

July 28, 2004

Mathematical Statistician
Division of Biostatistics, OSB

Final Statistical Review of PMA P040001, St. Francis Medical's X STOP® Interspinous Process
Distraction System (January 7, 2004)

Objective

The objective of the clinical study were to evaluate the safety and effectiveness of the X STOP Interspinous Process Distraction System in the treatment of neurogenic intermittent claudication secondary to mild or moderate lumbar spinal stenosis.

Study Design

A prospective, randomized, controlled, multi-center clinical study was conducted. 191 patients (100 X STOP and 91 controls) were enrolled and treated at 9 investigational sites. The X STOP was compared to non-operative therapy in the treatment of neurogenic; intermittent claudication secondary to mild or moderate lumbar spinal stenosis. Patients randomized to the control group received at least one epidural steroid injection, nonsteroidal anti-inflammatory (NSAIDs), analgesics, and physical therapy. The study was designed to prove that the X STOP is significantly more efficacious than the control utilizing the Zurich Claudication Questionnaire (ZCQ) as the primary outcomes measurement. Safety information was measured by an analysis of reported adverse events and additional surgeries.

Sample Size

The definition of study success is that 'the X STOP group has a higher success rate than the control group. The sponsor anticipated success rates for the X STOP and control of 60% and 37.5%, respectively. They found that for a 2-sided test of proportions with a 0.05 significance level and 80% power that 85 subjects per group were needed. In order to allow for a drop-out rate of 15%, 200 patients were planned to be enrolled in the study. *Note that the sponsor ultimately enrolled 229 patients, even though their IDE was for 200 subjects. Note that they only treated 191 subjects (100 X STOP and 91 control).*

Randomization

Block randomization of the patients, stratified by site, was used with block sizes of 2. **Note that with a fixed block size of 2 (1 treatment and 1 control patient for each pair of randomized patients) used by the company, the randomization assignment of patients could easily be subverted by the investigators.** That is they could always know the treatment assignment of the 2nd of the pair if so inclined, and thus, potentially pick the patient to receive that second assignment. The sponsor used a call-center to assign the treatment when each site called in the patient's information to verify eligibility. The date of surgery was considered as the treatment date for the X STOP patients and the date of the initial epidural injection was considered as the treatment date for the control patients. Masking of patients was not possible.

Enrollment

The sponsor enrolled 229 patients at 9 centers, with the highest-enrolling center enrolling 45 and treating 36 patients and the smallest enrolling 9 and treating 6 (page 44). Of these 229 patients, 191 were treated (100 X STOP and 91 control). 14 X STOP and 24 control patients withdrew or were excluded before treatment. Reasons were related to health (3 X STOP/2 Control), failure to meet inclusion/exclusion criteria (1/1), and voluntary withdrawal (16/26). Eight subjects withdrew after randomization to the control group. Details are provided below.

Table 1: Patients Enrolled but not Treated

Reason for Withdrawal/Exclusion	X STOP	Control
Health related		
Cancer	1	1
Coronary artery bypass surgery	0	1
Hand infection with fever	1	0
Renal complications w/nephrectomy	1	0
Failure to meet inclusion/exclusion criteria		
Stenosis at > 2 levels	1	0
Age < 50	0	1
Patient unable to schedule surgery/treatment		
Work schedule	1	0
Family illness	1	0
Delayed scheduling treatment until study closed	0	2
Patient voluntarily withdrew from study		
Elected to undergo laminectomy	1	4
Elected to forego X STOP implantation	7	0
Elected to forego epidural injection	0	4
Elected to withdraw after randomization to control group	0	8
Elected to withdraw after loss of spouse	0	1
Patient refused to make/keep appointments	0	2
TOTAL	14	24

The sponsor performed a series of sensitivity analyses in order to assess the effect that the exclusion of untreated patients might have had on estimates of the study success rate. Analyses were performed based on a) the assumption that all untreated patients would have been failures, b) the assumption that all untreated patients would have been successes, c) the use of available data from untreated patients (24 month ZCQ scores were obtained from untreated patients wherever possible), and d) multiple imputation methods. As shown in **Appendix F**, these analyses did not materially affect comparison to the control. Success rates were always statistically significantly higher in the X STOP group than in the control group.

Demographic Data

A total of 191 patients were enrolled in the X STOP clinical trial: 100 patients were treated with the X STOP and 91 received non-Operative care. Demographic information pertaining to the patients participating in the study is presented in Table VI. Note that there was very little difference between the 2 groups with respect to baseline characteristics.

	X STOP Device	Control	p-value
Age (yr.) Mean (Range)	70.0 [50-94]	69.1 [50-88]	0.513
Weight (lbs.) Mean (Range)	177.1 [105-265]	180.2 [98-293]	0.569
Height (in.) Mean (Range)	67.3 [56-74]	66.3 [56-75]	0.117
Gender: Male	57 (57.0%)	46 (50.5%)	0.387
Female	43(43.0%)	45 (49.5%)	
Spondylolisthesis Present	35 (35.0%)	24 (26.7%)	0.272

Patient Accountability

Patient accountability data are summarized below in Tables 5 and 6 of the Clinical Report (page 47 of Volume I of the PMA). Of the 100 treated patients for the X STOP group, 88 had 2-year follow-ups. Of the 12 patients who didn't, 6 had additional LSS surgery, 4 died, 1 had an implant removal, and 1 voluntarily withdrew. Of the 91 treated control patients, only 58 had 2-year evaluations. Of the remaining 33 patients, 24 had LSS surgery, 3 died, 1 had the procedure aborted, and 1 voluntarily withdrew.

Table 2: Patient Accountability by Follow-up Visit - X STOP Patient Group

	Treated	6 wk	6 mo	12 mo	24 mo
Evaluable Patients (Cumulative)					
?? Theoretically due	100	100	100	100	100
?? Additional LSS surgery with device removal (reached defined failure endpoint < 24 months)			2	4	6
Discontinued Patients (Cumulative)					
?? Deaths		1	2	2	4
?? Implant removals		1	1	1	1
?? Lost to follow-up		-	-	-	-
?? Voluntary withdrawal		-	1	1	1
Expected Patients*		98	95	93	89
Missed visits**		4	6	4	0
Patients Evaluated		94	88	88	88
Percent Follow-up		95.9%	92.6%	94.6%	98.9%

* Includes patients who voluntarily withdrew from the study in the total # of "expected" patients.

24 month follow-up visits outside of the protocol-defined window are treated as protocol deviations and are presented in **Attachment 5 (Listing 1.5).

Table 3: Patient Accountability by Follow-up Visit - Control Patient Group

	Treated	6 wk	6 mo	12 mo	24 mo
Evaluable Patients (Cumulative)					
?? Theoretically due	91	91	91	91	91
?? LSS Surgery (reached defined failure endpoint < 24 months)*		1	13*	17	24
Discontinued Patients (Cumulative)					
Deaths		-	1	2	3
Procedure Aborted	1	1	1	1	1
Lost to follow-up		-	-	-	-
Voluntary withdrawal		2	2	2	5
Expected Patients**		89	77	72	63
Missed visits***		17	11	1	0
Patients Evaluated		70	64	69	58
Percent Follow-up		78.7%	83.1%	95.8%	92.1%

*One control patient (#817) in this group died secondary to complications of Parkinson's disease 20 months following laminectomy

** Patients who voluntarily withdrew from the study are included in the total # of "expected" patients.

***24 month follow-up visits outside of the protocol-defined window are treated as protocol deviations and are presented in **Attachment 5 (Listing 1.5)**.

Data Analysis and Results

The primary efficacy variable was analyzed using an Intent-to-Treat (ITT) methodology where the ITT population was defined as all patients receiving treatment or attempted treatment (i.e., treatment was initiated but aborted) in the X STOP or control group. Patients who were enrolled but not treated (i.e., declined to participate or were unable to participate) were not included.

For all primary analyses, missing data were not imputed. Patients reaching any endpoint before completing the study that was defined in the protocol as a treatment failure (e.g., a laminectomy) were included as failures in the computation of the study success rate, however.

Secondary efficacy variables, safety variables and demographic and baseline variables were similarly analyzed except that ITT analysis was not used. Categorical variables were analyzed using the Fisher exact test. Continuous variables were analyzed using analysis of variance (ANOVA). Two-sided p-values were calculated and considered to be statistically significant when $p < 0.05$.

Effectiveness Analysis

As described above, the primary Efficacy variable was ZCQ success at the 24 month follow-up. Additionally, for treatment with the X STOP to be considered successful, distraction had to be maintained at 24 months, and there could be no device-related complications or re-operation.

Table VIII below summarizes the clinical results at the 24 month follow-up. The X STOP group had statistically significantly improved ZCQ scores over baseline compared to the control group, and a significantly greater number of X STOP patients were satisfied with their treatment. Overall treatment success was 45.7% (42/92) for the X STOP group compared to 4.9% (4/81) for the control group ($p < 0.001$).

	X STOP		Control		p-value
	n/N	%	n/N	%	
ZCQ Success Rate by Domain					
Physical Function	53/93	57.0%	12/81	14.8%	<0.001
Symptom Severity	56/93	60.2%	15/81	18.5%	<0.001
Satisfaction	68/93	73.1%	28/78	35.9%	<0.001
Percent of Patients Meeting All 3 ZCQ Success Criteria					
ZCQ Success	45/93	48.4%	4/81	4.9%	<0.001

From the table above for the overall success rate, 8 of the 100 treated X STOP and 10 of the 91 Control patients have been removed. As presented by the sponsor and shown by the results in Table 18.5 (Amendment 1), 2-year results for 10 (4 X STOP and 6 Controls) of these patients with results for over 1-year can be reasonably imputed. As stated by the sponsor “Variability in success rate estimates can be substantially reduced if some reasonable imputations are performed. There are 10 “missing” patients for whom the two-year outcome can reasonably be inferred from available data. These are listed below in **Table 18.5.**”

Table 18.5: Imputed Results for 10 Patients with Missing 24 Month Data

Group	Patient	Last Visit Day	Result (Imputed)
X STOP	0346*	384	Failed on all ZCQ domains
X STOP	0718	559	Failed on ZCQ severity
X STOP	1026	710	Success on all ZCQ domains
X STOP	1403	382	Failed on ZCQ physical function
Control	0124	343	Failed on ZCQ severity and function
Control	0125	427	Failed on all ZCQ domains
Control	0501	581	Failed on all ZCQ domains
Control	0715	559	Failed on ZCQ severity
Control	0806*	389	Failed on ZCQ severity and function
Control	1210	355	Failed on all ZCQ domains

* Died prior to two-year visit.

“Patient #1026 would almost certainly have been a success. That patient showed success on all ZCQ domains at two years and was only treated as missing because of the absence of two-year radiographs which prevented the evaluation of maintained distraction. However, one-year radiographs did indicate that distraction had been maintained.”

“The remaining patients were all failures at one year. Of the 88 patients (in both the X STOP and control groups) who were ZCQ failures at one year and were evaluable at two years, only 4 (4.5%) became successful at two years. Therefore it is reasonable to impute two-year failure for these patients.”

As can be seen from Table 18.7 below, the remaining 8 patients (4 X STOP and 4 controls) who were removed from the database didn’t have 1-year follow-ups. In fact all but one were lost to follow-up by 3 months. Note that 5 of these patients died within the 1st year (3 X STOP and 2 controls).

Table 18.7: Available Results for 8 Patients with Missing Data at the 24 Month Follow-up

Group	Patient	Last Visit Day	Result (Imputed)
X STOP	0120*	73	ZCQ success but incomplete
X STOP	0508*	-	Died of pulmonary edema on Day 2
X STOP	0803	41	Failed on ZCQ physical function
X STOP	0834*	90	Success on all ZCQ domains
Control	0319*	45	Success on all ZCQ domains
Control	0509	-	No data available - refused to return
Control	1014	38	No ZCQ data - worsening pain scores
Control	1017*	192	Failed on ZCQ severity and function

* Died prior to two-year visit.

The sponsor has demonstrated statistical superiority of their device over the control with respect to their primary endpoint. Given the information in Tables 18.5 and 18.7 above the overall success rates can be presented as follows:

Table of Overall Success Rates with Imputed Results on 10 Patients Compared to Lowest Possible & Highest Possible Success Rates

	X STOP	Control	p-value
Intent-to-Treat results *	43/100 (43.0%)	4/91 (4.4%)	< 0.001
Evaluable results**	43/96 (44.8%)	4/87 (4.6%)	< 0.001

* Results for all patients with 1-year data use last observation carried forward (see Table 18.5) and patients with less than 1-year data (Table 18.7) treated as failures.

** Results for patients with less than 1-year data (Table 18.7) excluded.

Success Rate by Site

The sponsor also presented success rates by center on page 72, Table 36 (provided below). The center that treated the most patients had an 85% success rate (17/20) for the X STOP. All other centers had success rates of 50% or less. In fact the next 2 largest centers had X STOP success rates of 28% (5/18) and 14% (2/14). In an attempt to dismiss this discrepancy, they state that the Breslow-Day test for homogeneity was not statistically significant (p=.474) and used this as evidence that there was not a significant difference among sites. **But when the X STOP success rates among centers is tested, it is found to be highly significantly different (p<0.005 for exact Chi-square and Likelihood tests). Most of this difference is attributable to the significantly higher success rate at site 2, the site which treated the most patients. The sponsor addressed this issue in their amendment by claiming that significant site to site differences are common in clinical trials, that the X STOP success rate is greater than the control at all sites, and that the X STOP success rate will still be significantly greater than the control even if this site is removed.**

From Table 36 (page 72): Success Rates by Investigational Site

Investigational Site Number	X STOP		Control	
	n/N	%	n/N	%
1	5/18	27.8%	1/16	6.3%
2	17/20	85.0%	2/15	13.3%
3	2/14	14.3%	0/14	0.0%
4	5/11	45.5%	0/9	0.0%
5	4/8	50.0%	0/9	0.0%
6	4/9	44.4%	0/8	0.0%
7	2/4	50.0%	0/5	0.0%
8	2/6	33.3%	1/5	20.0%
9	1/2	50.0%	0/1	0.0%

Also note on page 191 and 193, it is stated that the principal investigators from site 2 hold a significant equity interest in the Company. The reason for the discrepancy in success rates is unknown. Note that in Section V (the Financial Disclosure section) on pages 191 and 193, the sponsor also describes measures taken to minimize bias into the study at all sites. Investigators didn't administer questionnaires to patients and quantitative measurements were made by physicians not involved in the treatment of patients. An independent radiologist measured the maintenance of distraction in each patient as well as other anatomical measurements that were analyzed as part of the study results.

Safety

Aside from laminectomy (6 X STOP and 24 Control patients), the peri-operative and treatment related adverse events were minimal in each group (Table 49, page 86). The sponsor also lists in Table 50 (page 87) a number of adverse events that they claim are unrelated to the device and claim are not statistically different from the control. *Note that many of the events in the table below occur with greater prevalence in the X STOP group.*

Table 4: Adverse Events Unrelated to Device or Treatment

Adverse Event	X STOP (N = 100)			Control (N = 91)			p-value
	# of Events	# of Patients	%	# of Events	# of Patients	%	
Body as a Whole							
Accidental injury	14	11	11.0%	4	4	4.4%	0.110
Cancer	4	4	4.0%	1	1	1.1%	0.371
Weight gain	1	1	1.0%	0	0	0.0%	1.000
Cardiovascular System							
Cardiovascular disorder	5	5	5.0%	1	1	1.1%	0.214
Peripheral vascular disorder	1	1	1.0%	0	0	0.0%	1.000
Endocrine							
Diabetes	1	1	1.0%	0	0	0.0%	1.000
Gastrointestinal/Genitourinary System							
Gastrointestinal disorder	4	3	3.0%	0	0	0.0%	0.248
Genitourinary disorder/infection	4	4	4.0%	0	0	0.0%	0.123
Hematologic							
Anemia	1	1	1.0%	0	0	0.0%	1.000
Hepatobiliary							
Gallstones	1	1	1.0%	0	0	0.0%	1.000
Immunologic							
Allergy	1	1	1.0%	0	0	0.0%	1.000

Adverse Event	X STOP (N = 100)			Control (N = 91)			p-value
	# of Events	# of Patients	%	# of Events	# of Patients	%	
Musculoskeletal System							
Back, lower	21	16	16.0%	7	7	7.7%	0.118
Back, upper	4	4	4.0%	0	0	0.0%	0.123
Back, unspecified	3	3	3.0%	0	0	0.0%	0.248
Extremity, lower	16	13	13.0%	5	3	3.3%	0.018*
Extremity, upper	4	4	4.0%	2	2	2.2%	0.685
Groin	1	1	1.0%	2	2	2.2%	0.606
Hip	13	11	11.0%	3	3	3.3%	0.052
Rib	1	1	1.0%	0	0	0.0%	1.000
Unspecified	1	1	1.0%	0	0	0.0%	1.000
Nervous System							
Headache	1	1	1.0%	0	0	0.0%	1.000
Neurological disorder	1	1	1.0%	1	1	1.1%	1.000
Neuropathy	4	4	4.0%	0	0	0.0%	0.123
Neuropsychological disorder	5	5	5.0%	1	1	1.1%	0.214
Stroke	1	1	1.0%	1	1	1.1%	1.000
Respiratory System							
Respiratory disorder/infection	3	3	3.0%	1	1	1.1%	0.623

P-values determined using the Fisher exact test

* indicating a level of significance < 0.05

Laminectomy Data

The sponsor provided the following results for a statistical comparison of 36 laminectomy patients with the X STOP patients from their clinical trial. The laminectomy patients come from control (26) and X STOP (6) patients who failed this study, 5 control and 2 X STOP patients from the unwelded study and 7 untreated patients from this study. Of these, outcomes were available for 36 patients.

ZCQ data were collected for patients in the clinical trial who underwent laminectomy surgery. Data were also available for patients undergoing laminectomy from a feasibility study that preceded the Pivotal Clinical Trial. These ZCQ data are summarized in Table X on page 17 of the PMA. Mean follow-up for laminectomy patients was 1.2 years (range: 76 days to 2.71 years). According to the sponsor mean improvement scores in symptom severity and physical function were not significantly different. The mean satisfaction score for patients undergoing laminectomy was significantly higher (indicating less satisfaction) than the mean score in X STOP patient at the 24 month follow-up (p=0.006).

Mean Change Scores	Laminectomy Group (N=316)		X STOP Patients (N=86)		p-value
	N	Mean (SD)	N	Mean (SD)	
Symptom Severity	36	0.87 1.07	86	0.99 0.87	0.532
Physical Function	36	0.74 0.81	86	0.76 0.79	0.898
Patient Satisfaction	36	2.20 1.03	86	1.70 0.85	0.006*

P-value determined using an ANOVA

According to the sponsor, there were no significant differences in the proportion of patients achieving a threshold level of improvement defined as clinically significant in any of the three ZCQ domains (see Table XI, page 18. When all three ZCQ criteria are combined, 44.4% (16/36) of laminectomy patients had a successful outcome compared to 45.7% (42/92) of X STOP patients who met all study criteria for treatment success (p=0.899).

Table XI: Treatment Success					
	Laminectomy		X STOP		p-value
	n/N	%	n/N	%	
Clinically Significant Improvement in Symptom Severity	21/36	58.3 %	56/93	60.2%	0.841
Clinically Significant Improvement in Physical Function	23/36	63.9 %	53/93	57.0%	0.465
Percent of Patients Satisfied	22/36	61. %	68/93	73.1%	0.180
Treatment Success	16/36	44.4%	42/92	45.7%	0.899

Note that there are problems with using this collection of patients who failed the pilot and pivotal study as laminectomy comparison group for the X STOP device. Such a non-randomized comparison could introduce bias into the comparison which cannot be quantified and thus would make the statistical results suspect.

Sponsor’s Conclusions Drawn from the Studies (SSE, pages 12-13)

“Results of the Pivotal Clinical Trial confirm the safety of the X STOP. The implant resulted in a low percentage of complications, which resolved without significant clinical sequelae and no neurological injuries occurred in the study.

The efficacy of the X STOP was demonstrated by outcomes at the 24 month follow-up. A significantly greater proportion of X STOP patients achieved clinically significant improvement in symptom severity and physical function compared to control patients. A significantly greater proportion of X STOP patients were also satisfied with their treatment compared to control patients.

Symptom Severity

At the 24 month follow-up, 60.2% of X STOP patients achieved clinically significant improvement in symptoms compared to 18.5% of patients in the control group.

Physical Function

At the 24 month follow-up, 57.0% of X STOP patients achieved clinically significant improvement in physical function compared to 14.8% in the control group.

Patient Satisfaction

At the 24 month follow-up, 73.1 % of X STOP patients were satisfied with their treatment, compared to 35.9% of control patients.

When patients were required to meet all success criteria as defined in the study protocol, 45.7% of X STOP patient were treatment successes, compared to 4.9% of control patients. X STOP outcomes are comparable to results for laminectomy surgery.” **Note that these rates excludes several patients with missing 24 month data (3 who were failures at 1-year, 1 who was a success on all domains but didn’t have 2-year radiographic results, 3 who died before 1 year, and 1 who was lost to follow-up after day 41).**

According to the sponsor (page 13 of SSE), “The results of this clinical study demonstrate that the X STOP can offer patients an alternative after non-operative therapy and epidural injections are no longer effective and before a decompressive laminectomy is performed. The X STOP provides a conservative, safe, effective treatment option for patients with mild to moderate symptoms of lumbar spinal stenosis.”

Issues of Concern

1. Note that with a fixed block size of 2 (1 treatment and 1 control patient for each pair of randomized patients) as used by the company, the randomization assignment of patients could easily be subverted by the investigators. That is they could always know the treatment assignment of the 2nd of the pair if so inclined, and thus, potentially pick the patient to receive that second assignment.
2. The sponsor presented success rates by center on page 72, Table 36. The center that treated the most patients (site # 2) had an 85% success rate (17/20) for the X STOP. All other centers had success rates of 50% or less. In fact the next 2 largest centers had X STOP success rates of 28% (5/18) and 14% (2/14). In an attempt to dismiss this discrepancy, they state (page 72) that the Breslow-Day test for homogeneity was not statistically significant ($p=.474$) and used this as evidence that there was not a significant difference among sites. But when the X STOP success rates among centers is tested, it is found to be highly significantly different ($p<0.005$ for exact Chi-square and Likelihood tests). Most of this difference is attributable to the significantly higher success rate at site # 2. Furthermore, on page 191 and 193 it is stated that the investigators at this site, who also happen to be the inventors of the device, hold a significant equity interest in the company.
3. In Table 4 above (also Table 50, page 87 of the PMA), the sponsor presents a number of adverse events they claim are unrelated to the device and are not statistically different from the X STOP, yet many of them which could be related to the device (e.g., lower back, lower extremity, hip, etc.) occur much more frequently in the X STOP group.
4. The sponsor provided the results for a statistical comparison of 36 laminectomy patients with the X STOP patients from their clinical trial (see Tables X and XI on pages 16 and 17) and a discussion of the results on pages 101-3 of the PMA. The laminectomy patients come from control (26) and X STOP (6) patients who failed this study, 5 control and 2 X STOP patients from the unwelded study and 7 untreated patients from this study. Of these, outcomes were available for 36 patients. They use the results of the patients and Tables X and XI to support the contention that X STOP is comparable to results from decompressive laminectomy. Note that there are problems with using this collection of patients who failed the pilot and pivotal study as laminectomy comparison group for the X STOP device. Such a non-randomized comparison could introduce bias into the comparison which cannot be quantified and thus would make the statistical results suspect. If the sponsor wants to make the claim that that the X STOP is comparable to laminectomy, they should adequately, if possible, address the bias introduced into the comparison of the results since the laminectomy subjects are comprised of failed patients of 3 studies and the study is not randomized.

Conclusion

The sponsor demonstrated that the X STOP device was superior to the non-operative therapy control with respect to the chosen primary endpoint and the X STOP device appeared to be reasonably safe when compared to the control. Note that the observed success rates of 43% for the X STOP and 4.4% for the control are substantially lower than the expected success rates of 60% and 37.5% the design of the study was Powered (sized) for. Issues of the appropriateness of the control group and the relatively low effectiveness rate for the device should be addressed clinically. It should be noted that the sponsor randomized patients by using block sizes of 2 stratified by site which can allow investigators to potentially subvert the randomization, and thus potentially bias the study results. In addition, we noted that there were discrepancies in the success rates among centers. In Table 50 (page 87), the sponsor presents a number of adverse events they claim are unrelated to the device and are not statistically different from the X STOP, yet many of them occur much more frequently in the X STOP group. The comparison of the X STOP patients to the laminectomy patients is questionable since the study was not randomized and the laminectomy patients comprise patients from 3 different studies who failed their assigned treatment.