

Report of United States Clinical Study Results (G010188) – PRESTIGE® Cervical Disc System ¹

EXECUTIVE SUMMARY

A prospective, randomized, controlled IDE clinical study was conducted to evaluate the safety and effectiveness of the PRESTIGE® 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 541 patients participated, with 276 receiving the PRESTIGE (investigational) device and 265 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 very similar to the control group rate. This was also true for serious adverse events. Investigational patients had a lower rate of adverse events that were classified as implant- or implant/surgical procedure-associated. The radiographic reviewers noted no implant migration or fractured/broken implants in the investigational group, while some did exist for control patients. Of particular note, the investigational group had statistically lower rates of second surgical procedures related to revisions and supplemental fixations. The rate of removals was also lower, but not statistically different. The investigational group neurological success rate was statistically higher than the control group.

In terms of effectiveness measures, NDI score improvement following surgery was dramatic for both treatment groups, exceeding 30 points at 12 and 24 months. The mean improvements in NDI scores for the investigational group at these periods were nearly 2 points higher than those of the control group. A comparison of the NDI success rates showed that the investigational group was non-inferior to the control treatment. Neck and arm pain results also showed non-inferiority of the investigational treatment to the control.

¹



Radiographically, mean angular motion values for investigational patients were very similar before and 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. FSU (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 for investigational patients was 45 days, which was 16 days shorter than the time for control patients. This difference approached statistical significance.

At 24 months following surgery, the overall success rate for the investigational group was 80.5% and approximately nine percentage points higher than the 71.3% rate for the control group. A similar difference of 11 percentage points was also seen at 12 months. Overall success rates were also calculated by adding FSU (disc height) success to the formula. Again, at 12 and 24 months following surgery, the investigational group overall success rates were impressively higher than the control group rates, with the 24-month rate being nearly 17 percentage points higher. Regardless of the definition used, 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 PRESTIGE® Cervical Disc System is as safe and effective as the current standard of care, fusion, for treating cervical disc disease.

I. Introduction

In July of 2001, Medtronic Sofamor Danek filed an application for an Investigational Device Exemption (IDE) (G010188) with the FDA to study the use of the Artificial Cervical Disc in patients with symptomatic cervical disc disease. [REDACTED]

Clinical study surgeries were performed during a period from October 3, 2002 to August 19, 2004. A total of 541 surgeries were completed, consisting of 276 investigational patient surgeries and 265 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.

For future marketing purposes, the Artificial Cervical Disc device will be known as the PRESTIGE® Cervical Disc System. Hence, the device will be referred to by the latter name or as the investigational device in this report of clinical study results.

II. Clinical Study Description

A. Clinical Study Goals and Design

The goal of the IDE clinical study of the PRESTIGE® Cervical Disc System 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 PRESTIGE Cervical Disc were through direct clinical data comparisons between data collected from patients implanted with the PRESTIGE device and an equivalent group of patients who received a surgical fusion utilizing structural allograft bone with plate stabilization. The investigational and control treatments were randomized in a 1:1 manner.

Per the protocol, patient evaluations occurred preoperatively (within 6 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). Evaluations continued annually thereafter until the last subject enrolled in the study was seen for his/her 24 month evaluation.

The effectiveness of the PRESTIGE 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. An alternate overall success assessment was made including functional spinal unit (FSU) height maintenance or improvement along with the aforementioned criteria. 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, 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, one supplement increased the sample size from 450 to 550 patients, and another submission redefined disc height success to be based on the measurement of functional spinal unit height. 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 patients enrolled in the original IDE study occurred and the PMA application was being prepared and processed. Approximately 60 patients have been enrolled in the continued access program and information concerning them can be found in **Section IV.C**, "Other Relevant Clinical Data", in this module. All of these patients are within the first 12 months postoperative at this time.

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 2a**. These tables also provide patient evaluation distributions as a function of time within each study period. A total of 276 patients received the investigational device, and the control group had a total of 265 patients. The date of database closure for analyses was May 2, 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 88% at 24 months. The 24-month

follow-up rate for the investigational group was 93.4%, compared to a control group rate of 82.4%.

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 2a further examines the overall success information as a function of windowing. It also shows when “out of window” patients were observed. At 12 and 24 months following surgery, over 85% 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 3**. 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 except for the use of alcohol. Fewer investigational patients reported alcohol use than control patients (43.5% vs. 53.2%, $p = 0.025$). This finding is believed to be of little clinical significance to the ultimate study outcomes, especially considering the nature of the question asked of the patients, which was simply a yes/no question as to whether alcohol was consumed. Information concerning the number of drinks per week was not sought. Summarily, the investigational and control patients were demographically similar.

3. Preoperative Medical Status

a. Prior Medical History

Table 4 is a summary of the preoperative medical history information for investigational and control patients, based on over 40 medical history questions pertaining to the cardiovascular, endocrine, gastrointestinal, genitourinary, musculoskeletal, and respiratory systems of the body. There were no statistically significant differences ($p < 0.05$) in responses between investigational and control patients.

b. Preoperative Medical Condition and Medication Usage

Summaries of the patients' preoperative medical conditions and medications are provided in **Table 5**. These included the time interval from the initiation of symptoms to surgery, the number of prior neck surgeries, and the types of medications being used to alleviate the symptoms. There were no statistically significant differences ($p < 0.05$) between treatment groups for any of the variables.

c. Preoperative Evaluations of Clinical Endpoints

Table 6 summarizes the preoperative status of key effectiveness endpoints for the treatment groups. These analyses focused on NDI, SF-36 component summaries, and arm and neck pain. There were no statistically significant differences ($p < 0.05$) between the treatment groups for any of the examined variables.

d. Summary

The analyses presented above provide a robust examination of the preoperative medical status of investigational and control patients. None of the comparisons showed a statistically significant difference. Coupling this information with the demographic assessments, one can easily conclude that the patients in the two treatment groups were very similar. This is noteworthy in drawing ultimate conclusions from the study regarding the safety and

effectiveness of the investigational treatment, because it demonstrates that the results are based on a treatment effect rather than confounding factors.

4. Consented Patients Who Declined Participation Prior to Surgery

Eighty-four (84) patients were randomized but declined participation in the study prior to receiving the assigned treatment. Thirty-six (36) of these patients would have received the investigational treatment, while 48 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	11	10
Condition Improved	5	8
Dissatisfied with Randomization	2	11
Inclusion/Exclusion Criteria Not Met	7	4
Combination*	0	1
Other**	7	12
Unknown	4	2
Total	36	48

* Combination of Condition Improved and Dissatisfaction with Randomization

** "Other" includes the following: needed 2-level ACDF, decided not to participate, went to another surgeon, no-show for surgery, wanted larger settlement/seeking new lawyer, posterior/lateral approach required, waiting on attorney to approve surgery, EMG and nerve conduction study indicated carpal tunnel syndrome and no radiculopathy, "got cold feet", and not cleared for surgery.

The aforementioned demographic and preoperative medical status data were collected for these patients, and statistical comparisons were made to compare these patients to those who did receive the study treatment. These data are presented in **Table 7** parts a-d (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 over 65 comparisons made between study and non-study patients for each of the two treatment groups, and over 90% of them did not produce statistically significant differences ($p < 0.05$).

For investigational patients, study and non-study patients had statistically different results in six assessments. The non-study patient cohort had a higher percentage of non-Caucasians, more Worker's Compensation patients, and more tobacco users. The study cohort had more patients who were working at the time of study enrollment and a

higher percentage of patients with complaints of headaches. The educational level findings were mixed. The study cohort had a higher percentage of high school graduates, while the non-study group had more patients who either did not finish high school or who had post high school training.

For control patients, there were four assessments that revealed statistical differences. Non-study patients were statistically taller, had higher SF-36 MCS scores, and had lower (better) neck pain scores than study patients. In addition, more non-study patients were working at the time of study enrollment.

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 preoperative medical 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

Seventy-two (72) investigators and co-investigators from 34 sites performed surgeries in this clinical study. No single investigational site contributed more than 10% 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** in this module for financial disclosure forms for the investigators; this section only contains information 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 overall successful 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 focused on determining if the investigational group demonstrated superior overall success results, as compared to the control group. These

objectives were also focused on determining 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, analyses were proposed to determine if the investigational group had superior outcomes when compared to the control group for those individual variables, if non-inferiority were established. 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

As pre-specified in the Statistical Considerations of the Clinical Investigational Plan (CIP), three different analysis datasets (that is, primary, per-protocol, and missing-equals-failure datasets) were defined. The primary dataset consisted of all the patients who received study devices² and completed surgical procedures. In the rare event that 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 was one such case in this study where a patient was randomized to receive the investigational device but received the control treatment). Primary statistical comparisons were based on the observed data and missing data due to lost-to-follow-ups were not

²

[REDACTED]

[REDACTED]

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 and those who received wrong study treatments (patients who were randomized as control but actually received the investigational treatment or vice versa), or 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 way 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 reasons, 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



[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

3

[REDACTED]

[REDACTED]

[REDACTED]

D. Bayesian Interim Analysis

The Statistical Considerations pre-defined that the data would first be analyzed after a total of approximately 250 patients (investigational and control combined) had follow-up visits at 24 months. At that time point, all the patients were 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.

[REDACTED]

This PMA application is primarily based on the pre-defined interim analysis criteria, using the first 250 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 the patients were also included and presented. Those data have been monitored in an appropriate manner, cleaned, and verified.

Because of the time required for data monitoring and cleaning, more than 250 patients have had 24-month visits as of the cutoff date, May 2, 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.7 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 8 provides summaries of information related to the surgical procedures and postoperative hospitalizations of patients. The results of the statistical analyses between the investigational and control groups are provided in **Attachment D**. The mean operative times for the investigational and the control treatment groups were 1.6 hours and 1.4 hours, respectively. These mean operative times were statistically different based on Bayesian analyses. This statistical difference is due to overpowering from the large sample sizes. Arguably, a mean difference of only 0.2 hours, or 12 minutes, is of little clinical significance. This is especially true considering that the investigational device surgical technique was new to the investigators, which likely contributed to the slight difference observed.

Investigational patients were found to have similar estimated blood loss to the control group patients (60.1 ml versus 57.5 ml). The median blood loss was 50 ml in each treatment group. We believe there is no clinical difference in the estimated blood loss values.

The mean hospital stays of patients in both treatment groups were approximately one day. However, due to the large sample size, the mean hospital stay for investigational patients was found to be statistically higher than that for the control patients (1.1 days vs. 1.0 days, respectively). Both treatment groups had a median hospital stay of one day. Again, this difference is believed to be of little clinical relevance, especially considering the frequency of inpatient classifications. Of the investigational patients, 70.3% were inpatients, which compares well to the 69.1% rate for control patients.

Even though statistical analyses were not performed, it is evident that the distributions of the patients in the two treatment groups for the variables of treatment level, external orthosis use, and operative approach were very similar. Over 90% of the patients in both treatment groups had procedures at either C5-C6 or C6-C7.

In summary, investigational device patients had similar operative and discharge time values to control group patients. These findings are considered beneficial for the clinical trial since they indicate that both treatment groups had similar procedures and were treated similarly postoperatively.

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, or complications, 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 21 categories of adverse events, such as neurological, infection, dysphonia/dysphagia, etc. If the underlying cause of the adverse event is known, it was categorized accordingly. If the underlying cause is 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 were graded as a “3” or “4”, it was considered “serious”; otherwise, it was considered “non-serious”. Typically, adverse events that result in an emergency room visit or a hospitalization were regarded as “serious”.

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 would cause them to be considered a second surgery failure are identified with specific language in the adverse event narrative.

Table 9 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.⁴ Statistical

⁴ 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

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 226 (81.9%) investigational patients had at least one adverse event. Similarly, the number of patients in the control group with any adverse event was 212 (80.0%). 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. 6.0%) categories. Such events were not possible for investigational patients. Investigational patients also had a lower rate of spinal events (6.2% vs. 11.3%). These are adverse events that can occur at any location in the spine, including the treated level. In addition, the rate of urogenital adverse events statistically favored the control group over the investigational group (1.9% vs. 5.4%).

As shown in **Table 10**, the number of patients having serious adverse events, i.e., those with a WHO grade of 3 or 4, in the investigational group was less than that found in the control group (27.9% vs. 29.8%). This difference was not statistically different. The treatment group rates for the various categories were fairly similar. The largest gap was noted for cardiovascular related events, which favored the control group over the investigational group (0.8% vs. 3.6%). A more specific discussion of these events is provided later in this report. There were more reports of cancer in the investigational group (1.4% vs. 0.8%) than in the control group. None of these was considered related to the treatments. These events will also be discussed below in more detail.

Table 11 summarizes the adverse events that are classified as implant-associated or implant/surgical procedure-associated.⁵ The number of patients with these types of

regarding adverse events, if any occurred, after 24 months. Information regarding adverse events that occurred preoperatively (after 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**.

⁵ 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.

adverse events was lower in the investigational group than in the control group (3.3% vs. 9.8%). Upon closer examination, these event rates were fairly similar for both treatment groups in most of the specific categories. The main exception was pending non-union events, where the control group rate was 6.0%, as compared to a 0.0% rate for investigational patients. This gap is believed to be the main contributor to the overall rate difference.

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 three other categories of adverse events (cancer, cardiovascular, and deaths) due to the nature of the events. The urogenital category is also discussed because there was a statistical difference between the two treatment groups.

Neurological

A total of 78 neurological events occurred in 66 patients in the investigational group (23.9%). Twenty (20) of these events were rated Grade 1, 50 were rated Grade 2, eight were rated Grade 3, and none were rated Grade 4.

Out of the 78 events, the most commonly reported event among investigational patients was numbness (22 events). Of these 22 events, 17 involved the upper extremities (arms, hands and fingers). There was one report each of general numbness, numbness on the right side in a C6 distribution, and tongue numbness. In addition, there were two events that occurred in the lower extremities – numbness in the left foot and numbness that affected the left little toe, heel area and right ankle.

The next most frequently reported neurological events in investigational patients involved paresthesia, tingling, numbness and tingling, numbness and pain, and radiculopathy. There were five events of paresthesia affecting the arm, feet, and hands; two events affecting the hand; and one event of nocturnal paresthesia in both hands. There were four events of tingling affecting the hand, the elbow to the left hand, bilateral arms, and the hands and

fingers. There were seven events of numbness and tingling affecting the bilateral upper extremities, left hand, and the right middle finger, and two events affecting the left arm, the hands, and the wrist. There were five events of numbness and pain affecting the hand and elbow, fingers, neck, shoulder, and back, and two events affecting the arm. There were six events of radiculopathy in the C8 area, two events affecting the arm, two events affecting the left side, and one event involving neck spasms.

There were four events of weakness and two events of paresthesia and pain. The four events of weakness affected the elbow, shoulder, or arm (two events). The two instances of paresthesia and pain involved arm/cervical radiculopathy and the left arm or neck. There were two events of numbness and weakness in the leg and foot or arms and hands.

There were 21 events that only occurred once in the investigational group. These events included: C5 radiculopathy associated with shoulder pain; paresthesias/hypoesthesias; numbness and paresthesias; sciatica; Bell's Palsy; axillary pain/bilateral; bodily shocking sensations; burning and tingling in the hand; radiating pain in the neck; dizziness and numbness; decreased sensation in the little finger with neck, right shoulder, and right arm pain; numbness in the hands and arms with neck spasms; restless legs; pain with "needles" in the left arm and right foot; "pins and needles" in the thumbs and forearms; left trapezius shooting pain and weakness; neck swelling and stiffness with left hand numbness; numbness and weakness in the foot with low back pain; numbness, tingling, and pain the arm; pain radiating in the right arm associated with numbness; and back and leg pain thought to be associated with a previous L5-S1 disc bulge.

A total of 65 neurological events occurred in 55 patients in the control group (20.8%). Twenty (20) of these events were rated Grade 1, 39 were rated Grade 2, six were rated Grade 3, and none rated Grade 4.

Out of the 65 neurological events that occurred in control patients, the most frequently reported event was numbness. There were 15 events that involved the upper extremities (arms, elbow, hands and fingers). In addition, there were two occurrences of chin numbness. There was one instance involving numbness of the bilateral lower extremities, and

one report of general numbness involving the right side at C6-C7.

The next most frequently reported neurological event involved numbness accompanied by tingling, pain, tightness, and/or burning. There were four events involving numbness associated with tingling and pain. These events involved the right arm, shoulder and neck, finger and hand with shoulder pain, and thumb with neck pain. There were seven events that involved only numbness and tingling. These events included one instance involving the bilateral upper and lower extremities; two instances involving the bilateral hands; one instance involving the arms, hands, and feet; one instance involving the right arm and fingers; one instance involving the low back; and one instance involving the neck.

There were three events involving numbness associated with radiating pain. These events included one occurrence each in the groin; the arms and legs; and in the back, hip, thigh, and foot. There were two events that involved numbness and pain. These events involved numbness in the thigh with general spine pain, and numbness with pain in the shoulder and finger. There was one event involving numbness and burning of the left shoulder. There was also one event involving numbness of the arm and tightness of the scapula.

The next most commonly reported events involved paresthesia, weakness, and tingling. There were five events involving paresthesia. These were located in the thigh, calf, and foot, the left leg and foot, the right hand, the arm and hand, and the fingers of the left hand. There were five events that involved weakness of the right tricep, left arm, right arm, left deltoid, and bicep. There were three events involving tingling. These events included tingling in the right leg, the right hand, and the left hand.

Finally, there were two events that involved sciatica, of which one included numbness and weakness in the left leg. There were 13 events that only occurred once. These events included: hypersensitivity, myelopathy, hyperpathia, decreased pin prick, dysesthesia of the third and fourth digits, ulnar neuropathy, muscle hyperexcitability and twitching, hemisensory loss of the left side along with walking difficulty and visual obscuration, seizure, transverse myelitis, involuntary movements of the thumb and body, dizziness, and cervical myalgia/paresthesia.

Spinal Events

A total of 18 spinal events occurred in 17 patients in the investigational group (6.2%). The most frequently reported events were lumbar-associated (14 events). These events included the following: three herniated discs, three reports of degenerative disc disease, two cases of spondylolisthesis/lithesis, two cases of stenosis, two disc bulges, one collapsed disc, one transition syndrome, and one post-laminectomy syndrome. Additionally, there was one disc herniation in the thoracic spine.

Also reported were three cervical-associated events. There was one report of disc herniation and one report of degenerative disc disease that occurred adjacent to the level at which the investigational device was implanted. One patient was reported to have a mild, degenerative spondylolisthesis at the treated level 12 months postoperatively and was continuing treatment with medications and physical therapy.

In the control group, there were 32 spinal events noted in 30 patients (11.3%). Again, the most frequently reported events were lumbar-associated (15 events). These included seven reports of degenerative disc disease with stenosis or herniated disc, five herniated discs, two cases of stenosis, and one spondylolisthesis. There were four thoracic-associated events: three herniated discs and one degenerative disc disease.

In addition, there were 13 cervical-associated events reported. These included three herniated discs, three herniated discs and disc bulges, and two cases of stenosis. Finally, each of the following occurred once: degenerative disc disease, ossification at an adjacent level, stenosis/herniated disc, kyphosis/stenosis, and spurs/disc bulge.

As can be seen from the descriptions above, the nature of these adverse events was quite variable, as would be expected considering the non-specific nature of this category. However, it is noteworthy that the rate of adverse events categorized as “spinal events” was statistically lower for the investigational group. Furthermore, the incidence of cervical spine-related events was lower in the investigational group than in the control group (3 vs. 13 events).

Trauma

A total of 69 trauma events occurred in 59 patients in the investigational group (21.4%). The most frequently reported events were motor vehicle accidents (22 events) and falls (13 events). Also reported were five work-related injuries, two assaults, three lacerations, two patients with “jarred necks”, two reports of lifting injuries, two back injuries, and two events caused by sudden movements. Additionally, there were 16 events that occurred only once and were reported as trauma. These included: dog bite; snake bite; running injury; right hand tendon injury; torn left wrist tendon; rotator cuff tear; tractor accident; amusement park ride trauma; injury when hit with a heavy gate; “head popped”; severed digits; contact injury (hugged too hard); an injury due to yoga; and injuries to the eye, hand, and shoulder.

In the control group, there were 47 trauma events noted in 40 patients (15.1%). Similarly, the most frequently reported were falls (18 events) and motor vehicle accidents (10 events). There were seven reported work-related injuries. Two injuries were secondary to over activity and exercise-induced, and there were two meniscal (knee) tears. Additionally, there were eight events that occurred only once and were reported as trauma. These included: eye injury, hernia, muscle strain, neck strain, pulled muscle, “jammed head”, paddle boat falling on the shoulder, and a finger slammed in the door.

As evidenced by the above information, trauma events vary in cause and severity. The rate of incidence in investigational patients was observed to be higher than that of the control group (21.4% vs. 15.1%); however, it was not statistically different.

Neck and/or Arm Pain

A total of 190 events occurred in 138 patients in the investigational device group (50.0%). The events included the following: 42 neck pain; 35 shoulder pain; 20 neck and shoulder pain; 11 arm pain; 10 neck spasms; 10 neck and arm pain; eight interscapular/scapular pain; seven shoulder and arm pain; five epicondylitis; four neck and headache; four trapezius pain; three wrist pain; three rotator cuff events; three neck, shoulder and arm pain; three hand pain; and two shoulder tendonitis.

Additionally, there were 19 events that occurred only once and were reported as neck and arm pain. These included acromial/clavicle pain; deltoid pain; elbow pain; jaw pain,

muscle strain; cervical strain; elbow, wrist and hand pain; neck and ear pain; scapular pain with migraine headache; neck, shoulder, and upper back pain; shoulder, arm and hand pain; submastoid and arm pain; T1-T2 pain; shoulder joint degeneration; neck and shoulder pain with arm numbness; neck and scapular pain; neck and thoracic pain; shoulder impingement; and trapezius pain.

By comparison, a total of 173 neck and/or arm pain events occurred in 127 patients in the control group (47.9%). These events included 58 neck pain; 24 shoulder pain; 20 arm pain; 11 neck and shoulder pain; 10 neck and arm pain, nine neck spasms; nine interscapular/scapular pain; eight trapezius pain; six rotator cuff events; four neck pain with headache; three epicondylitis; two neck, arm, and shoulder pain; two elbow pain; and two shoulder impingement.

Additionally, there were five events that occurred only once and were reported as neck and arm pain. These included neck and scapular pain, radiating pain, neck and upper back pain, glenohumeral joint pain, and clavicle/scapular pain.

The reporting of neck and arm pain as an adverse event in a study such as this is questionable since they were measured effectiveness endpoints. However, these were recorded in the spirit of providing a complete picture of the study treatments. Cumulatively, nearly half 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 cumulative rates for the two treatment groups were very similar and not statistically different.

Other Pain

A total of 88 events classified as “other pain” occurred in 69 patients in the investigational group (25.0%). The most frequently occurring categories were back pain (28 events) and headaches (22 events). In addition, there were 10 reports of back and leg pain, seven reports of hip pain, six reports of knee pain, four reports of leg pain, four reports of thoracic pain, and two reports of back and hip pain. Each of the following types of pain was reported once: back/thoracic, hip/leg, bursitis, knee/ankle, and flank.

A total of 68 events classified as “other pain” occurred in 56 patients in the control group (21.1%). Again, the most

frequently reported events were back pain (27 events) and headaches (16 events). In addition, there were four reports of back and leg pain, four reports of knee pain, three reports of back and hip pain, three reports of leg pain, and two reports of hip pain. Each of the following types of pain was reported once: flank/pelvis, flank/knee, foot, abdominal, back/pelvis, and leg/hip. Finally, there was one occurrence each of thoracic pain, incision pain, and sacroiliitis.

None of the events in this category were related to the cervical spine. The rate for the investigational group was higher than that of the control group, but the difference was not statistically significant.

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 109 events in occurred in 70 investigational patients (25.4%), and 82 events occurred in 66 control patients (24.9%). The nature of these events for both treatment groups is presented in **Attachment G**. The rate of “Other” adverse events for the investigational group was not statistically different from that of the control group.

Cancer

The incidence rate of cancer in the investigational group was 1.8% (5 patients). This compares to a 0.8% rate (2 patients) for the control group. The rates were not statistically different. None of the cancers in either group were judged to be related to the study treatment.

Of the five investigational patients, the first was a non-Hodgkin’s lymphoma, which occurred approximately 26 months following surgery. The diagnosis was made based on a biopsy of a nasal mass found on a CT performed to investigate the patient’s complaint of hearing loss. At 28 months postoperative, the patient was admitted to the hospital for the placement of a Mediport. At 30 months postoperative, the patient underwent a PET scan, an additional CT, and a bone marrow biopsy.

The second report of cancer was basal cell carcinoma reported at the patient’s 24-month visit. The patient reported having undergone biopsies and numerous excisions and

topical treatments. The event is considered resolved since it was reported that the excision produced clean margins.

The third report of cancer occurred at 17 months postoperative when the patient was diagnosed with colon cancer. At 18.5 months, the patient underwent surgery to remove the colon cancer.

The fourth report of cancer was reported at the 24-month visit. The patient reported being seen by an ENT who diagnosed a thyroid cancer. The patient had several second opinions and received the same diagnosis. The patient refused surgery and is being treated holistically with herbs and minerals.

The fifth cancer reported in the investigational group was a breast carcinoma occurring approximately 17 months following surgery. The patient had a right breast lumpectomy, biopsy, and chemotherapy.

Of the two patients in the control group for whom a cancer was reported, the first was at approximately 23 months following surgery when the patient reported a reoccurrence of skin cancer. The patient had an excision of the malignant neoplasm on the left cheek with complex closures of the acquired defect. It is noted that the patient had a history of skin cancers.

The second report of cancer in the control group occurred about 7 months following surgery. The patient was found to have a brain tumor, and subsequently a craniotomy was performed to remove the tumor. The tumor was reported by the investigational site to be a low grade, non-metastatic astrocytoma unrelated to the cervical spine. The patient is continuing follow-up with the neurosurgeon who performed the surgery.

Urogenital

A total of 16 events occurred in 15 patients in the investigational group (5.4%). There were two events reported as incontinence. The other 14 events only occurred once in the investigational group. These events included testicular pain, abdominal pain, urinary urgency, urinary dribbling, urinary difficulties, stress incontinence, epididymitis and orchitis, prolapsed uterus, hysterectomy, erectile dysfunction secondary to hyperprolactinemia, irregular

menstrual bleeding secondary to an ovarian cyst, endometriosis, and problems with urination and impotence.

A total of 6 events occurred in 5 patients in the control group (1.9%). The most commonly reported event was incontinence (3 events). There was one report each of irregular bleeding, endometriosis, and impotence.

As stated earlier in this report, the incidence rate of urogenital adverse events was statistically lower in the control group. However, as can be seen from the nature of the specific adverse events detailed above, these occurrences are not believed not to have been caused by the study treatments, either investigational or control.

Cardiovascular

A total of 15 cardiovascular events occurred in 14 patients in the investigational group (5.1%). These events included the following: four incidents of chest pain, three arrhythmias, two myocardial infarctions (not resulting in death), and two cases of carotid artery disease. Each of the following was reported once: hypertension, an extra heartbeat, coronary artery disease, and leg claudication.

In the control group there were nine cardiovascular events noted in eight patients (3.0%). These events included three occurrences of hypertension and three reports of palpitations. Each of the following was reported once: deep vein thrombosis, myocardial infarction (not resulting in death), and arrhythmia (atrial fibrillation/flutter).

Deaths

There were no deaths in the investigational device group.

In the control group, there were three deaths reported (1.1%). The first was a 68-year-old female who suffered a fatal myocardial infarction at home 3 months following surgery. The patient had a previous history of cardiac problems. The second death occurred approximately 11 months postoperatively when the patient suffered an acute myocardial infarction. The patient was taken to the emergency room, where he received multiple treatments for cardiac arrest. The patient subsequently died as a result of the cardiac arrest. The third death occurred at about 24 months postoperative when the patient went to the emergency room complaining of right arm pain, chest pain,

and shortness of breath. The patient was kept overnight and discharged the next day. The patient was instructed to take an aspirin a day. A few days later, the patient had a fatal cardiac arrest.

2. Radiologist Findings

The radiographs from the clinical study were evaluated by independent reviewers. These individual reviewers were asked to report any observations of bent, broken, or fractured implants, as well as implant migration. A summary of their findings is presented in **Table 12**.

In the investigational group, the radiographic reviewers did not note any findings of implant bending, breakage, migration, or fracture. In the control group, implant migration was noted in three patients (██████████).⁶ None of these resulted in an apparent second surgery procedure associated with a migration. Patient ██████ did have a procedure classified as “other” to evacuate a hematoma following the study surgery.

Also in the control group, five patients (██████████) were reported to have broken or fractured bone grafts. Three of these patients did have their implants removed or replaced in subsequent revision (one) and removal (two) procedures. The revision procedure was due to the need to fuse an adjacent level. One of the removals was elective, while the other removal occurred in a subsequent procedure related to an infection.

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

⁶ Please note that this list of patients 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.

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. Other surgical procedures are ones that do not fit into the previously mentioned categories and may not even involve the cervical spine.

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 were no reported revision procedures in the investigational group. There were five occurrences (1.9%) in the control group. The investigational group rate was found to be statistically lower to that of the control group (0.0% vs. 1.9%). For control patients, four of the revision procedures involved an adjacent level fusion. The remaining revision procedure occurred shortly after the original procedure to remove residual disc material.

Likewise, the investigational device group had a statistically lower rate of supplemental fixation procedures when compared to the control group (0.0% vs. 3.0%). Eight control patients had nine procedures. Seven of these procedures were due to suspected non-unions arising from the original procedures. Non-unions were not a consideration for investigational patients, because they did not receive a fusion procedure. The other two procedures were performed in response to pain and neurological symptoms. Six of the nine 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 lower than that for the control group (1.8% vs. 3.4%); however, it was not statistically different. The removals in the investigational group were primarily due to the treatment of symptoms such as pain and neurological complaints. Fusion procedures

followed these removals. Seven of the nine control implant removals were non-elective, while two were elective removals. The non-elective removals were often associated with additional fusion procedures in which different implants were used. One non-elective removal in the control group occurred shortly after the original procedure in the treatment of an esophageal abscess.

Five (5) investigational patients had implant removal procedures. Histological and metallurgical analyses for three of these cases have been performed, and final reports are provided in **Attachment J**.⁷ A final metallurgical report and preliminary histological analyses are provided for a fourth investigational patient. No analysis information is available for the fifth patient since the removal was not performed by a study investigator. The histological analyses found tissue responses consistent with those typically seen in proximity to metal-on-metal arthroplasty devices. Slides documenting the histological analyses are available on request. In the metallurgical analyses, most of the implant surfaces showed only superficial wear patterns, and there was no evidence of fracture or damage that would suggest a manufacturing or processing defect.

Investigational patients experienced higher rates of reoperations and surgical procedures classified as “other” (1.4% vs. 0.6% and 21.0% vs. 16.6%, respectively). In neither comparison was the difference in rates found to be statistically different.

If a study patient had a revision, removal, or supplemental fixation procedure, he/she was then classified as a second surgery “failure”.⁸ 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 12 for the control group. One of the “failures” in the investigational group occurred after 24 months postoperative. At or before 12 months, there were two second surgery failures in the investigational group and nine in the control group.

7



⁸ Exceptions to this consideration were possible in control patients, where an adjacent level fusion was not considered to be second surgery failures if the originally treated level was judged to be fused or it was too premature for it to fuse. These patients could have their original plate replaced with a longer one to span the adjacent level.

4. Neurological

The neurological status of the patients participating in the clinical study was assessed preoperatively and postoperatively at every follow-up visit. The neurological status questionnaire addressed motor and sensory function as well as reflexes. Investigators judged if the patients were “normal” for these categories and, if not, specific measurements (elements) of the abnormal findings were required. Neurological success for each of the three indicators was based on maintenance or improvement of condition postoperatively as compared to the preoperative status for each element. Overall neurological status success was based on demonstrating maintenance or improvement in all three indicators.

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 the investigational group. The overall neurological success rates for the control group were lower than those of the investigational group. In fact, the success rates at 24 months postoperative were 93.8% and 86.8% for the investigational and control groups, respectively. Differences favoring the investigational group were noted in the sensory, motor, and reflex categories at 24 months.

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% and also a posterior probability of superiority of 97.1%. These results indicate that the overall neurological success rate for the investigational group was not only non-inferior, but also statistically superior, to the rate for 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 serious adverse events. Investigational patients had a

lower rate of adverse events that were classified as implant- or implant/surgical procedure-associated. The radiographic reviewers noted no implant migration or fractured/broken implants in the investigational group, while some did exist for control patients. Of particular note, the investigational group had statistically lower rates of second surgical procedures related to revisions and supplemental fixations. The rate of removals was also lower, but not statistically different. These findings resulted in a lower second surgery failure rate for investigational patients. The investigational group's neurological success rate was statistically higher than 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)⁹ 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 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.

⁹ Vernon, H. and Mior, S. "The Neck Disability Index: A study of reliability and validity," *J. Manipulative Physiol. Ther.* 1991; 14(7): 409-415.

The mean NDI scores for the investigational and control patients for the different clinical study periods are provided in **Table 15** and 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 34.8 and 35.2, respectively. These values are greater than the mean improvement scores of 32.8 and 33.6 for the control group.

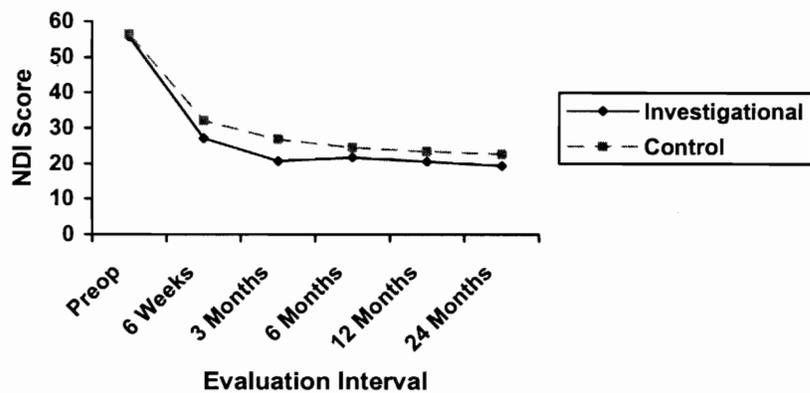


Table 16 shows the distributions of patients demonstrating successful NDI outcomes. The NDI success criterion is a function of the preoperative NDI score. A 15-point or greater NDI score improvement following surgery was required to be deemed a successful outcome. The table indicates that the investigational group rates at all postoperative periods were greater than the corresponding control group rates. Bayesian statistical analyses showed that the posterior probability of non-inferiority of investigational group to the control is 98.5%, thus demonstrating statistical non-inferiority.

2. Neck Pain

Numerical rating scales¹⁰ 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 multiplying the numeral rating scores

¹⁰ McDowell, I. and Newell, C. *Measuring Health: A Guide to Rating Scales and Questionnaires*, Oxford University Press, New York, 1996.

from the intensity and duration scales. 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 were similar for the two treatment groups.

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 neck pain success rates of 94.7% and 93.8%, respectively. The control group rates were 95.5% and 99.2%, respectively.

The 24-month control group success rate is numerically higher than the investigational rate. The Bayesian statistical analyses showed that the posterior probability of non-inferiority of the investigational device to the control at 24 months is 99.2%, i.e., statistically non-inferior.

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.

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 89.4%, as compared to a 92.3% rate for the control group. At 24 months, the rates were 90.6% and 94.2%, respectively.

Bayesian statistical analyses showed that the posterior probability of non-inferiority of the investigational group to the control is 98.1%, thus demonstrating non-inferiority.

4. General Health

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 6-, 12-, and 24-month 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 health summary (PCS) is based primarily on the physical functioning, role-physical, bodily pain, and general health scales of the SF-36 survey. The mental health 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 (12.8 and 12.9) compared very favorably to those values for the control group (11.2 and 11.4, respectively). The mean improvements in MCS scores from preoperative to 12 and 24 months postoperative for the investigational patients (7.7 and 7.1) were also comparable to those values for the control group (6.1 and 8.5).

Table 22 presents the proportions of patients who demonstrated maintenance or improvement in SF-36 results postoperatively as compared to the preoperative condition. For the PCS results at 12 and 24 months, the success rates for the two treatment groups exceeded 85% and were very similar. The MCS results were intriguing since the

investigational group rate at 12 months exceeded that of the control group by nearly eight percentage points. However, at 24 months, the success rates reversed with the control group rate being almost eight percentage points higher. There is no obvious reason for this MCS reversal, and it was not demonstrated in the PCS group.

Bayesian analyses were performed comparing both the 24-month PCS and MCS results of the investigational group to the control group. For the PCS results, the posterior probability of non-inferiority was found to be 97.9%. The posterior probability of non-inferiority comparing the MCS success results of the investigational group to the control group was 87.5%. Therefore, statistical non-inferiority was demonstrated for the PCS comparison, but not for MCS. The MCS finding is of limited importance considering the nature of the variable as a specific indicator of the effectiveness of the investigational treatment. Further, the mean MCS improvement scores (preoperative vs. 24 months) were 7.1 and 8.5 points for the investigational and control groups, respectively. This small difference of 1.4 points was not statistically different ($p=0.480$, t-test). Finally, in the 24-month Bayesian statistical analysis involving “all currently available data”, which will be presented later in this report, non-inferiority was established for the investigational group at 24 months.

5. Global Perceived Effect

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, 81.0% and 85.1%, respectively, of the investigational patients indicated that they had either “completely recovered” or were “much improved”. These rates were higher than the 74.9% and 81.0% rates, respectively, for the control group.

6. Doctor’s Perception of Results

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, 90.9% of the doctors responded

that investigational patients were in “excellent” or “good” condition. This rate is higher than the 87.5% value for the control group. At 24 months postoperative, 94.5% of the investigational device and 91.7% 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.¹¹ Please refer to **Attachment L** for information pertaining to the reviewers and their training. Financial disclosure forms for these individuals are provided in **Section IV.B** along with those for the clinical investigators.

1. Functional Spinal Unit (FSU)

Measurements pertaining to the functional spinal unit (FSU) height were made to evaluate whether the disc height had been maintained during the postoperative course. Using FSU height was a surrogate for evaluating the maintenance of disc height or directly determining whether the implant had subsided. Measuring the actual disc height using vertebral endplate distances is particularly difficult in investigational patients since the implant can obscure measurement landmarks. In control patients, formation of a solid fusion can also obscure the measurement landmarks.

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 can determine if the FSU

¹¹



height had changed. A notable decrease in FSU height over time is considered indicative of a decrease in disc space height.

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 6-week postoperative measurement. FSU height measurements were performed by two teams of radiographic reviewers. If their determinations of FSU height success differed, a third reviewer was used to break the tie.

Despite using FSU height methodology as an indicator of disc space height maintenance, measurements were still encumbered by the inability to visualize the area of interest or a poor-quality film. This was especially true for patients having procedures at C6-C7, where the shoulders can obscure the area of interest. The issue is compounded since both 6-week and 24-month measurements are needed to obtain FSU success. If either is missing, the FSU success result will be missing. For the interim analysis cohort, success/failure determinations could be made for approximately 75% of the patients, and approximately two-thirds of the missing FSU results were in patients having C6-C7 procedures. The missing FSU success results were spread fairly evenly between the investigational and control groups.

The rates of FSU height maintenance or improvement at 3, 6, 12 and 24 months following surgery are presented in **Table 25**. The FSU height maintenance or improvement rates were high, exceeding 95%, for both treatment groups at the four postoperative time periods. Bayesian analyses comparing the overall investigational FSU height 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**).

2. Treated Level Measurements

a. Investigational Group

Angular motion was measured at each study period by comparing lateral flexion and extension radiographs, and the results are given in **Table 26**. Two independent radiologists made the

measurements. The mean angular motion value prior to surgery was 7.55°. This level of motion was maintained following the implantation of the investigational device. At 12 and 24 months postoperative, the mean angular motion values were 7.59° and 7.87°, 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 27**) 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 0.40 mm.

Radiographic success for the investigational group was based on 1) the existence of flexion/extension angular motion in a range of $>4^\circ$ to $\leq 20^\circ$, and 2) no evidence of bridging trabecular bone forming a continuous connection between vertebral bodies. If the two primary radiographic reviewers yielded conflicting success outcomes for a patient, a third reviewer was used for adjudication.

Table 28 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 69.9% and 72.6%, respectively. The primary contributor to these success rates was the angular motion component since bridging bone was not observed in many patients – only one at 24 months. The angular motion component yielded success rates between 70 and 77% at the 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 7 to 8 degrees. 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 6.36° to 6.80°. These results are shown in **Table 29**.

b. 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 ($\leq 4^\circ$) and no radiolucent lines covering more than 50% of the implant surface.¹² Fusion observations were performed by two review teams of radiographic reviewers. If their determinations of fusion status differed, a third reviewer was used to break the tie.

Table 30 presents the fusion rates for the patients in the control group at the various postoperative intervals.¹³ At 12 and 24 months following surgery, the success rates were very high at 98.7% and 98.8%, respectively.

3. Adjacent Level Measurements

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 6 weeks.¹²

Angular Motion

Table 31 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 11.17° and 10.77°, respectively. At 12 months, the angular motion values had increased slightly to 11.94° and 12.07°, respectively. The mean 24-month angular motion value for the level above in investigational device patients was 12.05°, as compared to 11.63° 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 8.32° and 7.77°, respectively. The motion values remained fairly constant, with perhaps only a slight 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 9.47° and 9.07° 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, both the mean values for the level above and below were higher than the angular motion value for the treated segment at 24 months (12.05° vs. 9.47° vs. 7.87°, respectively).

Translational Motion

Results for translational motion at the adjacent levels, which was measured at all of the study periods, are given in **Table 32**. The mean values for both treatment groups were very similar and remained fairly constant over time. For the level above the treated segment, the mean values from preoperative through 24 months following surgery were in a range of 1.18 mm to 1.47 mm for the investigational and control groups.

The translational motion values for the level below the treated segment were consistently lower than the level above the treated level by approximately 0.50 mm. The mean values from preoperative through 24 months following surgery were in a range from 0.76 mm to 0.96 mm for the investigational and control groups.

For investigational patients at 24 months postoperative, both the level above and below had higher translational motion values than the 0.28 mm treated segment value.

We believe that these results indicate that the levels adjacent to the treated cervical segment were stable with regard to translation movement over the postoperative course, and the motion levels were similar to those before treatment.

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 cannot 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 33** provides this information for the two treatment groups at 3, 6, 12, and 24 months following surgery.

At 24 months following surgery, the overall success rate for the investigational group was 80.5%, as compared to a 71.3% rate for the control group. Similarly, an approximately 11 percentage point difference was also 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 95.9%.

Overall success rates were also calculated using the same criteria mentioned above with the addition of FSU (disc height) success, although Medtronic Sofamor Danek does not believe that it is meaningful to include FSU height as a component of overall success. Over the course of seeking the FDA approval of the IDE, Medtronic Sofamor Danek failed to persuade FDA not to consider FSU height success as one of the components for overall success,

the primary endpoint of the study. There is a precedent for disc height not being a component of overall success in at least one other IDE study of an artificial disc replacement device, and disc height was not a primary measurement for a lumbar disc device which has already received PMA approval. Therefore, there is certainly not a consensus that disc height is a primary descriptor of the safety and effectiveness of a spinal motion device. Some further rationales for excluding FSU success from the overall success determination are provided as follows.

There is an inherent difficulty with interpreting films in the cervical spine, especially the lower portion where shoulder obscuration can occur. Therefore, this measurement can be prone to yield missing data. This argument is further supported by the fact that, as mentioned earlier, approximately two-thirds of the missing FSU height data in this study were in patients having C6-C7 procedures. The issue is compounded since both 6-week (baseline) and 24-month radiographic measurements are needed to determine FSU success at 24 months. If either is missing, then the FSU success result will be missing. In addition, radiographs have historically been some of the most difficult data to obtain from investigational sites since the radiographs may be taken at hospitals and radiographic centers away from the investigator's office. There can also be retrieval and copying issues. For these reasons, FSU success status could be determined only in approximately 75% of the patients in the interim analysis cohort who had NDI or neurological status data (the other two overall success components). The missing FSU success results were spread fairly evenly between the two treatment groups.

In patients whose FSU success results were missing, their overall success outcomes were also missing when FSU success was considered, even if all of the other components, which were more important safety and effectiveness measurements, showed positive outcomes. Therefore, the safety and effectiveness determinations arising from such an overall success definition do not reflect their outcomes.

In this study, FSU success rates were very high in both treatment groups, exceeding 95%. A few failures in both groups are possibly due to measurement variation in combination with the stringent success criterion used. Supporting this argument is the fact that the investigational group had slightly higher success rates at both 12- and 24-month evaluations than the control group, in which the treatment is an instrumented fusion and any subsidence (FSU failure) is unusual. Thus, essentially every patient in both groups has a successful FSU status when data were available. Adding

FSU to overall success does not provide meaningful information, but it only reduces important safety and effectiveness information revealed by the other components because of the high percentage of missing values in FSU.

Clearly, it is not a good scientific approach to include FSU success as one of the overall success components for assessing safety and effectiveness of the devices.

Nevertheless, overall success rates which have FSU success as a component are also presented in **Table 33**. At 24 months following surgery, the overall success rate for the investigational group was nearly 17 percentage points higher than the control group (81.1% vs. 64.4%).

Despite the smaller sample size due to missing FSU success values, Bayesian statistical analyses of these modified overall success rates yielded a posterior probability of non-inferiority at 24 months of essentially 100%. The posterior probability of superiority was found to be 99.7%.

In summary, 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, regardless of the definition used. Therefore, the clinical study objective was met, thus indicating that the PRESTIGE® 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.¹⁴ 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 34**. Less than 80% of the patients had "normal" gait scores preoperatively for both treatment groups. These values climbed considerably postoperatively, with 99.2% of the investigational patients and 98.3% of the control patients having "normal" values at 24 months following surgery. The

¹⁴ Nurick, S. The pathogenesis of the spinal cord disorder associated with cervical spondylosis. *Brain* 1972; 95: 87-100.

success rate results, which were similarly high for both treatment groups, are given in **Table 35**.

2. Foraminal Compression Test

The foraminal compression test was performed on patients at every study period. The test is performed by applying a force to the top of the head while the patient laterally flexes his/her head. If the patient feels pain in the upper extremities, it is likely due to nerve root compression and is considered a “positive” result. The desirable outcome is “negative” – an absence of any sensation.

The foraminal compression test results are presented in **Table 34**. Preoperatively, over 50% of the patients in both treatment groups had “positive” responses. At 24 months, the rates of “negative” outcomes for both treatment groups were virtually identical and in excess of 95%.

3. Work Status

Table 36 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 36**, it is evident that the work status of the investigational patients was no worse than that of the control patients.

Preoperatively, approximately 66% of the investigational patients were working, as compared to a 63% rate for control patients. At 6 weeks through 6 months following surgery, the difference in working rates increased, favoring investigational patients. At 12 months, the gap returned to the preoperative level of approximately three percentage points. However, at 24 months following surgery, the percent of working patients in the investigational group (78.1%) bettered that of the control group (71.9%) by over six points.

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 **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 45 days, as compared to 61

days for control patients. This 16-day difference approached statistical significance (Log-Rank Test $p=0.094$, Wilcoxon Test $p=0.022$).

4. 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 37**. At 12 and 24 months following surgery, the results were fairly similar for both the investigational and control groups, and the 24-month results were at least as good as, if not better than, the 12-month postoperative results for both treatment groups. At 24 months postoperative for the first question, 89.0% of the investigational patients and 90.1% of the control patients responded either “definitely true” or “mostly true”. For the second question, 85.0% of the investigational patients and 85.1% of the control patients thought that they were helped as much as expected from their surgeries. Finally, 87.4% of the investigational patients said that they would have the surgery again, as opposed to an 84.3% rate for the control group.

In summary, the rates were not dramatically different, and the investigational patients appear to be at least as satisfied with their procedures as the control group patients.

5. Medication Summaries

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

6. Metal Ion Testing

Metal ion testing was not part of the original IDE study protocol. In addition, throughout most of the course of patient enrollment in the study, FDA never issued an advisory on the need for this data for this product. To our best recollection, the first mention of a potential regulatory need for ion data arose in mid-2004 in an IDE regulatory submission for a different spinal motion device.

Subsequently, both the company and FDA agreed that the collection of this information in a limited number of patients would provide some useful information. Notwithstanding, we do not believe metal ions present a safety concern for this product due to the long-standing use of stainless steel in orthopedic implants, the favorable preclinical animal injection study results,¹⁵ and the lack of definitive information that links metal ions to a clinical concern. Therefore, metal ion level studies are currently being conducted in a subset of patients enrolled in the Continued Access arm of the Artificial Cervical Disc pivotal IDE clinical trial.¹⁶ According to the protocol, patients are required to provide blood samples preoperatively and at 3, 6, 12, and 24 months following surgery. The blood samples collected as part of this study will be analyzed at Rush University Medical Center, Chicago, Illinois, for the presence of chromium and nickel ions. Testing will be done using analytical chemistry instruments, and metal ion quantities at each postoperative evaluation will be compared to the preoperative measurement.

Patients participating in the metal ion study must meet both the inclusion and exclusion criteria for the main Continued Access study as well as specific criteria for the ion study. These criteria specific to the ion study exclude patients who have metal implants, are taking certain medications or nutritional supplements, or experience occupational exposure to metal particles.

The metal ion study was approved for up to 25 patients, and all of them have been enrolled and undergone surgery. Any available information concerning these patients is presented with the Continued Access data in **Section IV.C**. Ion results will be provided to FDA in the future.

15

16

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, FSU, and overall success results across the sites. There were no statistically significant differences noted in any of the comparisons, thus indicating that the results were consistent among different sites. These outcomes provide confidence in 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. Patients who were excluded from the “per protocol” analysis had major protocol deviations, i.e., did not meet the inclusion/exclusion criteria or received the wrong study treatment, or other major protocol deviations that could potentially affect clinical outcomes.

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

“Per Protocol” Success Rates			
	Investigational	Control	Post. Prob. of Non-inferiority
NDI	82.5%	83.0%	97.1%
Neurological	93.7%	86.6%	100.0%
FSU Height	97.8%	95.1%	100.0%
Overall Success (without FSU)	80.2%	72.6%	100.0%
Overall Success (with FSU)	81.7%	65.5%	100.0%

Like the previous analyses, every statistical comparison for the “per protocol” dataset yielded a posterior probability of non-inferiority of at least 95%. Further, the investigational group was found to have statistically superior outcomes, as compared to the control group for neurological status and overall success with FSU as one of the criteria. The investigational group overall success rate without FSU in the definition was nearly eight percentage points higher than the

control group rate, which approached statistical superiority (probability of superiority 92.4%).

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 overall success outcomes were still non-inferior to those of the control group and superiority was approached.

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. The overall success rates for the investigational group were found to be non-inferior to those of the control group and superior if FSU is considered in the definition.

3. “Missing Equals Failure” Results

The “missing-equals-failure” data presentations for various study periods are included in **Attachment Q**. For this presentation, secondary surgery failures, 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 (without FSU) for the investigational group was higher than that of the control group (75.2% vs. 58.8%).

The same is also true for the investigational group with FSU in the definition (56.2% vs. 39.2%).

4. Sensitivity Analysis for Assessing Missing Values

As mentioned earlier in this report, there was a disparity in follow-up rates at 24 months between the investigational and control group. In the interim analysis cohort, nine (6.6%) of 137 investigational patients did not have overall success outcomes, as compared to 26 (17.6%) of 148 control patients. To assess the impact of lost-to-follow-ups 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 ($p=0.0411$). When 50% of missing investigational outcomes and 60% of the missing control outcomes are assumed to be successes (which favors the control group and could perhaps be closer to the actual situation), the superiority of the investigational treatment to the control can still be claimed ($p=0.0363$). 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	88.1%	87.0%	88.7%	86.9%
Neurological	92.9%	88.8%	92.7%	88.0%
FSU Height	97.8%	97.5%	98.9%	98.7%
Overall Success (without FSU)	84.9%	83.6%	85.5%	83.3%
Overall Success (with FSU)	84.4%	80.7%	86.4%	80.8%

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**. At 6 weeks following surgery, there was no significant correlation between these measurements. However, at later postoperative intervals, statistically significant correlations were noted. At 12 and 24 months following surgery, the three indicators of pain, i.e., NDI, neck pain, and arm pain, were found to be negatively correlated to angular motion, and in all comparisons, the correlations were statistically significant, although the magnitude of the correlations was very moderate. This means that lower, or better, pain scores are associated with higher angular motion values. This finding is intuitive since one would expect that more motion would create less stress at the treated level, which would, in turn, generate less pain.

7. Financial Disclosure Information and Analyses

Financial disclosure information pertaining to the investigators and co-investigators who participated in IDE G010188 is provided in **Attachment U**. The information indicates that 20 of 72 (28%) 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 187 patients to both treatment groups. At 12 and 24 months postoperative, there were no statistically significant differences in any of the outcome comparisons between the patients of surgeons with a financial interest versus those without.

Therefore, it is apparent that the existence of an investigator financial interest, as defined in 21 CFR 54, did not impact the results or 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 250 patients (ordered by the surgery date) who had valid overall success outcomes

(without FSU) at 24 months.¹⁷ 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**.

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	82.8%	81.8%	83.0%	80.3%
Neurological	93.8%	86.8%	92.8%	84.3%
FSU Height	96.8%	95.5%	96.5%	93.8%
Overall Success (without FSU)	80.5%	71.3%	79.3%	67.8%
Overall Success (with FSU)	81.1%	64.4%	80.4%	63.0%

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 and superior to the control group for neurological success and overall success (both definitions). “Per protocol” analyses yielded similar conclusions.

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 safety information previously presented in this report is for “All Currently Available Data”.

9. Data Listings

Data listings for the PRESTIGE® device and control patients are provided in **Attachment W**.

V. Conclusions

The goal of the PRESTIGE® Cervical Disc System IDE clinical study (G010188) 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

¹⁷ 

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 PRESTIGE® Cervical Disc, compared very favorably to the control group results.

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 virtually identical to the control group rate. Even though not statistically significant, the number of patients having serious adverse events in the investigational group was less than that found in the control group. The radiographic reviewers reported no incidences of broken or migrated implants in the investigational group, while some such observations were made in control patients.

The investigational group had a statistically lower rate of secondary surgical procedures related to implant revisions and supplemental fixations. Investigational patients also experienced a lower rate of implant removals, but it was not statistically different. These findings resulted in a lower second surgery failure rate for investigational patients.

Maintenance or improvement in neurological status was found in greater than 90% of patients in the investigational group. Furthermore, the 24-month overall neurological success rate of 93.8% for the investigational treatment group was found to be statistically superior to the rate of 86.8% seen in 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 PRESTIGE® Cervical Disc.

In terms of effectiveness measures, NDI scores improved dramatically after surgery for both treatment groups, exceeding 30 points at 12 and 24 months. The mean improvements in NDI scores for the investigational group at these periods were nearly two points higher than for the control group. 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 non-inferiority 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 very similar both before and after 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 (disc) height was found to be a difficult measure to obtain in both treatment groups, due, in part, to anatomical obscuration. Nevertheless, the available measurements showed that 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 45 days, which was 16 days shorter than the time for control patients.

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 overall success rate for the investigational group was 80.5% and approximately nine percentage points higher than the 71.3% rate for the control group. A similar difference was also seen at 12 months. Overall success rates were also calculated adding FSU (disc height) success to the formula. Again at 12 and 24 months following surgery, the investigational group overall success rate was impressively higher than the control group rate, with the 24-month rate being nearly 17 percentage points higher. Regardless of the definition used, 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 PRESTIGE® 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 PRESTIGE® Cervical Disc System and should lead to the approval of the device.

<<Tables 1-37 redacted>>