FDA Summary Memorandum

This summary memo serves as an overview to the data presented in the PMA and the questions raised during the review process. The areas where the agency has questions related to the submission are identified in this memo. In general terms, these questions are related to the following:

- the manufacturing to the recombinant human protein;
- the definitions of overall success, clinical success and radiographic success;
- the clinical performance, *i.e.*, effectiveness and safety, of the combination product; and
- the immune response to the recombinant protein.

PRODUCT DESCRIPTION SUMMARY

The product under review is OP-1 Putty, a three component combination product containing device and drug components. Specifically, it consists of bovine Type I collagen derived from demineralized bone, a recombinant human bone morphogenetic protein (rhBMP-7, referred to as recombinant human osteogenic protein or OP-1) and carboxymethylcellulose (CMC). The three components are mixed with sterile normal saline to form a graft material with a putty-like consistency. The prepared material is placed alongside the vertebral bodies as a graft for posterolateral fusions.

A single box of the product contains two glass vials:

vial 1 (2oz vial) – 1g of sterile powder consisting of 3.5mg lyophilized OP-1 protein and ~1g collagen

vial 2 (10ml vial) – 230mg CMC.

During surgery, the contents of the two vials are combined with each other then mixed with 2.5ml of sterile normal saline (not provided). Two boxes of the submitted device are required for each patient – one for each graft site. Each batch of the graft material is prepared separately and placed in the typical location for posterolateral fusion grafts.

The recombinant protein is used initiate the cascade of events responsible for bone formation, *e.g.*, recruitment of mesenchymal stem cells and cell proliferation and differentiation into chondroblasts. The bovine collagen is used as a resorbable scaffold for new bone formation. In order for this to occur, the sponsor only includes collagen particles in the range of 75-425µm. It is their belief that particles outside of this range will not allow for bone formation. CMC is used as a thickener to alter the handling characteristics of the OP-1 Implant (collagen + rhBMP-7) and turn it into OP-1 Putty, the combination product.

The product is intended to be used in uninstrumented posterolateral lumbar spinal fusion procedures for the treatment of lumbar spondylolisthesis in patients who have failed at least six months of conservative, non-surgical treatment. Other than the use of OP-1 Putty instead of bone graft, there are no differences between the surgical approach and techniques typically employed for uninstrumented posterolateral lumbar spinal fusion procedures.

NON-CLINICAL STUDIES SUMMARY

Because this product is a combination product that contains biologically-derived components and device components, there are numerous concerns that would not be present for "typical" orthopaedic devices, *e.g.*, immunological response, ability to signal tissue formation, dosage, etc. As a result, the sponsor was required to perform numerous tests to characterize the complete combination product. The majority of these tests were provided as part of the clinical trial submission supporting the PMA or included in the PMA submission itself. Because the majority of these studies provided sufficient information to allow for initiation of the clinical study, they will not be addressed in this memo.

NON-CLINICAL PHARMACOLOGY/TOXICOLOGY STUDIES

OP-1 (BMP-7)

The BMP family of cytokines was first identified as a group of osteoinductive cytokines from the bone matrix. BMP-7 is expressed in various tissues and has an important role in the development of the skeleton, nervous system, eye, kidney, heart and germ cells. A role in tissue repair or regeneration upon injury in kidney and brain has also been described.

Pharmacology

Non-clinical efficacy studies with OP-1 Putty conducted in various animal models (dog, sheep, rabbit, rat) showed that OP-1 Putty was as effective or superior to autograft in bringing about spinal fusion. The results from the animal studies were the basis for the investigational human dose selection (3.5 mg/g). Because the dose response relationship in humans was not established, *i.e.*, no human dosing studies were performed, the optimal human dose is not known. As a result, a question is raised to whether the observed level of effectiveness of the product described below resulted from the selection of an improper dose or some other factor.

Toxicology

Systemic and local toxicity studies with OP-1 were performed in mice and rats and did not raise significant concern. Carcinogenic potential of OP-1 was addressed in *in vivo* and *in vitro* studies. Sarcomas formed in rats upon subcutaneous implantation of OP-1 appear to be probably due to solid state carcinogenesis study as a result of ectopic bone formation at the site of implantation.. Studies on the effect of OP-1 on tumor cell proliferation were inconclusive.

Immunogenicity

In the pivotal clinical trial the immunogenicity of OP-1 was demonstrated by a high incidence (94%) of anti-OP-1 antibodies including antibodies with OP-1 neutralizing activity. Titers decreased over time, but in many subjects they were still significant out to 24 months post-op. Neutralizing antibodies have the potential to interfere with the actions of endogenous BMP-7. This presence over time is of particular concern for women of childbearing potential. The sponsor submitted data from two reproductive toxicity studies in immunized rabbits to address the issue of developmental and pre/postnatal toxicity.

While non-clinical studies specifically designed to evaluate the effect of OP-1 antibodies on organ function or tissue repair were not performed, AEs related to these events were collected. Wound healing abnormalities were not observed in any of the surgical non-clinical studies and AEs associated explicitly with organ function, typical laboratory parameters, *e.g.*, urinalysis, serum chemistry, etc., and general safety were not reported.

In studies that included an assessment of OP-1 antibodies, they were detected.

Reproductive toxicity
In a fetal development study, female rabbits were immunized with OP-1, then mated and euthanized on Gestation Day 29 (GD29). Rabbits developed IgG and IgM antibodies that were transmitted to the fetal circulation. The antibody's neutralizing activity was not evaluated.
In treated does, fertility was not affected. In fetuses from treated does, there was an increased incidence in external malformations (open eye/absent eyelid, malrotated hindlimb), skeletal malformations (*e.g.*, misshapen rib, fused skull bone, fused sternebrae, misshapen thoracic centra) and skeletal variations (*e.g.*, misaligned sternebrae, unossified sternebrae, misshapen sternebrae).

In a pre- and postnatal study, female rabbits were immunized with OP-1, mated and allowed to deliver. Necropsy was performed on lactation day 28. OP-1 antibodies (IgG, IgM) were detected in serum and milk of treated does and kits. Data revealed an adverse effect on postnatal growth (reduced kit body weight at 3-4 weeks of age), but did not reflect the skeletal effects observed in the developmental study. However, a fetal histological examination was not performed.

Due to species differences in antibody transfer, the predictive value of the rabbit model is not clear. However, the results suggest there may be a risk for prenatal developmental effects in women treated with products containing OP-1.

ASSESSMENT OF IMMUNOGENICITY ASSAYS

The sponsor initially developed four assays to evaluate OP-1 immunogenicity:

- a screening ELISA that detects anti-IgG, IgM, IgA and IgE;
- a titering ELISA to confirm the results of the binding ELISA;
- a neutralizing assay based on the ability of anti-sera to inhibit the production of alkaline phosphatase in the ROS cell line; and
- an alternative quantitative polymerase chain reaction (QPCR) based neutralizing assay that evaluates the ability of anti-sera to inhibit the induction of the Id-1 gene in the A549 cell line.

The binding and titrating assays were determined to be acceptable by the agency. The data generated regarding the incidence of binding antibodies is believed to be accurate.

The ROS cell line based neutralizing assay, however, was determined to be irreparably flawed. The incidence of neutralizing antibodies reported for the pivotal trial was based on this flawed assay, therefore, the reported numbers are unreliable.

The sponsor developed a third neutralizing assay based on quantification of transcripitional activity from a promoter reporter construct. In this case, the sponsor linked a promoter element from the Id-1 gene to a luciferase reporter construct. Both the QPCR and the promoter reporter assays are suitable for generating accurate clinical data. As described below, the collected sera have not been reanalyzed using either of these new, validated assays. As a result, the actual rate of neutralizing antibodies (currently unknown) has not been definitively determined, but has the potential to be different from the reported rate (25.6%).

CMC SUMMARY

DESCRIPTION OF THE BULK OP-1

The active form of OP-1 is a disulfide linked dimer of two polypeptide chains. The molecular weight of the full-length (139 amino acid) OP-1 monomer is approximately 17 kDa. Bulk OP-1 is a heterogeneous population of both homo and heterodimers which represent species with differing degrees of glycosylation as well as species containing N-terminal truncation. Each OP-1 monomer contains seven cysteines. Six cysteines participate in three intramolecular disulphide bonds, with the remaining cysteine participating in an intermolecular disulphide bond with a second OP-1 monomer. The arrangement of intramolecular disulphide bonds has been termed the cysteine "knot".

Complexity

As indicated above, bulk OP-1 is a heterogeneous population of both homo and heterodimers caused by proteolytic activity during cell growth, cell harvest and conditioned media storage resulting in truncation of OP-1 from the N-terminus. N-terminal truncation is apparent with species present beginning with N-terminal at amino acids Ser^{8}, Met^{23}, and Ala^{24}. The full-length sequence (starting with Ser^{1}) is the predominant species in OP-1.

Additionally, OP-1 is glycosylated. Approximately 10% of the mass of OP-1 is attributable to N-linked carbohydrate structures found at Asn^{10}, Asn^{29} and Asn^{80}. The Asn^{80} site is heavily or completely glycosylated, whereas the occupancy at Asn^{10} and Asn^{29} is < 10%. Approximately 70% of OP-1 glycoforms contain high-mannose oligosaccharide units. The remaining 30% contain a mixture of non-sialylated and sialylated complex biantennary and triantennary oligosaccharide units.

Biological Activity Assay

The *in vitro* Biological Activity assay is a bioassay performed using rat osteosarcoma (ROS) 17/2.8 cells. ROS cells are clonal osteosarcoma cells which induces alkaline phosphatase production in a dose dependent manner. Alkaline phosphatase (ALP) is a membrane-bound metalloenzyme which catalyzes the hydrolysis of phosphomonoesters at an alkaline pH. While alkaline phosphatase (ALP) is a commonly used marker of bone formation, it lacks sensitivity and specificity because it is produced by osteoblasts (OBs) and neutrophilic granulocytes of the nonspecific liver/bone/kidney type ALP.[1] A moderately increased ALP level may be a reflection of growth of the skeleton in childhood, repair of a recent fracture, osteomalacia, Paget's disease of bone, hepatobiliary disease, malignancy or an effect of medication.[2] Therefore ALP induction is a relevant measure of biological activity.

OP-1 Manufacture

OP-1 drug substance is produced as a secreted protein in large-scale cell culture using a Chinese hamster ovary (CHO) cell line. After expansion, cells are grown in a production bioreactor. The harvest is subjected to a series of chromatographic and filtration steps, which serve to remove process-related impurities and adventitious virus. Once purified the OP-1 product is stored at 2° to 8°C until it is filled and vacuum dried.

Release specifications

The tests for release for OP-1 include peptide mapping (to confirm identity), SDS PAGE gels (to examine purity and N-terminal heterogeneity), size exclusion HPLC (demonstrating the product has low

[1] Miao D, Scutt A. "Histochemical localization of alkaline phosphatase activity in decalcified bone and cartilage," *J Histochem Cytochem*, 2002 Mar;50(3):333-40.
[2] http://www.merckmedicus.com/pp/us/hcp/diseasemodules/osteoporosis/text/dia_txtl6.jsp

levels of high molecular weight species, i.e. aggregates), reverse phase HPLC (to determine oxidation levels) and monosaccharide analysis (to show consistency of Glycosylation).

Degradation and Stability

OP-1 modes of degradation include an increase in aggregates, deamidated and oxidized species. Real time stability data indicate that the product appears to have a robust stability profile. Aggregation has been shown to decrease purity in SDS-PAGE at higher temperatures. Stryker demonstrated that the following assays can detect degradation:

- Percent oxidation by reduced rpHPLC;
- Aggregation by Size Exclusion Chromatography and Analytical Ultracentrifugation;
- Purity by reduced and non-reduced SDS-PAGE;
- *In Vitro* Biological Activity;
- Isoaspartate content (deamidation by ISOQUANT assay); and
- Sialic Acid content

DESCRIPTION OF COLLAGEN MATRIX

Bovine long bone midshafts are used in the production of bone powder. The bone powder is subsequently demineralized by hydrochloric acid treatment, and then treated with guanidine hydrochloride and hot acetic acid to remove non-collagen proteins. The resultant collagen matrix consists of predominantly Type I collagen. Although some of these treatments serve to reduce potential viral contaminants, it is important to note that the Collagen Matrix is not sterilized prior to incorporation into OP-1 Implant. Release tests for Collagen Matrix include SDS-PAGE analysis, bioburden, particle size, loss on drying and percent SDS-extractable proteins (Purity).

DESCRIPTION OF OP-1 IMPLANT

OP-1 drug substance is reconstituted with a target concentration of 0.39 mg/mL prior to mixing with the collagen matrix. The Collagen matrix is hydrated and then mixed into the OP-1 solution to form a suspension which is subsequently vacuum dried. After drying, the OP-1/Collagen matrix is powder-mixed. The final vials are filled and stoppered under vacuum, capped, labeled, placed in secondary packaging and shipped to a vendor who performs terminal sterilization by Gamma irradiation at > 24-kGy. The sterilized vials are shipped back to Stryker for release testing, final container packaging and labeling.

Release specifications

The tests for release for OP-1 Implant include SDS PAGE gels (to examine purity and N-terminal heterogeneity), size exclusion HPLC (demonstrating the product has high molecular weight species, i.e. aggregates), reversed-phase HPLC (to determine oxidation levels), *In vitro* biological activity assay (demonstrating potency), pH, testing for residual TFA and ethanol, endotoxin and sterility/container closure integrity.

CMC CONCERNS

The primary concern with the CMC data is associated with a single aspect of the processing – the terminal radiation sterilization. The sponsor has chosen to terminally sterilize the combination product using high levels of gamma irradiation (24.5 to 31.5 kGy). The effects of radiation on protein structure and function, especially at this high level, are known. The characterization data for the dried protein component clearly indicate that protein oxidation, aggregation, and truncation. have occurred. No data

have been presented which would indicate if the reported high antibody response rate was related to these observed changes. In addition, it is not known if the level of effectiveness results described below are the result of these changes or changes in the potency or stability resulting from the irradiation.

To address this issue, the sponsor was requested to provide an analysis of data derived from *in vitro* biological activity and additional product quality characterization assays (including peptide mapping, performed on OP-1 extracted from implant exposed to the full range of irradiation doses as specified by the SOP. It is possible that additional data, representing doses outside of this range, may be able to support the argument that there is no correlation between gamma irradiation dose and biological activity or clinical efficacy. The requested data would also confirm or negate that the chosen irradiation exposure limits ensure both implant sterility and protein quality; because exposure of proteins, in general, to high levels of irradiation is known to cause significant protein damage and amino acid modification. No data addressing this concern were provided.

In a recent PMA submission, the sponsor provided a brief data summary and several graphs from various non-clinical studies and a current IND study to address concerns related to antibody production and exposure of OP-1 to gamma irradiation. Because complete test reports from these studies were not provided in the amendment, it is not possible to fully correlate the results from these non-clinical evaluations to the observed results from the clinical studies described below.

In addition to this issue, other CMC issues that remain unaddressed, but are not the subject of any proposed Panel discussion, include:

- the need for an appropriate calculation of release and stability specifications for the OP-1 drug substance;
- OP-1 extraction efficiency from the OP-1 Implant;
- the adequacy of the removal of adventitious organisms from the collagen matrix;
- the adequacy of control over parameters affecting the % truncation of the OP-1 molecule during fermentation;
- the determination of protein yields for each process step for the OP-1 drug substance; and

the lack of an assay at release and stability measuring sialic acid content. Sialic acid content was shown to decrease with time and high temperature.

CLINICAL STUDY SUMMARY

The data in this PMA were collected as part of three distinct clinical studies:

a pilot study (referred to by the sponsor as S99-01US);
a pivotal study (referred to by the sponsor as (S01-01US); and
a set of additional clinical data (referred to by the sponsor as "the Extension Clinical Study", 06-UPLF-01).

The pivotal study was the primary source of clinical data with the extension study providing additional data to address specific issues raised during the review of the submission. Because of differences in study design between the pilot and pivotal studies, as well as the small number of enrolled subjects, the pilot study data are summarized below, but were not used either by the sponsor or the agency as part of the evaluation of safety and effectiveness of the identified product.

PILOT STUDY

general information

This study was a prospective, randomized, controlled, multicenter clinical trial designed to evaluate the safety and effectiveness of OP-1 Putty as a grafting material and aid to fusion in the treatment of subjects with single level (L3-S1) degenerative spondylolisthesis (Grade 1-2) and spinal stenosis undergoing decompression and uninstrumented posterolateral lumbar fusion. The original study design was to evaluate OP-1 Putty used by itself compared to OP-1 Putty mixed with autograft as an graft extender. The protocol was revised midway through the study by replacing the OP-1 Putty/autograft extender group with an autograft alone control group.

Subjects were skeletally mature men and women < 81 years of age and were randomized in a 2:1 fashion to investigational (OP-1 Putty) or control (autograft) groups. Surgery consisted of posterior decompression and posterolateral intertransverse process arthrodesis. Follow-up evaluations took place at 6 weeks and 3, 6, 9, 12 and 24 months post-op, then annually until the last subject enrolled in the pivotal study had received a 24 month evaluation.

The primary efficacy endpoint was a composite consisting of radiographic and clinical parameters. Overall success was defined as radiographic success <u>and</u> clinical success. Radiographic success was defined as:

- presence of bridging bone on AP or lateral x-rays at the treated spinal level;
- $\leq 5^{o}$ angular motion; <u>and</u>
- $\leq$ 2 mm translational motion.

Clinical success was defined as

- $\geq$ 20% improvement in ODI (Oswestry Disability Index); and
- absence of reoperation intended to promote fusion at 24 months.

Safety was assessed by comparing the rate and type of complications and the neurological status between the investigational and control subjects. Analysis of treatment success was calculated both with and without a "last observation carried forward" (LOCF) approach.

Secondary efficacy endpoints included:

SF-36 Health Outcomes Survey scores;
leg/buttock pain (measured by VAS);
donor site pain (measured by VAS);
radiographic disc height; and
degree of angular and translational motion, determined radiographically.

Because of the presence of the recombinant human protein, all subjects had blood drawn pre-operatively and at 6 weeks, and 6 months post-op for serum antibody testing.

Due to the nature of second-site surgery to harvest the grafts in the control group (or the original OP-1 Putty/autograft extender group), blinding of the subjects and investigators was not possible. Radiographic assessments were performed by 2 independent, blinded radiologists with discrepancies resolved by a third reviewer. It should be noted that these assessments are the only blinded data in the study.

A total of 57 subjects were enrolled. Of these 48 were treated and included in the analysis (24 investigational subjects, 12 OP-1 Putty/autograft subjects and 12 autograft control subjects). Because of the small sample sizes in each group, statistical analyses were not performed. The following effectiveness and safety results were reported:

effectiveness

	OP-1 Putty		OP-1 Putty + autograft		autograft control	
	n	# (%) success	n	# (%) success	n	# (%) success
overall success with LOCF	24	11 (45.8)	12	5 (41.7)	12	4 (33.3)
overall success without LOCF	18	10 (55.6)	9	4 (44.4)	9	3 (33.3)
radiographic success	19	11 (57.9)	10	5 (50.0)	10	4 (40.0)
bridging bone	19	15 (78.9)	10	7 (70.0)	10	9 (90.0)
angulation success	18	12 (66.7)	10	5 (50.0)	10	5 (50.0)
translation success	18	16 (88.9)	10	6 (60.0)	10	7 (70.0)
ODI	18	17 (94.4)	9	8 (88.9)	10	6 (60.0)
no retreatment	24	24 (100)	12	11 (91.7)	12	12 (100.0)

Because these subjects were followed until completion of the pivotal study, additional 36 and 48 month overall success data were available (missing data were not imputed):

	OP-1 Putty		OP-1 Putty + autograft		autograft alone	
	n	# (%) success	n	# (%) success	n	# (%) success
36 months	18	9 (50.0)	8	3 (37.5)	5	1 (20.0)
48 months	16	10 (62.5)	7	3 (42.9)	6	2 (33.3)

safety

The following safety observations were made:

- AEs occurred in 100% of subjects across all treatment groups:
  149 in the 36 subjects in the OP-1 Putty groups (with or without autograft)
  51 in the 12 subjects who received autograft alone

- Four subjects across treatment groups reported 8 malignancies, none of which were attributed to the use of OP-1 Putty.

- One death, due to carcinomatosis, was reported.

- Three subjects treated with OP-1 Putty and 1 subject treated with autograft developed heterotopic bone formation radiographically, but no AEs were associated with this finding.

- Four wound infections were noted in subjects treated with OP-1 Putty only, but none in the other 2 groups.

- Pseudarthrosis was reported predominantly between 6 and 12 months post-op. This AE only occurred in the subjects who received OP-1 either alone (10 subjects) or in combination with autograft (3 subjects). Only 2 of these 13 subjects required a surgical retreatment due to pseudarthrosis. Pseudarthrosis was observed in 4 of the 7 (57%) subjects with neutralizing antibodies.

- Post-operative wound infections occurred in 4 subjects who received OP-1 Putty alone and did not occur in subjects who received an autograft.

- There was no evidence of neurological deterioration post-operatively in any treatment group.

- 83% (10/12) of subjects in the OP-1 Putty/autograft group and 91.7% (22/24) of subjects in the OP-1 Putty alone group had antibody titers at 6 months. Of these, zero subjects in the combination therapy group developed neutralizing antibodies, while 7 of 24 subjects in the OP-1 Putty alone group developed neutralizing antibodies at 6 weeks with only 1 subject still having neutralizing antibodies at 6 months. The overall success rate in subjects with neutralizing antibodies was similar to the success rate in subjects without them in the OP-1 treatment groups.

- Seven subjects in the OP-1 Putty group had neutralizing antibodies at either 6 weeks or 6 months post-op, however no pattern of immunologically-mediated AEs was observed. The overall success rates of OP- 1 Putty subjects with or without neutralizing antibodies appeared comparable (42.3% vs. 44.8%.) In this trial, no relationship between antibody occurrence and clinical events was observed.

PIVOTAL STUDY

general information

This study was a prospective, multicenter, randomized, open-label, non-inferiority trial designed to evaluate the safety and effectiveness of OP-1 Putty as a replacement for autograft in subjects with single level (L3-S1) degenerative spondylolisthesis (Grade 1-2) and spinal stenosis undergoing decompression and uninstrumented posterolateral lumbar fusion. Subjects were skeletally mature men and women < 85 years of age. Enrollment took place at 25 institutions in the US and Canada and subjects were randomized in a 2:1 fashion to either OP-1 Putty (investigational group) or autogenous iliac crest bone graft (control group). Due to the nature of second-site surgery to harvest the grafts in the control group, blinding of the subject and investigators was not possible. Radiographic assessments were performed by 2 independent, blinded radiologists with discrepancies resolved by a third reviewer. It should be noted that these assessments are the only blinded data in the study.

The primary efficacy endpoint was overall treatment success at 24 months, defined as a composite of 5 criteria:

- ≥ 20% improvement in ODI (Oswestry Disability Index);
- radiographic spinal fusion defined as :
    - the presence of bridging bone;
    - ≤ 5° angulation on flexion/extension radiographs at the affected level; <u>and</u>
    - ≤ 2mm translational motion on flexion/extension radiographs at the affected level);
- absence of a decrease in neurological status (muscle strength, reflexes, sensory, and straight leg raise) unless attributable to a concurrent medical condition or to the surgical procedure by a blinded independent neurological reviewer;
- absence of retreatment; and
- absence of treatment-related serious adverse events (SAEs).

In order to be considered as a treatment success, a subject had to be successful in each of the individual primary clinical and radiographic endpoints.

Secondary efficacy endpoints included:

- overall success at 12, 24, and 36 months without imputation of missing data;

- components of overall success (ODI, absence of treatment-related SAEs, absence of retreatment to promote fusion, neurological success, and overall radiographic success at 12, 24, and 36 months without imputation of missing data); and
- overall radiographic success at 24 months with missing data imputed.

Safety endpoints included analyses of any reported AEs, clinical laboratory evaluations and neurological status.

Additional information collected for each subject included:

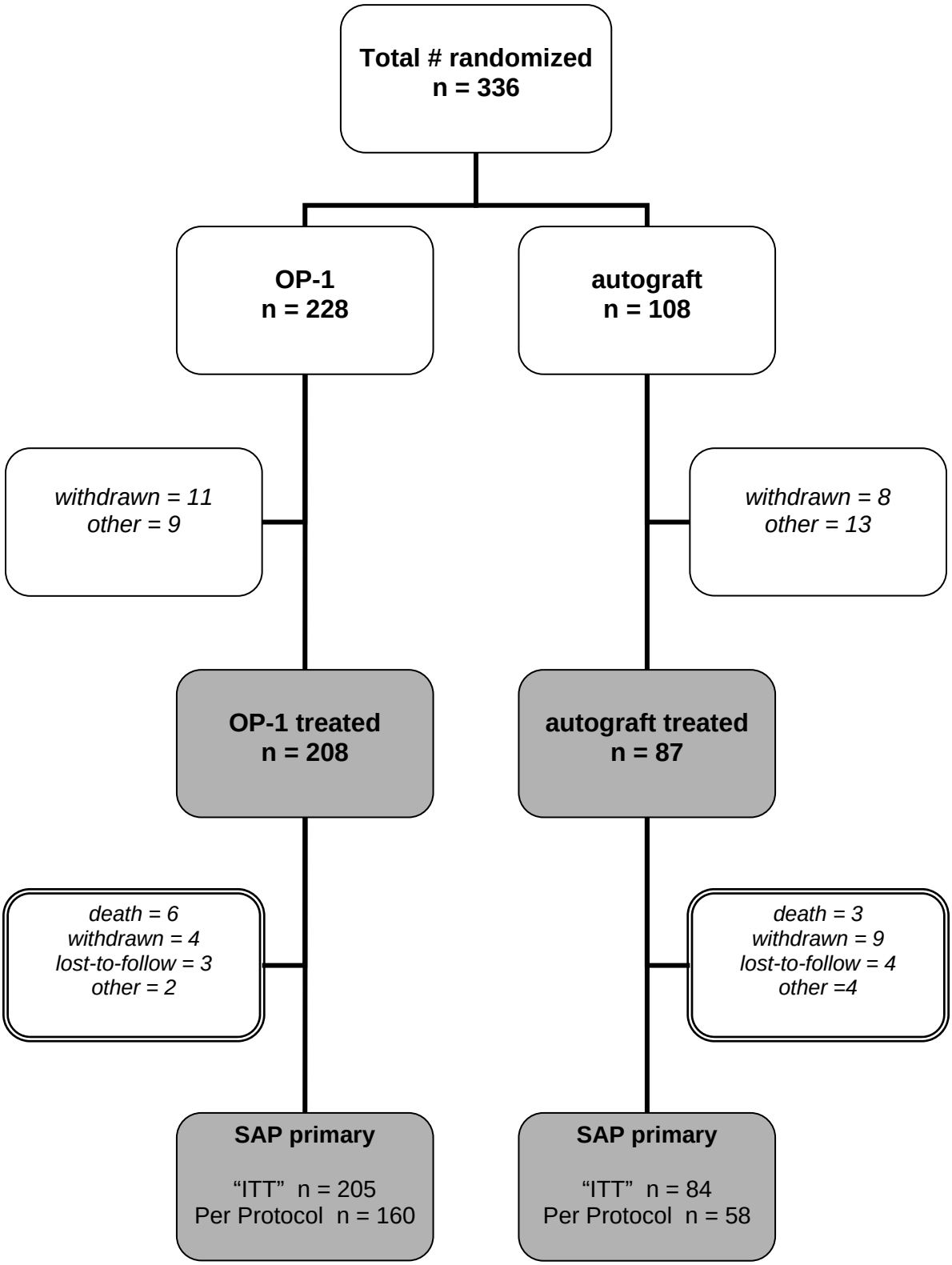
- VAS results for pain assessment;
- donor site pain (control subjects only);
- medication use;
- hospitalization data; and
- general health survey (SF-36).

In addition to the plain film radiographic data collected at each post-op evaluation, a CT evaluation was to be performed at 9 months post-op in order to assess bridging bone formation and pseudoarthrosis. These data were not to be used as part of the assessment of radiographic success or considered as a study endpoint, however.

Subject assessment was performed pre-operatively, at the time of hospitalization (just prior to surgical procedure), and at the following post-op timepoints –within 72 hours of surgery, at 6 weeks and at 3, 6, 9, 12, and 24 months, then annually until the last subject enrolled in the study had received a 24 month evaluation. In order to determine the immune response to the recombinant human protein, blood was drawn pre-operatively, immediately post-op and at 6 weeks and 3, 6, 12, and 24 months post-op.

The study planned to treat 312 subjects (208 OP-1 Putty and 104 autograft) in order to achieve 80% power at one-sided alpha level of 0.05 under the expectation of 53% overall success rate in the OP- 1 Putty group and 47% for the autograft group with approximately 15% of the treated subjects not being evaluable at 24 months). The maximum allowable difference between the treatment groups that can be used to conclude that OP-1 Putty is not inferior to autograft was set at 10%, *i.e.*, the , non-inferiority margin is 10%.

As shown in the figure below, a total of 336 subjects were enrolled and randomized. Of these, 295 were ultimately treated - 208 received OP-1 Putty and 87 received autograft. The remaining 41 randomized subjects did not receive any study treatment due to withdrawal or other reasons prior to study treatment (investigational: 20/228 = 9% vs. control: 21/108 = 19%). Through the post-operative 24 months, the control group had a much higher percentage of withdrawals (control: 20/87 = 23% vs. investigational : 15/208 = 7%) mainly due to the differences in voluntary withdrawal, withdrawals by investigator, and lost-to-follow-up.



In accordance with the approved IDE protocol, the analysis of 24 month overall success rate was to be the basis for the test of non-inferiority. The primary null hypothesis was to be tested by calculating the difference between the proportion of success in the two treatment groups (P_{auto}- P_{OP-1}), the associated standard error and the upper 95% confidence limit (1.64 times the standard error.) Both per-protocol and

intent-to-treat analyses were planned with subjects to be classified based on the definition of individual subject success/failure.

The per-protocol analysis was to exclude subjects who violated the inclusion/exclusion criteria. Subjects were also to be excluded if they were missing an Oswestry assessment at 24 months, if their 24 month radiographic results were missing or not evaluable, or if the subject was missing a neurological assessment.

The intent-to-treat analysis was to consist of all treated subjects. Subjects who were considered no longer participating in the study at the 24 month follow-up visit or who had missing data during the 24 month follow-up interval were to be considered as missing. For this analysis, subjects with missing data were to be initially classified using the last observation carried forward approach (LOCF) followed by a sensitivity analysis to examine the stability of the conclusions to alternative classification methods

effectiveness discussion

Prior to submission of the PMA (after all clinical data had been collected but prior to closure of the database), the sponsor submitted a modification to the statistical plan that proposed a change to:

- the definition of overall success;

  The overall radiographic success was changed for that described listed above to the following:

  - presence of bone (rather than bridging bone**)**;
  - angulation of ≤5°, and
  - translational movement of ≤ 3mm (rather than ≤ 2mm)

- the definition of the efficacy populations for analysis;

  The efficacy population analyses were conducted on a modified intent-to-treat population (mITT) which included all treated subjects with at least one post-treatment follow-up visit. The analyses approved in the original protocol incorporating the original success definition were to be performed on the intent-to-treat population. Both efficacy analyses were also to be performed on a per protocol population.

- a modification to the fixed non-inferiority margin, as well as other minor clarifications and changes.

  The non- inferiority margin approved in the original IDE protocol was specified at 10%. In the proposed change, this was modified to be variable with a maximum of approximately 14% depending on the success rate in the control group. The non-inferiority margins for all other variables remained at 10%. In addition, the overall success and overall radiographic success endpoints at 24 months imputation was changed from LOCF to multiple imputations.

As part of the PMA submission, the sponsor included a *post hoc* analysis by introducing a second change of the overall success definition, *i.e.*, referred to as the "overall clinical success", by removing all radiographic data from the pre-defined composite primary endpoint. The sponsor's rationale for performing the *post hoc* analysis was that "…[e]vidence for presence of bone on X-ray was inconsistent with the beneficial effects noted in angular and translational movement…"

For this dataset, overall success was defined as:

- ODI success;
- absence of retreatment;

- absence of serious treatment-related adverse events; and
- neurological success.

In addition, this analysis did not incorporate data from all treated subjects or include a sensitivity analysis to evaluate the potential impact of this exclusion.

The agency has questions related to the first change to the success definitions due to the clinical relevance of the new definition of radiographic success. The agency also has concerns related to the second definition of overall success used in the *post hoc* analysis.

First, subjects who receive the investigational product are expected to form a solid fusion mass in the absence of any bone graft. Bone formation relies solely on the ability of the recombinant protein signal to elicit cellular differentiation and proliferation. Short of performing another invasive surgery to explore the surgical site for the presence of a fusion mass, the agency is unaware of another reliable way to detect the presence of bone other than by radiographic techniques. Because the natural history of spondylolisthesis progression remains unclear, it is better to consider radiographic evidence of bone formation, especially bridging bone, as an important indicator of whether or not an investigational treatment achieved its goal of causing bony fusion.

Second, the sponsor's rationale for the *post hoc* analysis, *i.e.*, inconsistency between the bone formation and spinal stability, was made without supporting data. If this type of inconsistency were present, it would have been possible to confirm it through an analysis assessing for a correlation between bone formation and spinal stability. Although requested, this type of analysis was not provided.
Third, the sponsor's *post hoc* analysis was subject to bias in favor of the investigational treatment because:

(1) the primary endpoint was redefined retrospectively;
(2) the radiographic component was the sole blinded endpoint;
(3) a much higher percentage of the excluded subjects were from the control group (percentage excluded: control: 15/86 = 17% vs. investigational: 16/207 = 8%); and
(4) most of the excluded control subjects were considered successes at 12 months post-op.

Finally, the Type I error rate could be seriously inflated without any adjustment for the *post hoc* nature of changing the primary endpoint and its associated multiple analyses. All of these issues were expressed to the sponsor. Even acknowledging the questions raised by these changes, the agency reviewed and analyzed the data from the pivotal study using these multiple success definitions.

Using the original definition of overall success approved in the clinical study, both the intent-to-treat analysis (ITT, Table 1a, p = 0.824) and the per-protocol analysis (Table 1b, p = 0.942) failed to show that OP-1 was non-inferior to autograft in terms of the overall success at 24 months (these analyses were performed in accordance with the original protocol approved in the clinical trial (referred to as the pre-specified statistical analysis plan (SAP)). Consistent with the sponsor's non-inferiority tests, the agency's analyses based on the reported data showed that the control treatment was actually superior to the investigational treatment through the 24 month follow-up in terms of the overall success rate. At 24 months, the overall success rate in the control group was significantly higher than that in the investigational group (ITT: control = 48% vs. investigational = 32%, Fisher's exact test P = 0.015, 90% CI = [5%, 26%]; per-protocol: control = 57% vs. investigational = 35%, Fisher's exact test P = 0.005, 90% CI = [10%, 34%]).

Table 1 and and b: Overall success rates per the originally approved protocol (pre-specified SAP)

ITT[1] (*Missing data were imputed using last observed value carried forward approach)*

Time	Control (n=87)	investigational (n=208)	P-value for non-inferiority	Difference (P_{Auto} - P_{OP})	90% CI (P_{Auto} - P_{OP})	P-value for (P_{Auto} - P_{OP})
3 months	42% (33/79)	20% (39/196)	0.971	22%	[12%, 32%]	0.0004
6 months	37% (31/83)	30% (61/204)	0.340	7%	[-3%, 18%]	0.264
12 months	48% (40/84)	31% (63/205)	0.863	17%	[6%, 27%]	0.010
24 months	**48% (40/84)**	**32% (65/205)**	**0.824**	**16%**	**[5%, 26%]**	**0.015**

Per-protocol analysis *(Missing data were not imputed)*

Time	Control (n=87)	Investigational (n=208)	P-value for non-inferiority	Difference (P_{Auto} - P_{OP})	90% CI (P_{Auto} - P_{OP})	P-value for (P_{Auto} - P_{OP})
3 months	47% (26/55)	22% (33/153)	0.982	25%	[13%, 38%]	0.0005
6 months	43% (24/56)	32% (50/157)	0.553	11%	[8%, 15%]	0.145
12 months	60% (31/52)	31% (48/153)	0.990	29%	[15%, 41%]	0.0005
24 months	**57% (33/58)**	**35% (56/160)**	**0.942**	**22%**	**[10%, 34%]**	**0.005**

Note:
1. In the intent-to-treat analysis (ITT), patients without of any post-operative data were excluded (*e.g.*, 3 subjects in the autograft group and 3 in the OP-1 group were excluded from the 24-month analysis. In the per-protocol analysis, subjects with missing data or with violation of inclusion/selection criteria were all excluded.
2. Sponsor's reported p-values based on one-sided two sample t-test with a non-inferiority margin of 10%.
3. FDA calculated 90% confidence intervals (CI) for the difference between the two group success rates. P-values were based on two-sided Fisher's exact test for the difference.

Using the second definition of overall success, where the radiographic endpoint was modified, the investigational treatment was also not found to be non-inferior to the control treatment (modified ITT [mITT], P = 0.331 with multiple imputations, control = 49% vs. investigational = 39%, 90% CI = [1%, 23%]).

It should also be noted that the overall radiographic success rate was lower for the investigational treatment compared to the control treatment. This large difference in overall radiographic success rate was primarily driven by the differences in bone formation (bridging bone in accordance with the original radiographic success definition compared to any bone formation in the second radiographic success definition.) The proportion of subjects with the presence of bone on plain film at 24 months was significantly lower in the investigational treatment compared to the control treatment (61.7% for the investigational treatment vs. 83.1% for the control treatment, $P < 0.001$). In addition, the investigational treatment was not demonstrated to be non-inferior to the control treatment with respect to angular success at 3, 6, 12 and 24 months (mITT population with missing data imputed using LOCF, upper bound of 95% CI for the difference = 11.5%, $P = 0.087$).

Using the third definition of overall success, which excluded all of the radiographic data, the sponsor claimed that the investigational treatment was statistically non- inferior to the control treatment (investigational: 136/191 = 71.2% vs. control: 49/71 = 69.0%, unadjusted $P = 0.029$), and later agreed with the agency that such claim of statistical significance was not appropriate due to the concerns outlined above.

The following tables provide additional information concerning the overall success and radiographic success outcomes based on the various definitions of success:

Table 2: Success defined as ODI improvement, absence of retreatment and of serious AEs and neurological outcome **(ITT population[1] per the first definition of overall success)**

Endpoint	Time	Control (n=87)	Investigational (n=208)	P-value for non-inferiority[2]	Difference (P_{Auto} - P_{OP})	90% CI[3] (P_{Auto} - P_{OP})	P-value[3] for (P_{Auto} - P_{OP})
ODI	12 months	84% (65/77)	83% (155/187)	0.045	1%	[-7%, 10%]	0.857
	24 months	**85% (53/62)**	**80% (144/179)**	**0.180**	**5%**	**[-4%, 14%]**	**0.448**
No retreatment	12 months	96% (78/81)	95% (190/200)	<0.001	1%	[-3%, 6%]	0.763
	24 months	93% (62/67)	9% (179/193)	**0.003**	**0%**	**[-6%, 6%]**	**1.000**
No serious treatment related AEs	12 months	96% (78/81)	91% (181/200)	0.078	6%	**[1%, 11%]**	0.141
	24 months	**96% (64/67)**	**89% (172/193)**	**0.141**	**7%**	**[1%, 12%]**	**0.145**
Neurological	12 months	97% (73/75)	98% (189/193)	<0.001	-1%	[-4%, 3%]	0.673
	24 months	94% (62/66)	100% (189/189)	**<0.001**	**-6%**	**[-11%, -1%]**	**0.004**
Radiographic	**Control treatment was significantly superior to investigational treatment (see Tables 4-6)**						

Note:
1. Missing data were not imputed.
2. Sponsor's reported p-values based on one-sided two sample t-test with a non-inferiority margin of 10%.
3. FDA calculated 90% confidence intervals (CI) for the difference between the two group success rates. P-values were based on two-sided Fisher's exact test for the difference.

Table 3: Overall radiographic success rates using original radiographic success definition

ITT[1] *(Missing data in plain film were imputed using last observed value carried forward approach)*

Time	Control (n=87)	Investigational (n=208)	P-value for non-inferiority[2]	Difference (P_{Auto} - P_{OP})	90% CI[3] (P_{Auto} - P_{OP})	P-value[3] for (P_{Auto} - P_{OP})
3 months (plain film)	53% (41/78)	25% (48/193)	0.997	28%	[17%, 38%]	0.00003
6 months (plain film)	48% (39/81)	37% (76/203)	0.543	11%	[0%, 21%]	0.109
12 months (plain film)	61% (50/82)	40% (81/204)	0.961	21%	[11%, 32%]	0.0016
24 months (plain film)	**65% (53/82)**	**40% (82/204)**	**0.989**	**25%**	**[14%, 35%]**	**0.0002**

Per-protocol analysis[1] *(Missing data were not imputed)*

Time	Control (n=87)	Investigational (n=208)	P-value for non-inferiority[2]	Difference (P_{Auto} - P_{OP})	90% CI[3] (P_{Auto} - P_{OP})	P-value[3] for (P_{Auto} - P_{OP})
12 months (plain film)	73% (38/52)	38% (58/152)	1.000	35%	[23%, 47%]	0.00002
24 months (plain film)	**74% (43/58)**	**40% (64/160)**	**1.000**	**34%**	**[23%, 46%]**	**0.00001**

Note:

1. In the intent-to-treat analysis (ITT), subjects without of any post-operative data were excluded (e.g., 4 subjects in the control group and 5 in the investigational group were excluded from the 24-month analysis). In the per-protocol analysis, subjects with missing data or with violation of inclusion/selection criteria were all excluded.
2. Sponsor's reported p-values based on one-sided two sample t-test with a non-inferiority margin of 10%.
3. FDA calculated 90% confidence intervals (CI) for the difference between the two group success rates. P-values were based on two-sided Fisher's exact test for the difference.

Table 4: Overall Radiographic Success Rates using the second radiographic success definition

	mITT[1]					
Time	**Control (n=87)**	**Investigational (n=208)**	**P-value for non-inferiority[2]**	**Difference (P_{Auto} - P_{OP})**	**90% CI[3] (P_{Auto} - P_{OP})**	**P-value[3] for (P_{Auto} - P_{OP})**
12 months (plain film)	73% (52/71)	49% (88/179)	0.985	24%	**35%]**	0.0006
24 months[1a] (plain film)	75% (44/59)	52% (87/166)	0.961	23%	[11%, 34%]	0.0034
24 months[1b] (MI)	**69% (59/86)**	**53% (110/207)**	**0.622**	**16%**	**[5%, 26%]**	**0.019**

Per-protocol analysis[1] *(Missing or non-evaluable data were excluded)*

Time	**Control (n=87)**	**Investigational (n=208)**	**P-value for non-inferiority[2]**	**Difference (P_{Auto} - P_{OP})**	**90% CI[3] (P_{Auto} - P_{OP})**	**P-value[3] for (P_{Auto} - P_{OP})**
12 months (plain film)	74% (52/70)	49% (87/177)	0.990	25%	[15%, 36%]	0.0004
24 months (plain film)	**75% (44/59)**	**52% (86/164)**	**0.960**	**23%**	**[11%, 34%]**	**0.0034**

Note:

1. The modified intent-to-treat population (mITT) included all treated subjects with at least one post-treatment follow-up visit: 1a-missing or non-evaluable data were excluded; 1b-missing data were imputed and the percentage of success was based on estimates from multiple imputations. In the per-protocol analysis, subjects with missing data or with violation of inclusion/selection criteria were all excluded.
2. Sponsor's reported p-values based on one-sided two sample t-test with a non-inferiority margin of 10% in cases of no missing data imputation. In the case of multiple imputations, P-value was based on the test in the angular scale with 14% non-inferiority margin.
3. FDA calculated 90% confidence intervals (CI) for the difference between the two group success rates. P-values were based on two-sided Fisher's exact test for the difference.

Table 5: Presence of bone through the post-operative follow-up (mITT population[1])

Time	**Control (n=87)**	**Investigational (n=208)**	**P-value for non-inferiority[2]**	**Difference (P_{Auto} - P_{OP})**	**90% CI[3] (P_{Auto} - P_{OP})**	**P-value[3] for (P_{Auto} - P_{OP})**
3 months (plain film)	83% (67/81)	52% (104/200)	1.000	31%	[22%, 40%]	0.000001
6 months (plain film)	87% (71/82)	60% (122/204)	1.000	27%	[18%, 35%]	0.000007
9 months (CT-scan)	*99% (73/74)*	*85% (156/184)*	*0.929*	*14%*	*[9%, 19%]*	*0.00074*
*9 months (**Bridging** Bone CT-scan)*	*54% (40/74)*	*31% (57/184)*	*Not reported*	*23%*	*[12%, 34%]*	*0.00066*
12 months (plain film)	84% (70/83)	62% (126/205)	0.994	22%	[14%, 31%]	0.0001
24 months (plain film)	**83% (69/83)**	**62% (126/205)**	**0.990**	**21%**	**[13%, 30%]**	**0.0003**

Note:
The mean scores of all reviewers are used in the analysis.

1. Missing data were imputed using the last observation carried forward approach for 6, 12, 24 months.
2. Sponsor's reported p-values based on one-sided two sample t-test with a non-inferiority margin of 10%.
3. FDA calculated 90% confidence intervals (CI) for the difference between the two group success rates. P-values were based on two-sided Fisher's exact test for the difference.

In summary, the pivotal study did not show that OP-1 Putty was non-inferior to autograft in terms of subject overall success rate based on any of the previous success definitions. This finding held true for the success definition which excluded the radiographic data (and as described above would have been potentially biased in favor of the investigational treatment). The overall success rate in the control group was higher when compared to the investigational group and was primarily due to the higher rate of overall radiographic success in that control group. The lack of bone formation, especially bridging bone, appears to be the primary source of the poor radiographic outcome in the investigational group. Control subjects had a higher rate of bone formation compared to the investigational subjects at 3 months post-op and this remained higher throughout the remainder of the follow-up period.

EXTENSION STUDY

The sponsor agreed with the agency that there were potential issues or biases with the *post hoc* analyses and that certain aspects of the data may not have been adequately addressed by the analyses. The sponsor also agreed that excluding the radiographic data from the definition of overall success may not be clinically appropriate. Based on a *post hoc* reanalysis of the 24 month plain films and the 9 month CT scans on a subset of the treated subjects, the sponsor proposed an additional analysis that utilized a fourth definition of overall success.

From this analysis, the sponsor determined that the original radiographic assessment had been faulty as demonstrated by their determination that in many cases the 24 month plain films did not exhibit evidence of bone formation, while the 9 month CT scans did. The sponsor stated that two points could be derived from this evaluation. The first was that the use of plain films was inappropriate for the evaluation of their product. The second was that the initial radiographic reviewers had been looking in the wrong location for the fusion mass. This was based on their observation that the implants had migrated laterally. Because the radiographic reviewers were expecting to find bone formation in a more lateral location, the sponsor believed that the evaluators missed the more medial bone.

In order to address their concerns with the 24 month plain films and the altered location of the resulting fusion masses, the sponsor designed and conducted an extension study. The purpose of this new prospective study was to collect a single CT scan image from any available subjects. Because the sponsor had already completed the follow-up on the treated subjects several years previously, the new CT scan data would represent a time period ranging from 36+ months post-op to almost 72 months post-op. Specifically, each available subject would provide a single CT scan that could have been taken anywhere from 36+ months to almost 72 months post-op. The resulting dataset would not be representative of any single post-op timepoint and did not include data from all treated subjects. In addition to this new radiographic data, the sponsor also collected a set of clinical data, including whether or not a subject had had a reoperation.

Using the additional CT scan and reoperation data, the sponsor proposed a fourth definition of overall success. This definition combined the 24 month clinical outcome data with the new 36+ month CT scan/re-operation data. The actual amount of time that elapsed between the original study surgery and the extension study evaluation was 1633.3 ± 149.02 days (mean ± SD, range = 1374-1995 days) for the control group compared to 1599 ± 147.31 days (mean ± SD, range = 1344-1980 days) for the investigational group. With approximately 80% of eligible subjects being accounted for in the extension

study (control: 58/74 vs. investigational: 144/183), the sponsor's analysis showed that the investigational treatment was statistically non-inferior to the control treatment (investigational = 47.2% vs. control = 46.8%, unadjusted P = 0.025)

While this new dataset had the potential to provide some longer-term outcome information on the available enrolled subjects, the agency expressed concerns related to:

- the *post hoc* nature of the reanalysis that determined that original data were faulty
- the evidence of migration of the implants that had not been previously observed in either the non-clinical animal studies or the previous clinical experience, *e.g.*, the pilot study;
- the combination of different follow-up time points for different elements of the success definition;
- the reliance on longer-term data that is not consistent with our understanding of clinical practice, *e.g.*, determining the need for intervention some time between 6 and twelve months post-op, therefore raising questions why the 9 month CT scan did not indicate the presence of successful fusion and why it would be necessary to rely on a 36+ month CT scan before being able to determine that a subject had had a successful fusion; and
- the post hoc re-definition of the primary endpoint and its associated statistical deficiencies with respect to the sponsor's inappropriate claim of non-inferiority based on their mITT analysis of the extension study (investigational: 47.2% vs. control=46.8%,unadjusted p = 0.025 with multiple imputation for missing data):
    - The confirmatory nature of the original pivotal trial was compromised due to the change of the primary endpoint after analyzing the data. This would make it difficult to tightly control the Type I error rate under the pre-specified 5% level. Therefore, the extension study results could only potentially be used to justify the value of conducting a future confirmatory trial incorporating CT scans as a primary radiographic tool for detecting the presence of bone.
    - Because the upper bound of the 95% confidence interval for the difference in success rate was beyond the original pre-specified non-inferiority margin of 10%, a statistical non-inferiority claim could not be made, even without any adjustment to the significance level (alpha) for the *post hoc* change of the primary endpoint.
    - All the other analyses without imputation by the sponsor suggested that statistical non-inferiority could not be claimed even without any alpha adjustment, e.g., SPPP: control = 44.2% (23/52) vs. investigational = 42.6% (55/129) and mITT without multiple imputation: control = 39.7% (23/58) vs. investigational = 37.7% (55/146).
    - The underlying statistical assumption of "Missing at Random" (MAR) for the multiple imputation to derive the unadjusted p-value of 0.025 might not be valid because of the high percentage of missing data (29.5% in the investigational group and 32.6% in the control group).

SAFETY DISCUSSION

The safety analysis was divided into two types of analyses – a "traditional" safety analysis that would be applicable to any product intended as an aid to spinal fusion and an analysis focused on potential immune responses to the recombinant protein and bovine collagen components of the product.

"traditional" safety analysis

With respect to the "traditional" analysis, the types of AEs reported for the two groups of subjects were not significantly different from each other or from the types of AEs that would be expected for a spinal fusion surgery. It should be noted, however, that, although not statistically significant, the investigational treatment had a higher proportion of subjects with treatment-related serious AEs (investigational: 25/208 = 12% vs. control: 6/87 = 6.9%, 95% CI= [-1.8%, 12%], Fisher's exact test P=0.22). These are highlighted in the tables below. This is consistent with the finding that the investigational treatment was inferior to the control treatment with respect to the success rate based on the absence of serious treatment-related AEs.

Treatment-Emergent Adverse Events (Safety Population)

Parameter	Investigational (n=208)		Control (n=87)	
	Number (%) of subjects with	95% CI	Number (%) of subjects with	95% CI
	Events		Events	
Any AE	201 (96.6)	[93%, 99%]	82 (94.3)	[87%, 98%]
Severe AE	43 (20.7)	[15%, 27%]	17 (19.5)	[12%, 29%]
Treatment-related AE	54 (26.0)	[20%, 33%]	23 (26.4)	[18%, 37%]
Unanticipated AE	**6 (2.9)**	**[1%, 6%]**	**0 (0.0)**	**[0%, 4%]**
Serious AE	104 (50.0)	[43%, 57%]	43 (49.4)	[39%, 60%]
Serious and Unanticipated AE	**5 (2.4)**	**[1%, 6%]**	**0 (0.0)**	**[0%, 4%]**
Treatment-related Serious AE	**25 (12.0)**	**[8%, 17%]**	**6 (6.9)**	**[3%, 14%]**
Neoplasms Benign, Malignant And Unspecified (incl Cysts And Polyps)	12 (5.8)	[3%, 10%]	8 (9.2)	[4%, 17%]
Death (not related to treatment)	7 (3.4)	[1%, 7%]	4 (4.6)	[1% , 11%]

Absence of Serious Treatment-Related Adverse Events Success Rate (Protocol Defined) at 12, 24, and 36 Months - ITT (*Missing data were not imputed)*

Time	Control (n=87)	Investigational (n=208)	P-value for non-inferiority[1]	Difference (P_{Auto} - P_{OP})	90% CI[2] (P_{Auto} - P_{OP})	P-value[2] for (P_{Auto} - P_{OP})
12 months	96% (78/81)	91% (181/200)	0.079	5.8%	[1%, 11%]	0.1405
24 months	**96% (64/67)**	**89% (172/193)**	**0.141**	**6.4%**	**[1%, 12%]**	**0.1453**

Note:

1. Sponsor's reported p-values based on one-sided two sample t-test with a non-inferiority margin of 10%.
2. FDA calculated 90% confidence intervals (CI) for the difference between the two group success rates. P-values were based on two-sided Fisher's exact test for the difference.

immunological safety analysis

Because the product contains a recombinant human protein which could elicit an immune response, a second safety analysis was performed. To address this concern, the sponsor was required to collect serum from all enrolled subjects (investigational and control) and perform assays evaluating for the presence of OP-1 antibodies and OP-1 neutralizing antibodies.

The sponsor developed assays to detect and binding and neutralizing capabilities of anti-OP-1 antibodies. A titering assay was also developed for the binding assay. Immunogenicity in the pivotal trial was high, with 94% of subjects testing positive for anti-OP-1 binding antibodies at some point in the trial. In addition, although antibody titers decreased over time, many subjects still had significant titers even out to 24 months post-op. This raised concerns with regard to the long-term effects of the antibodies.

The agency believes that the high incidence and long duration of anti-OP-1 antibodies raises questions because OP-1 (BMP-7) has been demonstrated to have important roles in fetal development as well as adult mammals. Specifically:

- studies using animal models indicate that OP-1 has non-redundant roles in female fertility until menopause[1,23];
- OP-1 is critical for fetal kidney and eye development, as well as skeletal patterning in mice. According to the literature, elimination of the OP-1 gene leads to neonatal renal function and fatality[3,4];
- in adult animal models, OP-1 is involved in tissue repair, particularly post-hypoxia/ischemia in the kidney and brain[5,6] ;
- membrane bound OP-1 receptors are constitutively expressed in adult rat kidney[5] and mature mammalian brain[6]; and
- OP-1 mRNA is expressed in normal rat kidney and at lower levels in most other tissues, including gonads.

As described previously, the assay used to detect the presence of neutralizing antibodies was unreliable and the existing serum was not retested using the new assay. Because of this, the actual rate of neutralizing antibodies (currently reported as 25.6%) has not been definitively determined and may be different from the reported rate. While there was no correlation between AEs and the presence of neutralizing antibodies, there was a trend towards decreased overall treatment success (29.8% vs. 40.7%) and decreased overall radiographic success (41.5% vs. 56.0%) in subjects with neutralizing antibodies compared to those with non-neutralizing antibodies, although this trend did not meet statistical significance. It should also be understood, however, that the mechanism of action of the recombinant BMP and the effect of neutralizing antibodies are not completely known. As a result, these observed

---

1) Knight PG, Glister C: TGF-b superfamily members and ovarian follicle development. Reproduction (2006) 132: 191 – 206.
2) Glister C, Richards SL, Knight PG: Bone morphogenic proteins (BMP) -4, -6, and -7 potently suppress basal and luteinizing hormone-induced androgen production by bovine theca internal cells in primary culture: could ovarian hyperandrogenic dysfunction be caused by a defect in BMP signaling? Endocrinology (2005) 146: 1883 – 1892.
3) Dudley AT, Lyons KM, Robertson EJ: A requirement for bone morphogenic protein-7 during development of the mammalian kidney and eye. Genes and Development (1995) 9: 2795 – 2807.
4) Luo G, Hofmann C, Bronckers ALJJ, Sohocki M, Bradley A, Kersenty G: BMP-7 is an inducer of nephrogenesis, and is also required for eye development and skeletal patterning. Genes and Development (1995) 9: 2808 – 2820.
5) Dube PH, Almanzar JJ, Frazier KS, Jones WK, Charette MF, Paredes A: Osteogenic Protein-1: Gene expression and treatment in the rat remnant kidney model. Toxicologic Pathology (2004) 32: 384 – 392.
6) Kawamata T, Ren JM, Chan TCK, Carette M, Finklestein SP: Intracisternal osteogenic protein-1 enhances functional recovery following focal stroke. Neuorpharmachology (1998) 9: 1441 – 1445.

immunological responses remain an area of possible safety concern that would need to be addressed as part of the Panel's discussion.

In light of the agency's questions, the sponsor was asked to provide additional information related to immunological response to the product:

- To assess patient anti-rhuOP-1 sera for their ability to cross react with native human OP-1.
  This assay is still in development but no data were submitted to the PMA
- To isotype patients' anti-OP-1 antibodies.
  These assays have been developed, but data from systematic reanalyses of the treated subjects have not been submitted to the PMA
- To investigate the effect of systemic anti-rhu-OP-1 antibodies on female fertility in an animal model.
  Refer to the pharmacology/toxicology analysis summarized above.
- To investigate the effect of systemic anti-rhu-OP-1 antibodies on renal health using an animal model.
  Refer to the pharmacology/toxicology analysis summarized above.

Based on their responses or lack of responses to these requests, it is not clear that the sponsor has provided adequate information to address the immunological safety questions the agency has regarding this product.

SUMMARY

As described above, the sponsor has submitted non-clinical and clinical data to support the safety and effectiveness of OP-1 Putty as a graft material and aid to fusion in the treatment of single level lumbar spondylolisthesis. Based on reviews of these data, the agency has questions whether or not this combination product can be considered relatively safe and effective when compared to the autograft control.

Based upon the currently submitted pivotal study data, the source of the effectiveness differences between the subjects treated with OP-1 Putty vs. the control treatment is not known and is left as a point of discussion for the Panel. It might be due to selection of the current human dose for the OP-1 component; due to differences in the ability of the product to promote bone formation in humans in contrast to the behavior observed in lower order animals; due to changes in the potency or stability of the recombinant protein after exposure to irradiation sterilization; or due to a clinical response to the oxidized, truncated and aggregated protein. In addition to the effectiveness questions, the agency believes that there are safety questions associated with the immune response to the product.

As with the level of effectiveness results, it is not clear if the observed antibody response (antibody rates, types and duration of effect) is due to the terminal radiation processing or some inherent characteristic of the protein. The agency has questions concerning this as well.