the superior or inferior shell of the device.

If the two primary radiographic reviewers yielded conflicting success outcomes for a patient, a third reviewer was used for adjudication.

**Table 29** presents the radiographic success rates for the investigational patients at the various postoperative intervals. The success rates at all time periods were similar. At 12 and 24 months following surgery, the success rates were 81.8% and 79.6%, respectively. The major contributor to lowering the success rates was the angular motion component since bridging bone and/or radiolucency were not observed in any patients. The angular motion component yielded success rates between 78.0% and 81.8% at all measured postoperative time periods. This level of success is indicative of the relatively constant level of angular motion both before and after surgery. The mean values were consistently in a range from 6° to 8°. Considering that the standard deviation was typically over half of the mean value, one would expect some excursions out of the angular

motion success range, especially on the low end. Considering this, the mean angular motion values are arguably a better indicator of the effect of the prosthesis rather than a fairly arbitrary success criterion.

Lateral bending was evaluated by comparing the angular movements from left and right neck bending films. Throughout the postoperative course, the mean results were very consistent in a range of 4.0° to 4.4°. These results are shown in **Table 30**.

Lateral bending success was defined as motion ≥ 4°. Success rates for this measurement, which are presented in **Table 31**, were 54.0% and 49.7% at 12 and 24 months postoperative, respectively. These low rates are not unexpected considering the mean lateral bending values and the high standard deviations. Also, the success criterion was based on FDA's recommendation after review of the IDE protocol. It is our belief that the clinical relevance of this success criterion is not well-understood.

## Control Group

Radiographic success for control patients was based on the presence of fusion of the treated spinal segment. To be considered fused, there had to be radiographic evidence of bone spanning the two vertebral bodies in the treated segment. Additional criteria for fusion included flexion/extension angular motion stability (≤4°) and no radiolucent lines covering more than 50% of the graft surface. Fusion observations were performed by two radiographic reviewers. If their determinations of fusion status differed, a third reviewer was used to break the tie.

**Table 32** presents the fusion rates for the patients in the control group at the various postoperative intervals. The fusion rate at 6 months postoperative was less than 50%, which is unexpectedly low. This low rate can be attributed to the bridging bone component of the fusion criteria. Obviously with a plated cervical fusion procedure involving a fairly inert interbody spacer, i.e., structural allograft bone, the detection of bridging bone at early postoperative times without the use of CT scans would be very

challenging. [redacted] Notwithstanding these efforts, the fusion rate was uncharacteristically low. For a better perspective, the fusion rate at 6 months was 100% without bridging bone as a criterion. Therefore, the lack of obvious bridging bone at 6 months was not associated with excessive angular motion and lucent lines, i.e., the treated levels appeared stable. At 12 and 24 months, the fusion rates in the control group increased. The 24-month rate of nearly 93% approximates an expected historical level.

Furthermore, it appears that bridging bone revealed from plain radiographs does not affect overall success outcomes. From 6 months to 24 months, fusion rates increased from below 50% to 93%, but overall success rates stayed relatively constant during this same period at approximately 70%.

## 9. Angular Motion Measurements at Adjacent Levels

In order to determine the effect, if any, of the study treatment on adjacent levels, the stability of the cervical segments above and below the treated level was assessed. The measurements were made from flexion/extension films preoperatively and postoperatively beginning at 3 months.

**Table 33** provides a summary of the angular motion results. For the level above the treated segment, the mean preoperative values for the investigational and control treatments were similar at 8.3° and 7.8°, respectively. At 12 months, the angular motion values had increased slightly to 9.8° and 8.7°, respectively. The mean 24-month angular motion value for the level above in investigational device patients was 9.1°, as compared to 8.9° for control patients.

The mean preoperative angular motion values at the level below the treated segment were consistently less than those above the segment. The preoperative values for the investigational and control groups were 5.0° and 5.2°, respectively. The motion values remained fairly constant,

with only a mild increase, throughout the postoperative course for both treatments. At 24 months following surgery, the angular motion levels had increased from preoperative, with mean values of 6.4° and 6.2° for the two respective treatment groups.

Therefore, in summary, it appeared that the two treatments showed similar adjacent level angular motion outcomes following surgery. Motion at the level above the treated level tended to be higher than the level below the treated level. However, both levels experienced only a modest increase in motion for the two treatment groups. For investigational patients at 24 months following surgery, the mean value for treated segment tended to be lower than the level above and higher than the level below at 24 months (7.7° vs. 9.1° vs. 6.4°, respectively).

## E. Overall Success

Overall success at 24 months is the primary endpoint for the clinical study and it is the parameter on which the success of the clinical study is determined. Overall success is based on a patient demonstrating successful outcomes for NDI and neurological status. Also, to be considered an overall success, a patient could not have had a serious implant-associated or implant/surgical procedure-associated adverse event or have undergone a second surgery classified as a "failure". Therefore, this parameter encompasses important safety and effectiveness aspects of the treatment. **Table 34** provides this information for the two treatment groups at the postoperative periods.

At 24 months following surgery, the overall success rate for the investigational group was 80.6%, as compared to a 70.7% rate for the control group. Similarly, an approximate 11 percentage point difference was seen at 12 months.

Bayesian statistical analyses yielded a posterior probability of non-inferiority at 24 months of essentially 100%. The posterior probability of superiority was found to be 96.9%.

The 24-month overall success rates for the investigational group were found not only to be statistically non-inferior to the control group rates, but also superior. Therefore, the clinical study objective was met, indicating that the BRYAN® Cervical Disc System is as safe and effective as the current standard of care for treating cervical disc disease.

## F. Other Measurements

### 1. Gait Assessment

Assessments of patients' gait were made preoperatively and postoperatively using Nurick's classification.[15] Nurick's classification is based on a scale of 0 to 5, with a higher score signifying more impairment due to neurological status. Patients with a normal gait without nerve root or spinal cord symptoms were classified as "normal." Gait assessment outcomes for each postoperative study period are provided in **Table 35**. Less than 80% of the patients had "normal" gait scores preoperatively for both treatment groups. These values climbed considerably postoperatively, with 96.2% of the investigational patients and 96.4% of the control patients having "normal" values at 24 months following surgery. The success rate results, which were similarly high for all postoperative periods in both treatment groups, are given in **Table 36**.

### 2. Work Status

**Table 37** shows the work status of patients at various time points in the clinical study. In many ways, the data are difficult to interpret since many factors affect whether a patient returns to work or not, as well as the nature of the work performed when they return to work. From **Table 37**, it is evident that the work status of the investigational patients was no worse than that of the control patients.

Preoperatively, approximately 65% of both the investigational and control patients were working. At 6 weeks and 3 months following surgery, the rates of investigational patients who were working were higher than those for the control group, especially at 6 weeks. The differences in rates between the two treatment groups lessened over time. At 12 months, the rates were virtually identical, and at 24 months following surgery, the percent of working patients in the investigational group (74.8%) was modestly higher than that of the control group (72.1%).

Perhaps a better way to examine work status is to analyze the number of days from surgery to work return using Kaplan-Meier life table methods. Please refer to

[15] Nurick, S. The pathogenesis of the spinal cord disorder associated with cervical spondylosis. *Brain* 1972; 95: 87-100.

**Attachment M** for the results of such analyses comparing the investigational and control group. As evident from the data, investigational patients appear to return to work faster than control patients. The median return to work value for investigational patients was 48 days, as compared to 61 days for control patients. This 13-day difference was statistically significant (Log-Rank Test p=0.044, Wilcoxon Test p=0.016).

### 3. Patient Satisfaction

At each postoperative time point, patients were asked to respond to three statements pertaining to their satisfaction with the study treatment. These statements were as follows:

1. I am satisfied with the results of my surgery.
2. I was helped as much as I thought I would be with my surgery.
3. All things considered, I would have the surgery again for the same condition.

Each statement had a series of possible responses ranging from "definitely true" to "definitely false".

Summaries of the responses to the questions are provided in **Table 38**. At 12 and 24 months following surgery, investigational patients appeared to be more satisfied with their treatments than control patients; however, the levels of satisfaction were good for both treatments. At 24 months postoperative for the first question, 95.5% of the investigational patients and 92.9% of the control patients responded either "definitely true" or "mostly true". For the second question, 89.8% of the investigational patients and 83.5% of the control patients thought that they were helped as much as expected from their surgeries. Finally, 94.3% of the investigational patients said that they would have the surgery again, as opposed to a 90.7% rate for the control group.

In summary, the rates in both treatment groups were very high, and the investigational patients appeared to be at least as satisfied, if not more so, with their procedures as the control group patients.

### 4. Medication Summaries

Summaries of the medications taken by investigational and control patients at the various study periods are summarized in **Attachment N**.

## G. Additional Data Presentations

### 1. Examination of Effectiveness Variables by Investigator and Justification for Pooling Data Across Investigational Sites

Information pertaining to the effectiveness results at 12 and 24 months by investigational site is presented in **Attachment O**. The Breslow-Day test was used to assess the homogeneity of NDI, neurological, and overall success results across the sites. There were no statistically significant differences ($p<0.05$) noted in any of the comparisons, thus indicating that the results were consistent among different sites. These outcomes provide justification for pooling the data across investigational centers.

### 2. "Per Protocol" Results

A "per protocol" data analysis was performed, and the results are presented in **Attachment P**. The "per protocol" dataset was a subset of patients who were included in the primary analysis dataset. This cohort was previously described in Section III.B of this report.

The following table summarizes the results at 24 months following surgery.

| "Per Protocol" Success Rates | | | | |
|---|---|---|---|---|
| | Investigational | Control | Post. Prob. of Non-inferiority | Post. Prob. of Superiority |
| NDI | 88.5% | 84.3% | ~100% | 90.6% |
| Neurological | 93.4% | 91.0% | ~100% | 68.1% |
| Overall Success | 84.7% | 76.9% | ~100% | 94.4% |

Like the previous analyses, every statistical comparison for the "per protocol" dataset yielded a posterior probability of non-inferiority essentially 100%. Further, the investigational group approached statistical superiority for NDI and overall success.

In addition, the "per protocol" dataset was further refined by excluding any "out of window" visits, and similar analyses

were performed on it. The results and Bayesian analyses for this dataset are also provided in **Attachment P**. The statistical analyses showed that the results and conclusions were very similar to those obtained with "out of window" data included, and investigational group NDI, neurological, and overall success outcomes were still non-inferior to those of the control group. Investigational group superiority for the overall success variable was also found.

In summary, despite the smaller sample size, the "per protocol" results mimic those of the larger primary dataset, thus attesting to the uniformity and consistency of the data. For the analyses involving the most stringent condition of excluding patients who were not in their visit windows, the overall success rates for the investigational group were found to be not only non-inferior to those of the control group, but also superior.

### 3. "Missing Equals Failure" Results

The "missing-equals-failure" data presentations for various study periods are included in **Attachment Q**. For this presentation, deaths, patients lost-to-follow-up, and missing observations due to other causes resulted in missing observations for the outcome variables, and, therefore, were included in the denominators of the calculated rates, i.e., considered as "failures." By treating these patients as treatment failures, the clinical outcome rates in the "missing-equals-failure" analyses were lower than those observed in the clinical data. The 24-month overall success rate for the investigational group was higher than that of the control group (76.8% vs. 60.0%).

### 4. Sensitivity Analysis for Assessing Missing Values

Postoperative follow-up rates were presented earlier in this report (Section II.C.1). From **Table 2**, there is an apparent 10-point difference in rates between the two treatment groups at 24 months. The difference favored the investigational group, which had a very high value of 95.2%. The control group rate was a respectable 85.4%. In order to understand the impact, if any, of this disparity, a sensitivity analysis was performed. In the interim analysis cohort, eight (4.8%) of 168 investigational patients did not have overall success outcomes, as compared to 25 (15.2%) of 165 control patients. To assess the impact of lost-to-follow-up patients on study conclusions, we performed a sensitivity

analysis of overall success at 24 months by various imputations for the missing outcomes, the results of which are presented in **Attachment R**. The analyses were focused on the 24-month data and used simple frequentist calculations.

The results show that even in the worst case scenario (where all missing investigational outcomes are assumed to be failures and all missing control outcomes are assumed to be successes), which is grossly biased against the investigational group, non-inferiority of the investigational treatment to the control can still be claimed (p=0.007). When 0% of missing investigational outcomes and 50% of the missing control outcomes are assumed to be successes (which still grossly favors the control group), the superiority of the investigational treatment to the control can be claimed (p=0.034). These results indicate that the study conclusions with regard to both non-inferiority and superiority are robust, even considering lost-to-follow-ups and the impact of missing observations.

## 5. Correlations between 12-Month and 24-Month Results

Analyses were performed to examine the relationships between certain key endpoints at 12 and 24 months postoperative. The results for the primary and "per protocol" dataset are presented in **Attachment S** and are summarized in the table below.

| **Percent Agreement Between 12- and 24-Month Data** | | | | |
|---|---|---|---|---|
| | Primary Dataset | | "Per Protocol" Dataset | |
| | Investigational | Control | Investigational | Control |
| NDI | 88.5% | 80.5% | 88.8% | 82.7% |
| Neurological | 93.6% | 90.9% | 93.3% | 92.8% |
| Overall Success | 86.1% | 78.8% | 85.9% | 80.4% |

It is readily apparent that there is good agreement between the 12- and 24-month outcomes. This means that there is a high likelihood of a patient in either treatment group having the same outcome at the two latter study periods. This is especially important for Bayesian analyses since it strengthens the inferences that can be made.

## 6. Correlation between Pain and Disability Outcomes and Angular Motion Measurements

The relationships between NDI, neck pain, and arm pain results and angular motion values were examined in

investigational device patients to determine if there was any correlation between the degree of segmental motion and pain. The results of this analysis are located in **Attachment T**. No statistically significant correlations were noted at any of the postoperative time periods.

## 7. Financial Disclosure Information and Analyses

Financial disclosure information pertaining to the investigators and co-investigators who participated in IDE [redacted] is provided in **Attachment U**. The information indicates that 14 of 65 (22%) surgeons who performed surgeries met the criteria for having a financial interest at some point during the course of the clinical study. These surgeons contributed 142 patients to both treatment groups. At 12 and 24 months postoperative, there were only two statistical differences in the outcome comparisons between the patients of surgeons with a financial interest versus those without. The difference was found in the NDI and overall success comparisons for investigational patients at 12 months postoperative. The investigational patients from surgeons with a financial interest had lower NDI and overall success rates than those from surgeons not meeting the financial interest criteria. This finding certainly shows that an interest did not tilt the results in favor of the new treatment. These statistical differences at 12 months did not surface in the 24-month comparisons.

Therefore, it is apparent that the existence of an investigator financial interest, as defined in 21 CFR 54, did not materially impact the results or cast any doubt over conclusions of the study.

## 8. Currently Available Data Presentations

As previously stated, this submission is primarily based on the pre-defined interim analysis criteria, using an interim analysis cohort of the first 300 patients (ordered by the surgery date) who had valid overall success outcomes at 24 months.[16] Data and analyses are available for the interim analysis patients plus those patients not in that cohort, referred to here as "all currently available 24-month data". These tables are included in **Attachment V**.

[16] Due to either missing baseline or postoperative elements or missed postoperative visits at 24 months, the first 333 patients yielded the 300-patient interim analysis cohort.

A comparison of the success rates of certain key variables for the two cohorts is made in the following table.

| 24- Month Success Rates for Interim Analysis Cohort vs. All Currently Available Data (Primary Dataset) | | | | |
|---|---|---|---|---|
| | Interim Analysis Cohort | | All Currently Available Data | |
| | Investigational | Control | Investigational | Control |
| NDI | 84.3% | 75.7% | 86.5% | 78.3% |
| Neurological | 93.7% | 91.4% | 93.2% | 90.3% |
| Overall Success | 80.6% | 70.7% | 82.7% | 72.6% |

As can be seen from this table, the outcomes are very similar regardless of which dataset is used. For the "all currently available 24-month data" cohort, the investigational group results were found to be statistically non-inferior to the control group in all comparisons. Statistical superiority was found for NDI, arm pain, and overall success. "Per protocol" analyses of NDI, neurological status, and overall success demonstrated investigational group non-inferiority for all these parameters as well as superiority for overall success.

Therefore, we believe the interim analysis cohort results and the conclusions drawn from them will apply to the total study population. This is especially true when one considers that the adverse event information previously presented in this report is for "All Currently Available Data".

### 9. Data Listings

Data listings for the investigational and control patients are provided in **Attachment W**.[17]

## V. Conclusions

The goal of the BRYAN® Cervical Disc System IDE clinical study was to evaluate the safety and effectiveness of the use of the device in the treatment of patients with symptomatic cervical disc disease when compared to the control treatment, a standard of care fusion procedure using structural allograft bone with an anterior cervical plate. As demonstrated in this report, the clinical results from the use of the investigational device, the BRYAN® Cervical Disc, compared very favorably to the control group results.

[17]

The investigational device was found to be at least as safe as the control treatment. The rate of investigational group patients having at least one adverse event was similar to the control group rate and not statistically different. Investigational patients had lower rates of serious adverse events that were classified as implant- and implant/surgical procedure-associated.

The investigational group had a statistically lower rate of secondary surgical procedures related to supplemental fixations. Investigational patients also had similar rates of revisions and implant removals to control patients. Overall, there was a lower second surgery failure rate for investigational patients.

Maintenance or improvement in neurological status was found in greater than 93% of patients in the investigational group, which was statistically similar to the control group. Based on the favorable neurological status outcome, as well as the adverse event and second surgery rates, the results of this study certainly support the safety of the BRYAN Cervical Disc.

In terms of primary dataset effectiveness measures, NDI scores improved dramatically after surgery for both treatment groups, exceeding 28 points at 12 and 24 months. The mean improvements in NDI scores for the investigational group at these periods were consistently higher than for the control group at all postoperative periods. A comparison of the NDI success rates (based on a 15-point improvement from baseline) showed that the investigational group had higher rates than the control treatment at all postoperative time periods. Statistical superiority to the control group was demonstrated at 24 months. Neck and arm pain results were statistically similar for both treatment groups.

Radiographically, mean angular motion values for investigational patients were slightly higher after surgery than prior to surgery, thus indicating that the device maintains the motion of the treated level. This is one of the intended functions of the device. In terms of adjacent level motion, the two treatments showed similar performance following surgery. Motion at the level above tended to be higher than the level below. FSU height was maintained postoperatively in a very high percentage of patients in both treatment groups (success rates exceeding 95%).

Investigational patients returned to work more quickly than control patients. The median time for investigational patients was 48 days, which was 13 days shorter than the time for control patients. This difference was statistically significant.

Overall success was the primary endpoint for the clinical study, and it is the parameter on which the success of the clinical study is determined.

Overall success is based on a patient having a successful NDI outcome and neurological status maintenance or improvement. Also, to be considered an overall success, a patient could not have undergone a second surgery classified as a "failure" or have had a serious adverse event that was judged as implant- or implant/surgical procedure-associated. Therefore, this parameter encompasses important safety and effectiveness aspects of the treatment. At 24 months following surgery, the primary dataset overall success rate for the investigational group was 80.6%, approximately 10 percentage points higher than the 70.7% rate for the control group. A similar difference was also seen at 12 months. The overall success rates for the investigational group were found not only to be statistically non-inferior to the control group rates, but also superior. Therefore, the clinical study objective was met, thus indicating that the BRYAN® Cervical Disc System is as safe and effective as the current standard of care, fusion, for treating cervical degenerative disc disease.

The data and information presented in this PMA application provide a reasonable assurance of the safety and effectiveness of the BRYAN Cervical Disc System and should lead to the approval of the device.