



Food and Drug Administration
Center for Biologics Evaluation and Research
Office of Biostatistics and Epidemiology
Division of Biostatistics

STATISTICAL REVIEW AND EVALUATION BLA

Date: November 19, 2009

BLA/Supplement Number: 125348/000

Product Name: Azfibrocel-T (previously Isolagen Therapy (IT))

Indication(s): Treatment of moderate to severe nasolabial fold wrinkles

Applicant: Fibrocell Technologies, Inc (previously Isolagen Technologies, Inc.)

Date(s): Paper submission of Module 1, Vol.1; Module 2, Vol.1; and Module 5, Vol.1-66 (letter date: March 6, 2009); electronic SAS data sets; and sponsor's response to 5/19/09 letter. PDUFA date: January 4, 2010

Review Priority: Standard (10-month)

Statistical Branch: CBER/OBE/DB/TEB (HFM-219)

Primary Statistical Reviewer: Shiowjen Lee, Ph.D.

Concurring Reviewer (1): Boguang Zhen, Ph.D., Team Leader
Therapeutic Evaluation Branch (HFM-219)

Concurring Reviewer (2): Ghanshyam Gupta, Ph.D., Chief
Therapeutic Evaluation Branch (HFM-219)

Medical Office/Division: CBER/OCTGT/DCEPT

Clinical Reviewer(s): Agnes Lim, MD; Yao Yao Zhu, MD.; Changting Haudenschild, MD (HFM-755)

Project Manager: Lori Tull (HFM-705)

Table of Contents

| | | |
|-------------------------------|---|-------------------------------------|
| 1. | EXECUTIVE SUMMARY | 3 |
| 1.1 | CONCLUSIONS AND RECOMMENDATIONS..... | 3 |
| 1.2 | BRIEF OVERVIEW OF CLINICAL STUDIES..... | 4 |
| 1.3 | MAJOR STATISTICAL ISSUES AND FINDINGS | 4 |
| 2. | INTRODUCTION | 5 |
| 2.1 | OVERVIEW AND BACKGROUND | 5 |
| 2.2 | DATA SOURCES..... | 7 |
| 3. | STATISTICAL EVALUATION | 7 |
| 3.1 | EVALUATION OF EFFICACY..... | 7 |
| 3.1.1 | Study Design and Endpoints..... | 8 |
| 3.1.2 | Patient Disposition, Demographic and Baseline Characteristics | 11 |
| 3.1.3 | Statistical Methodologies..... | 13 |
| 3.1.4 | Results and Conclusions | 15 |
| | 3.1.4.1 Co-Primary Efficacy Endpoints..... | 15 |
| | 3.1.4.2 Discussion of Results of Co-Primary Efficacy Endpoints..... | 16 |
| | 3.1.4.3 Secondary Endpoints | 20 |
| | 3.1.4.4 Other Results | 21 |
| | 3.1.4.5 Conclusions..... | 21 |
| 3.2 | EVALUATION OF SAFETY | 22 |
| 3.3 | GENDER, RACE, AGE AND OTHER SPECIAL/SUBGROUP POPULATIONS..... | 23 |
| 3.3.1 | Gender | 23 |
| 3.3.2 | Race | 23 |
| 3.3.3 | Age..... | 24 |
| 3.3.4 | Baseline Evaluator Wrinkle Severity Assessment..... | 24 |
| 3.3.5 | Baseline Subject Wrinkle Assessment..... | 24 |
| 3.3.6 | Baseline Wrinkle Length | 24 |
| 4. | SUMMARY AND CONCLUSIONS | 25 |
| 4.1 | STATISTICAL ISSUES AND COLLECTIVE EVIDENCE | 25 |
| 4.2 | CONCLUSIONS AND RECOMMENDATIONS..... | 25 |
| APPENDICES..... | | 27 |
| DISTRIBUTION LIST..... | | ERROR! BOOKMARK NOT DEFINED. |

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Results of two pivotal studies (IT-R-005 and IT-R-006) were submitted in support of the efficacy and safety claim of Isolagen Therapy (IT) for the treatment of moderate to severe nasolabial fold wrinkles in adults 18 years of age or older.

In agreement with the FDA, the co-primary efficacy endpoints are:

- a. Proportion of patients with at least 2-point improvement from baseline to 6-month post-treatment on both sides of face in Evaluator Wrinkle Severity Assessment.
- b. Proportion of patients with at least 2-point improvement from baseline to 6-month post-treatment in Subject Wrinkle Assessment.

Each study is declared as a success if IT is shown to be superior to vehicle with respect to each of the co-primary efficacy endpoints. Results of the co-primary efficacy endpoints based on the intent-to-treat (ITT) population (subjects with missing data were treated as treatment failure) are summarized in the following:

| No. of subjects (%) | Study IT-R-005 | | | Study IT-R-006 | | |
|---------------------------------|-----------------|----------------------|----------|-----------------|----------------------|----------|
| | IT (n = 100) | Vehicle (n = 103) | p-value | IT (n = 110) | Vehicle (n = 108) | p-value |
| Evaluator Wrinkle Assessment | 33 (33%) | 7 (7%) | < 0.0001 | 21 (19%) | 8 (7%) | 0.0075 |
| Subject Wrinkle Assessment | 57 (57%) | 31 (30%) | 0.0001 | 50 (45%) | 19 (18%) | < 0.0001 |

In summary, IT is statistically superior to vehicle with respect to each co-primary efficacy endpoints.

- The success rates in Evaluator Wrinkle Severity Assessment were 33% for IT vs. 7% for vehicle in study IT-R-005; and 19% for IT vs. 7% for vehicle in study IT-R-006.
- The success rates in Subject Wrinkle Assessment were 57% for IT vs. 30% for vehicle in study IT-R-005; and 45% for IT vs. 18% for vehicle in study IT-R-006.

Although there are a number of issues with respect to the subgroup analyses presented in Section 1.3, the overall results of the co-primary efficacy endpoints are generally robust over different statistical methods for data analyses. Since missing data rates ranged from 9% to 20% for treatment groups in the two studies, results were impacted in terms of statistical significance in particular under the worst case scenario for missing data handling, where the worst case scenario imputes missing in IT as failures and missing in vehicle as successes. However, the efficacy trend favors IT numerically in this scenario. Additionally, results of the secondary efficacy endpoints support outcomes of the co-primary efficacy endpoints. Note that only efficacy data up to 6-month post-treatment are evaluated, long-term efficacy beyond 6-month post-treatment has not been established.

For possible labeling inclusion, though the sponsor's pre-specified hierarchical testing of Subject Improvement Assessment, Evaluator Improvement Assessment and time-to-success on the co-primary endpoints in the Statistical Analysis Plan (dated 6/17/2008) satisfied statistical principle; the endpoints to be included in the labeling should be clinically relevant to the proposed indication at a minimum. Medical Division's input is essential regarding labeling inclusion.

For safety evaluation, two treatment groups are generally comparable with respect to the overall incidence rate (per patient basis) of treatment emergent adverse events (TEAEs) in both pivotal trials regardless of relationship to the study treatments. The overall incidence rates were 61% vs. 62% for IT vs. vehicle in study IT-R-005; and 63% vs. 57% in study IT-R-006. Most TEAEs were injection site reactions and were mild to moderate in severity. Medical Division should provide more in-depth safety review from clinical perspective.

1.2 Brief Overview of Clinical Studies

To demonstrate the safety and efficacy of IT in the treatment of moderate to severe nasolabial fold wrinkles, results of a total of seven studies (IT-R-001, IT-R-002, IT-R-003A, IT-R-003B, IT-R-005, IT-R-006, and IT-R-007) were submitted in Biologics Licensure Application (BLA). All seven studies were conducted in the US under an IND. Studies IT-R-005 and IT-R-006 are designated as pivotal trials and are the basis of evidence to support the approval of the product in the proposed indication. Others are supportive as IT dose and frequency in the studies as well as the indications treated were different to some extent. The focus of this memo is review of results from Study IT-R-005 and Study IT-R-006.

Studies IT-R-005 and IT-R-006 were identically designed as randomized, double-blind, vehicle-controlled, and multicenter. Both studies included primary evaluation phase (6-month post treatment) and patients would be followed up to 12 months. The primary evaluation phase of the studies was conducted at 7 and 6 US sites during 10/23/06 – 6/26/08 and 11/1/06 – 6/9/08, respectively. A total of 200 patients were pre-planned for each study. Totals of 203 and 218 patients were actually randomized in Studies IT-R-005 and IT-R-006, respectively. The randomization allocation ratio of IT to vehicle was 1:1. This resulted in 100 and 103 patients in IT and vehicle group, respectively, for Study IT-R-005; while 110 and 108 patients were in the respective group for Study IT-R-006. Each patient was to be treated with 3 treatments of either IT or vehicle at 5±1-week intervals at Visit 1, 2 and 3. Clinical evaluation and subject assessment were scheduled at Treatment 3 (Visit 3), 2- (Visit 4), 4- (Visit 5), and 6-month (Visit 6) post Treatment 3 with Visit 6 as the primary time point for efficacy evaluation.

1.3 Major Statistical Issues and Findings

Though the overall efficacy results of IT are superior to vehicle with respect to the two co-primary efficacy endpoints, the following issues are observed:

- In study IT-R-006, sites 6100, 6300 and 6600 had considerably lower success rate for IT treatment in Evaluator Wrinkle Severity Assessment than other sites. The success rates were 5.3% (1/19), 4.5% (1/22) and 10% (2/20) in the respective sites. Following the examination of relevant information, evaluator assessment may be a potential factor.
- The evidence of IT efficacy is limited for male and non-white subjects due to under-represented subgroups.

- Only a total of 5 subjects (1%) aged ≤ 34 years old; and a total of 25 subjects (6%) aged ≤ 40 years old were randomized in the pivotal trials. Although the efficacy trend favors IT, the evidence of IT efficacy is limited in this age group. Consequently, the proposed lower limit of age of 18 years old for the product may be an issue.
- Observed numerically reverse efficacy trend in different co-primary endpoints for the two studies in subjects aged 65 and older. These subjects accounted for 17% of study population.
- As only efficacy data up to 6-month post-treatment are evaluated, long-term efficacy beyond 6-month post-treatment has not been established.

2. INTRODUCTION

2.1 Overview and Background

The proposed product is azfibrocel-T, previously as Isolagen Therapy (IT). The seeking indication is the treatment of moderate to severe nasolabial fold wrinkles in adults 18 years of age or older. This memo will use IT to denote the proposed product throughout the entire review.

IT is a suspension of autologous fibroblasts, grown from a biopsy of each individual's own skin using the standard tissue culture procedures and sponsor's manufacturing process. The product is supplied as two, 2 mL vials of cells at a concentration of $1.0\text{-}2.0 \times 10^7$ cells/mL. The proposed dose frequency is 3-treatment at 5 ± 1 -week intervals. The route of product administration is intradermal injection into the nasolabial fold wrinkles at a dose of up to 2 mL at 0.1 mL/per linear cm.

Prior to FDA regulation of somatic cell therapies, IT was marketed in the US as a cosmetic treatment for facial contour deformities from December 1995 to February 1999. According to the BLA, approximately 1,100 subjects were treated with IT during 1995 and 1999 in the US, mainly for facial contour deformities. IT was also marketed as a cosmetic treatment for facial wrinkles in the United Kingdom (UK) and Europe from 1995 until January 2007.

The IND of IT was initially filed with the FDA on 10/12/1999. The following gives a chronological summary of key correspondence between the sponsor and the FDA regarding the IND:

- The IND was placed on clinical hold on 12/9/1999 due to chemistry, manufacturing, and controls (CMC) issues.
- Clinical hold was lifted on 5/3/2002 following satisfactory responses of CMC issues in submissions 002 and 003 (submitted on 4/5/2002 and 5/2/2002, respectively) and a t-con (4/29/2002).
- FDA Tissue Reference Group confirmed that IT must be regulated as a biologic product (8/29/2002).
- A guidance meeting was held on 4/9/2003. Meeting notes dated 5/8/2003 was provided to the sponsor, including the recommendation of 6-month as the primary time-point for clinical studies that support a BLA.
- FDA on 10/28/2003 provided comments on Study IT-R-002 protocol for the treatment of contour deformities.
- A meeting was held on 12/18/2003 regarding some preliminary Phase 2 data and the design of Study IT-R-002. Meeting notes dated 1/16/2004 was provided to the sponsor, including the inadequacy of study design of Study IT-R-002.
- FDA on 5/21/2004 provided comments on Study IT-R-003 for the treatment of contour deformities. FDA on 1/26/2005 provided comments on revised Study IT-R-003.

Table 1: Overview of the Seven Clinical Studies Conducted under IND

| Study (duration) | No. of sites | No. of subjects Treatment randomized | Study Design | Dosing Range and Frequency | Study Indication |
|---|--------------|---|---|--|-----------------------------|
| IT-R-001 (2/03 – 4/05) | 2 | N = 41 0.5x10 ⁷ : 10 1.0x10 ⁷ : 10 2.0x10 ⁷ : 10 Vehicle: 11 | Randomized, double-blind, vehicle-controlled; long-term phase till 11-month post treatment | Up to 1.2 mL – 0.5, 1.0, and 2.0x10 ⁷ cells/mL 3 treatments every 2-wk | Facial rhytids |
| IT-R-002 (5/03 – 5/05) | 10 | N = 158 2.0x10 ⁷ : 116 Vehicle: 42 | Randomized, double-blind, vehicle-controlled; long-term phase till 12-month post 1 st IT treatment | Up to 2.0 mL of 2.0x10 ⁷ cells/mL 3 treatments every 2-wk | Rhytids and facial scars |
| IT-R-003A (7/04 – 12/05) | 3 | N = 123 2.0x10 ⁷ : 61 Vehicle: 62 | Randomized, double-blind, vehicle-controlled at acute phase; long-term phase till 12-month visit | Up to 1.2 mL of 2.0x10 ⁷ cells/mL, 0.1 mL/linear cm 3 treatments every 7-14 days | Contour deformities |
| IT-R-003B (7/04 – 11/05) | 3 | N = 115 2.0x10 ⁷ : 58 Vehicle: 57 | Randomized, double-blind, vehicle-controlled at acute phase; long-term phase till 12-month visit | Up to 1.2 mL of 2.0x10 ⁷ cells/mL, 0.1 mL/linear cm 3 treatments every 7-14 days | Contour deformities |
| IT-R-005 (10/06 – 6/08) | 7 | N = 203 1.0-2.0x10 ⁷ : 100 Vehicle: 103 | Randomized, double-blind, vehicle-controlled | Up to 2.0 mL of 1.0-2.0x10 ⁷ cells/mL, 0.1 mL/linear cm 3 treatments every 5-wk | Nasolabial fold wrinkles |
| IT-R-006 (11/06 – 6/08) | 6 | N = 218 1.0-2.0x10 ⁷ : 110 Vehicle: 108 | Randomized, double-blind, vehicle-controlled | Up to 2.0 mL of 1.0-2.0x10 ⁷ cells/mL, 0.1 mL/linear cm 3 treatments every 5-wk | Nasolabial fold wrinkles |
| IT-R-007 (3/07 – 6/08) | 5 | N = 50 | Open-labeled, uncontrolled | Up to 6.0 mL of 1.0-2.0x10 ⁷ cells/mL, 0.05 mL/linear cm 2 treatments every 5-wk | Facial wrinkles and creases |
| Source: Sponsor's submission – Module 2, Section 2.5, p.16 and Section 2.7.6, p.10-46. | | | | | |

2.2 Data Sources

Data sources include paper submission (Module 1, Volume 1; Module 2, Volume 1; and Module 5, Volumes 1-66 dated 3/6/2009), electronic patient line listings in SAS transport file format and electronic patient case report forms (CRF) located in electronic document room (EDR); and sponsor's responses (dated 8/7/2009) to FDA's May 19, 2009 letter.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

Results of studies IT-R-005 and IT-R-006 (denoted as 005 and 006 hereafter) are the focus in the evaluation of efficacy of the proposed product Isolagen Therapy (IT). The study protocols were reviewed by FDA under the Special Protocol Assessment (SPA) requests; and the Statistical

Analysis Plan (SAP) was reviewed in subsequent submissions. The statistical reviewer of the protocols and SAP at the IND stage was Dr. Boguang Zhen.

3.1.1 Study Design and Endpoints

Study Design

Studies 005 and 006 were identically designed as randomized, double-blind, vehicle-controlled and multicenter. Both studies included primary evaluation phase (up to 6-month post treatment) and patients would be followed up to 12 months. The primary evaluation phase of the studies was conducted during 10/23/06 – 6/26/08 and 11/1/06 – 6/9/08 at 7 and 6 US sites, respectively. The enrolled patient population included those who were 23 years of age or older and had moderate to severe bilateral nasolabial fold wrinkles based on Evaluator Wrinkle Severity Assessment (i.e., grade of 3 or greater on a 6-point scale) at screening/baseline visit. Wrinkle on each side of face was ≤ 10 cm in length, for a total treatment area of ≤ 20 cm.

A total of 200 patients were pre-planned for each study. Each study (i.e., 100/per arm) was sized to detect a minimal difference of 20% treatment effect at two-sided significance level of 0.05 with a power of 80%, assuming the success rates in IT and vehicle of $\geq 40\%$ and $< 20\%$, respectively. Totals of 203 and 218 patients were actually randomized in studies 005 and 006, respectively. The randomization allocation ratio of IT to vehicle was 1:1. This resulted in 100 and 103 patients in IT and vehicle group, respectively, for study 005; while 110 and 108 patients were in the respective group for study 006. Each patient was to be treated with 3 treatments of either IT or vehicle at 5+1-week intervals at Visit 1, 2, and 3. Clinical evaluation and subject assessment were scheduled at Treatment 3 (Visit 3), 2-month (Visit 4), 4-month (Visit 5), and 6-month (Visit 6) post Treatment 3 with 6-month as the primary time point for efficacy evaluation.

Randomization

Since IT treatment is for autologous use, each patient must undergo three 3-mm² punch post-auricular skin biopsies in order to produce autologous fibroblasts. As a result, patients were only randomized upon biopsy acceptance by IT Quality Control (QC) representatives in the studies. The acceptance criteria pre-specified in the protocols were:

- Shipping package was correctly packed, shipped within 24 hours of biopsy and received intact.
- Vial is intact inside shipper and contents did not leak.
- Visual inspection of the vial does not suggest contamination.
- Biopsies are in the fluid of the vial and have not apparently been frozen or dried out.
- Information about patient is included and forms are correctly filled out, label on vial is correctly filled out with complete information about subject.
- Biopsy appears to generally conform to biopsy technique (3 small superficial punches, though there may be more pieces if a scalpel was necessary to remove the punch).

Once the skin biopsy was determined to be acceptable, it was then passed to IT Quality Assurance (QA) representatives who randomized the subject per the next randomization slot on the site's list prior to providing the biopsies to Manufacturing for production according to treatment assignment.

For subjects whose original biopsy was not acceptable, re-biopsy is permitted. A total of 6 (3%) and 2 (0.9%) subjects had re-biopsy in study 005 and 006, respectively, prior to treatment assignment. All but one subject had re-biopsy due to shipping errors. The remaining subject had re-biopsy due to vial not labeled. The biopsy and injection information are presented in Table A.1 of the Appendix.

Sponsor’s randomization utilized randomization block scheme and was stratified by study site. A total of 60 randomization codes were generated for each site prior to initiation of the trials. Patients were randomized to receive IT and vehicle in a 1:1 allocation ratio within each site. Some errors occurred in treatment assignment during the trials. They are presented in the table below.

| Study | Subjects Involved | Treatment actually received | Randomization (Code) Originally planned | Biopsy date (re-biopsied if 2 nd date) | Biopsy shipping date |
|--|-------------------|-----------------------------|---|---|----------------------|
| 005 | -(b)(6)- | Vehicle | Vehicle (516) | 12/21/06; 1/2/07 | 1/2/07 |
| | -(b)(6)- | IT | Vehicle (517) | 12/21/06; 1/4/07 | 1/4/07 |
| | -(b)(6)- | Vehicle | IT (518) | 12/21/06; 1/16/07 | 1/16/07 |
| | -(b)(6)- | Vehicle | IT (519) | 1/25/07 | 1/25/07 |
| | -(b)(6)- | Vehicle | Vehicle (609) | 12/21/06 | 12/21/06 |
| | -(b)(6)- | Vehicle | IT (610) | 12/21/06 | 12/21/06 |
| | -(b)(6)- | IT | Vehicle (611) | 12/21/06 | 12/21/06 |
| | -(b)(6)- | IT | IT (612) | 12/21/06 | 12/21/06 |
| 006 | -(b)(6)- | Vehicle | Vehicle (213) | 1/2/07 | 1/2/07 |
| | -(b)(6)- | Vehicle | Vehicle (214) | 1/4/07 | 1/4/07 |
| Source: Sponsor’s electronic SAS datasets and sponsor’s original randomization lists. | | | | | |

The summary is:

- For study 005, eight consecutive randomization codes (516-523) for site 5500 were skipped and four of them (516-519) were used by site 5600 (i.e., subjects -----(b)(6)----- . The codes of 516-519 were planned to be used for subjects -----(b)(6)----- who were the last four subjects randomized at site 5500. On the other hand, the randomization list at site 5600 for the remaining subjects after -(b)(6)- resulted in 15 and 16 in IT and vehicle, respectively, that are identical to the numbers of subjects actually received the respective treatment in the current trial. As a result, study 005 has 1 less IT subject than what was planned.
- For study 006, one randomization code 213, Vehicle, for site 6200 was used twice – subjects -(b)(6)-- and -(b)(6)--. Subject -(b)(6)-- was planned to be assigned with code 214 of Vehicle. Because of the error, totals of 6 and 9 subjects were treated with IT and vehicle, respectively, after subject -(b)(6)-. On the other hand, if no error of treatment assignment had occurred at site 6200, subjects treated with IT and vehicle after -(b)(6)-- would have been 7 and 8, respectively. As a result, study 006 has 1 less IT subject than what was planned.

Consequently, the impact of errors on the balance of treatment allocation is minimal in both trials.

Blinding

The studies were designed as double-blind. The following procedures were used to maintain study blinding:

- Each subject was assigned with two investigators, one as the Evaluator and the other as the Injector. The Evaluator was to determine the subject’s treatment area and to perform assessments of wrinkle severity. The Injector was to administer study treatment. Both investigators were instructed not to discuss patients’ wrinkle severity and treatment assignment.
- As skin biopsy is necessary to grow cells in IT treatment, subjects were paired (IT and vehicle) to be re-biopsied or to be withdrawn from studies if the manufacturing process could not produce sufficient product for subjects assigned to IT group.
- Data of blinding assessment were collected per FDA’s request. However, no statistical method was proposed or specified for analyses. Subjects were given a pre-printed postcard at Visit 1, 2 and 3 (i.e., Treatment 1, 2, and 3) to record their knowledge as to which treatment they believed

they had received. They were requested to mail back the postcard prior to the next visit. Evaluators were asked to record their opinion of which treatment each subject had received at Visit 3, 4 and 5.

A total of 1 and 4 subjects in study 005 and 006, respectively, had re-biopsy post randomization. The re-biopsy and injection information are presented in Table A.2 of the Appendix. Only one re-biopsy post-randomization was related to maintain study blinding (subject -(b)(6)-). Such an action might somehow unblind the patients. The impact however is minimal from this reviewer’s point of view, as the rate in IT group is 2% (= 4/210).

Study Endpoints

Two co-primary efficacy endpoints were pre-specified and agreed upon with the FDA:

- a. Proportion of patients with at least 2-point improvement from baseline to 6-month post treatment on both sides of face in Evaluator Wrinkle Severity Assessment.
- b. Proportion of patients with at least 2-point improvement from baseline to 6-month post treatment in Subject Wrinkle Assessment.

Each study is declared as a success if the superiority of IT to vehicle is established with respect to each of the co-primary efficacy endpoints.

The Subject Wrinkle Assessment is patient’s live evaluation of his/her wrinkles of the lower part of face based on a 5-point scale. The Evaluator Wrinkle Severity Assessment is investigator’s live evaluation of patients’ wrinkles based on a 6-point Lemperle Scale with photoguide. Wrinkle severity on each side of the face was evaluated for each individual in Evaluator Wrinkle Severity Assessment. Descriptions of the scales are in the following:

| Subject Wrinkle Assessment | | Evaluator Wrinkle Severity Assessment | |
|----------------------------|--|---------------------------------------|--|
| Grade | Description | Grade | Description |
| -2 | I am very dissatisfied with the wrinkles of the lower part of my face | 0 | No wrinkle visible |
| -1 | I am dissatisfied with the wrinkles of the lower part of my face | 1 | Just perceptible wrinkle |
| 0 | I am somewhat satisfied with the wrinkles of the lower part of my face | 2 | Shallow wrinkle |
| +1 | I am satisfied with the wrinkles of the lower part of my face | 3 | Moderately deep wrinkle (definite and distinct wrinkle) |
| +2 | I am very satisfied with the wrinkles of the lower part of my face | 4 | Deep wrinkle, well-defined edge (prominent wrinkle, well-defined edge) |
| | | 5 | Very deep wrinkle, redundant fold (very severe wrinkle, pronounced edge) |

Numerous secondary efficacy endpoints were proposed in study protocols. For possible inclusion of the secondary efficacy endpoints in the labeling, sponsor’s SAP proposed a hierarchical testing of endpoints that are stated in Section 3.1.3 Statistical Methodologies of the review.

The secondary endpoints pre-specified in the protocols were:

- Success rate in live Subject Wrinkle Assessment at Visits 3, 4 and 5, with a success of an individual defined as a 2-point or better improvement.

- Success rate in live Evaluator Wrinkle Severity Assessment at Visits 3, 4 and 5, with a success of an individual defined as a 2-point or better improvement in wrinkle severity on both sides of the face.
- Subject Improvement Assessment based on photos performed at Visit 6, with a response defined as better (+1) or much better (+2) improvement than before.
- Evaluator Improvement Assessment based on photos performed at Visit 6, with a response defined as better (+1) or much better (+2) improvement than before.

The Subject Improvement Assessment was patient’s self assessment that evaluated the post-treatment change in the wrinkles of the lower part of the face, using a 5-point improvement scale, comparing the photographs obtained at Visits 3 through 6 to the Baseline photographs. The Evaluator Improvement Assessment was investigator’s assessment that evaluated the post-treatment change in the bilateral nasolabial fold wrinkles using a 5-point improvement scale, comparing the patient photographs taken at Visits 3 through 6 to the Baseline photographs. Descriptions of the scales are shown below.

| Subject Improvement Assessment | | Evaluator Improvement Assessment | |
|--------------------------------|---------------------------------------|----------------------------------|------------------------------------|
| Grade | Description | Grade | Description |
| -2 | Appearance is much worse than before | -2 | Wrinkle is much worse than before |
| -1 | Appearance is worse than before | -1 | Wrinkle is worse than before |
| 0 | Appearance is the same as before | 0 | Wrinkle is the same as before |
| +1 | Appearance is better than before | +1 | Wrinkle is better than before |
| +2 | Appearance is much better than before | +2 | Wrinkle is much better than before |

Sponsor’s safety parameters included treatment exposure, vital signs and incidence of adverse events throughout each study.

Population Proposed and Analyzed in Protocols and BLA

The analysis populations included in the study protocols were intent-to-treat (ITT), Efficacy Evaluable (EE) and safety populations. An addition of modified intent-to-treat (MITT) population in data analyses was agreed upon with the FDA at a t-con held on 7/18/2008. The primary efficacy analysis is the ITT analyses; analyses based on EE and MITT populations are supportive. Safety analysis is based on the safety population. Definitions of the analysis populations are:

- ITT: all patients who were randomized.
- MITT: patients who were randomized and received at least one treatment.
- EE: patients who were randomized and received three injections of the study treatment, met inclusion/exclusion criteria, and did not have a major protocol deviation.
- Safety: Patients who received at least one treatment.

3.1.2 Patient Disposition, Demographic and Baseline Characteristics

To make a proper comparison between treatment groups in efficacy, patient disposition and the comparability of treatment groups at baseline are evaluated. Table 2 presents the patient disposition for studies 005 and 006. Results of study enrollment by site and patient demographic/baseline characteristics based on the ITT population are given in Tables A.3-A.5 of the Appendix.

As observed from Table 2, the study completion rates were 80% vs. 85% for IT vs. vehicle in study 005; and 85% vs. 91% in study 006. Approximately 76% and 87% in the respective group received all three treatments for study 005; and they were 78% and 90% for study 006.

Higher rates of patients in IT group had major protocol deviation as compared to vehicle (22% vs. 16% for IT vs. vehicle in study 005; and 19% vs. 5% in study 006). The most frequent major protocol deviation was the patient visit outside the specified protocol windows. IT group had higher rates of patients discontinued from each study (20% vs. 15% for IT vs. vehicle in study 005; and 15% vs. 9% in study 006). Majority of patients discontinued from each study due to subject withdrawal and sponsor request prior to treatment administration. As a result, IT group had lower rates of randomized patients included in the MITT, EE and safety populations.

For the category of “*Sponsor Request*” in Table 2, all patients but two (ID: -(b)(6)- and -(b)(6)-) were discontinued because study product could not be manufactured. None of these patients received any dose of treatment and consequently they were not followed. The two patients -(b)(6)- and -(b)(6)- in the category were discontinued due to misunderstanding of the exclusion criterion by the investigator and belligerent behavior, respectively.

Note that all 6 patients discontinued from study 005 due to “*Others*” reason were all from site 5300. Among them, 3 and 2 were in IT and vehicle group, respectively, had basal cell carcinoma (BCC) diagnosis at baseline visits (ID:------(b)(6)-----). Patient -(b)(6)- was discontinued because of patient’s history of a prolactin secreting pituitary tumor.

Table 2: Patient Disposition – Studies 005 and 006

| No. of subjects (%) | Study 005 (n = 203) | | Study 006 (n = 218) | |
|---|---------------------|------------|---------------------|------------|
| | IT | Vehicle | IT | Vehicle |
| Subject Randomized | 100 | 103 | 110 | 108 |
| Completed Study | 80 (80%) | 88 (85%) | 93 (85%) | 98 (91%) |
| Received all 3 treatments | 76 (76%) | 90 (87%) | 86 (78%) | 97 (90%) |
| Discontinued, total | 20 (20%) | 15 (15%) | 17 (15%) | 10 (9%) |
| Subject withdrawal | 6 | 6 | 3 | 4 |
| Sponsor request | 5 | 1 | 10 | 4 |
| Adverse event ¹ | 2 | 1 | 1 | 1 |
| Non-compliance with protocol | 3 | 1 | 2 | 1 |
| Lost to follow-up | 1 | 3 | 0 | 0 |
| Others ² | 3 | 3 | 1 | 0 |
| Subjects with one or more major protocol deviation | 22 (22%) | 16 (16%) | 21 (19%) | 5 (5%) |
| ITT population | 100 (100%) | 103 (100%) | 110 (100%) | 108 (100%) |
| MITT population | 83 (83%) | 92 (89%) | 98 (89%) | 99 (92%) |
| EE population | 60 (60%) | 76 (74%) | 66 (60%) | 88 (81%) |
| Safety population | 83 (83%) | 92 (89%) | 98 (89%) | 99 (92%) |
| Source: Table 5 on p.52 and Table 7 on p.55, Vol.43; Table 5 on p.51 and Table 7 on p.55, Vol.52; and SAS electronic datasets in EDR. | | | | |
| ¹ “ <i>Adverse event</i> ” included deaths – one death -(b)(6)- in vehicle group for study 005 due to heart attack after receiving three treatments; one death -(b)(6)- in vehicle group for study 006 due to cardiac arrest prior to treatment. | | | | |
| ² “ <i>Others</i> ” included history and diagnosis of basal cell carcinoma (BCC), history of prolactin secreting tumor and relocation. | | | | |

Patient enrollment is generally comparable between groups within each site for each study (Table A.3 of the Appendix). No single site had enrolled more than 25% of the study population. It however should be noted that sites 5100, 5300 and 5600 randomized a total of 133 patients which

accounted for 65.5% of study population in trial 005. The primary investigators in these sites had participated in other studies conducted under the same IND. Specifically, the primary investigator at site 5100 participated in earlier studies IT-R-001, IT-R-002 and IT-R-003A; the primary investigator at site 5300 participated in earlier studies IT-R-002 and IT-R-003A; and the primary investigator at site 5600 participated in study IT-R-007, where the timeline of study IT-R-007 was overlapped with that of the pivotal studies. Although the role of these investigators is stated in patient CRFs, it may nevertheless impact the efficacy results (e.g., unintentionally knew what treatments patients received due to experiences from earlier trials). Similarly, the primary investigator at site 6400 in study 006 participated in earlier studies IT-R-002 and IT-R-003B, and the site enrollment accounted for 16% of study population. This issue will be investigated further in Section 3.1.4 of the review.

For the treatment distribution by demographic and baseline characteristics, results in Tables A.4-A.5 generally suggest non-significant difference between groups for studies 005 and 006, respectively. Study 005 randomized patients whose age ranged between 35 and 78 with mean age of approximately 56.5 years old; while study 006 randomized patients aged between 23 and 81 with mean age of approximately 54.5 years old. About 90% of patients are women. Patients of white race accounted for about 95% and 89% of study population in trials 005 and 006, respectively. The average wrinkle length was comparable between groups within each study ($p= 0.4720$ and $p = 0.5818$, respectively) with wrinkle length ranged from 6.2 to 18.0 cm for study 005 and 4.0 to 17.0 cm for study 006. All patients had dissatisfied (-1) or very dissatisfied (-2) scores in Subject Wrinkle Assessment and were graded 3 and above on both sides of face in Evaluator Wrinkle Severity Assessment at screening/baseline visits.

As mentioned earlier in Table 2 regarding the category of “*Sponsor Request*”, all patients but two were discontinued from trials because study product could not be manufactured which was manufacturing failure. Manufacturing Failure occurred when subject’s skin biopsy failed to produce product or did not produce sufficient product for three treatments. Subjects whose skin biopsy did not produce products were treated as treatment failures in efficacy analyses; while subjects whose skin biopsy did not produce sufficient product were followed and evaluated for efficacy at 6-month. Results of the manufacturing failure are presented in the table below. The summary is:

- The IT manufacturing failure rate of not producing product is about 6% (= 13/210).
- The IT manufacturing failure rate of not producing sufficient product is about 5% (= 11/210).
- Consequently, the total IT manufacturing failure rate in the two pivotal trials is 11% (= 24/210).

| Manufacturing Failure | Type of Failure | Study 005 | | Study 006 | |
|-----------------------|-----------------------------|-----------------|----------------------|-----------------|----------------------|
| | | IT (n = 100) | Vehicle (n = 103) | IT (n = 110) | Vehicle (n = 108) |
| | No Product Produced, Total | 5 (5%) | 1 (0.97%) | 8 (7%) | 4 (3.7%) |
| | Insufficient Product, Total | 3 (3%) | 0 | 8 (7.3%) | 1 (0.93%) |
| | Received 1x | 1 (1%) | 0 | 4 (3.6%) | 0 |
| | Received 2x | 2 (2%) | 0 | 4 (3.6%) | 1 (0.93%) |

Source: Sponsor’s BLA submission (Module 5, Volume 46).

Reviewer’s Comments on Patient Disposition:

On page 49, Volume 52, the sponsor states that one subject -(b)(6)-(assigned to vehicle) was withdrawn prior to treatments due to a misunderstanding by the investigator of the exclusion criteria. Following reviewing SAS electronic data sets and CRF, subject -(b)(6)- had all three treatments administered and completed the

study. The referred subject should be -(b)(6)- (assigned to IT) who was discontinued per sponsor request and was categorized as “*Sponsor Request*” in Table 2.

3.1.3 Statistical Methodologies

The main statistical methods for data analyses in BLA submission were pre-specified in the protocols and Statistical Analysis Plan. They are summarized in the following:

- Each of the co-primary efficacy endpoint was tested for superiority of IT to vehicle using Cochran-Mantel-Haenszel (CMH) test stratified by site as the primary.
- Repeated measures analysis to incorporate data over visits with the assumption that data are missing at random (MAR) is performed as a sensitivity analysis of the co-primary efficacy endpoints.
- Secondary efficacy endpoints were analyzed using the primary analysis rule.
- Comparison of groups in time-to-success for each co-primary efficacy endpoint was based on log-rank test, where time-to-success evaluates the time to sustained response of at least 2-point improvement from baseline.
- Efficacy analysis based on the ITT population was the primary; and analyses based on the MITT and EE populations were secondary.
- For handling missing data in the efficacy analyses, imputation with treatment failure was the primary. Sensitivity analyses that evaluate the robustness of study results due to missing data included:
 - Treating missing as successes.
 - Worst case imputation – treating missing in vehicle group as successes and missing in IT group as failures.
 - Last observation carried forward (LOCF) method (i.e., carried forward to 6-month if there were data at 2- and 4-month post-treatment; otherwise, treatment failure at 6-month post-treatment).
- To preserve the possible inclusion of efficacy endpoints in the labeling, the sponsor stated in SAP (dated 6/17/2008) that the analyses were to be considered for hierarchical testing according to the order outlined in Table 3 below:

Table 3: Hierarchy Testing of Endpoints from SAP

| Order | Endpoint/Analysis | Type of Endpoint |
|-------|---|--------------------|
| 1 | Successful improvement in live Subject Wrinkle Assessment at Visit 6 (2 or more points better than Baseline) | Primary Efficacy |
| 2 | Successful improvement in live Evaluator Wrinkle Severity Assessment at Visit 6 (2 or more points better than Baseline on both sides of face) | Primary Efficacy |
| 3 | Success on Subject Improvement Assessment at Visit 6 (a score of +1 or +2) based on photos | Secondary Efficacy |
| 4 | Success on Evaluator Improvement Assessment at Visit 6 (a score of +1 or +2 on both sides of face) based on photos | Secondary Efficacy |
| 5 | Time-to-first improvement (2 or more points) on the Subject Wrinkle Assessment relative to Baseline | Secondary Efficacy |
| 6 | Time-to-first improvement (2 or more points) on the Evaluator Wrinkle Severity Assessment relative to Baseline | Secondary Efficacy |

Reviewer’s Comments on Hierarchical Testing of Endpoints:

1. Both primary efficacy endpoints must meet the objectives to declare a study success. There is no need testing the co-primary efficacy endpoints in a hierarchical order.

2. Though the hierarchical testing of endpoints was pre-specified in the SAP and satisfied the statistical principle, the endpoints to be included in the labeling should be clinically relevant to the proposed indication at a minimum. It should be noted that time-to-first improvement endpoints on the Subject Wrinkle Assessment and on the Evaluator Wrinkle Severity Assessment were included in the SAP for possible labeling inclusions; while the secondary endpoints specified in the protocol, success in live Subject Wrinkle Assessment and success in live Evaluator Wrinkle Severity Assessment at intermediate visits (i.e., Visit 3, 4 and 5) were not included for labeling inclusion. Medical Division’s input is essential regarding the inclusion of these endpoints in the labeling.

3.1.4 Results and Conclusions

To demonstrate the efficacy of IT, results of the co-primary efficacy endpoints, secondary endpoints, and additional analyses based on the ITT, MITT and EE populations are included in the sponsor’s BLA. As conclusions are in agreement, this review presents mostly the ITT outcomes, the primary analysis.

3.1.4.1 Co-Primary Efficacy Endpoints

Results of the co-primary efficacy endpoints based on the ITT (primary), MITT (secondary) and EE (secondary) populations are presented in Table 4. Missing data in the analyses are treated as treatment failures.

Table 4: Results* of the Co-Primary Efficacy Endpoints – Studies 005 and 006

| No. of subjects (%) | | Study 005 | | | Study 006 | | |
|------------------------------------|----------------------------|------------------------------|---------------|---------------|---------------|---------------|---------------|
| Type of Analysis | Endpoints | ITT | Vehicle | p-value | ITT | Vehicle | p-value |
| Primary (ITT) 005 – (100, 103) | Subject Wrinkle Assessment | 57 (57%) | 31 (30%) | 0.0001 | 50 (45%) | 19 (18%) | < 0.0001 |
| | 006 – (110, 108) | Evaluator Wrinkle Assessment | 33 (33%) | 7 (7%) | < 0.0001 | 21 (19%) | 8 (7%) |
| Secondary (EE) 005 – (60, 76) | | Subject Wrinkle Assessment | 44 (73.3%) | 27 (35.5%) | < 0.0001 | 35 (53.0%) | 17 (19.3%) |
| | 006 – (66, 88) | Evaluator Wrinkle Assessment | 25 (41.7%) | 4 (5.3%) | < 0.0001 | 16 (24.2%) | 7 (8.0%) |
| Secondary (MITT) 005 – (83, 92) | Subject Wrinkle Assessment | 57 (68.7%) | 31 (33.7%) | < 0.0001 | 50 (51.0%) | 19 (19.2%) | < 0.0001 |
| | 006 – (98, 99) | Evaluator Wrinkle Assessment | 33 (39.8%) | 7 (7.6%) | < 0.0001 | 21 (21.4%) | 8 (8.1%) |

Source: Module 5 – Vol.43 (p.69-72, Table 14.2.6.2, Table 14.2.8.3, Table 14.2.11.2, and Table 14.2.13.3); and Vol.52 (p.68-71, Table 14.2.6.2, Table 14.2.8.3, Table 14.2.11.2, and Table 14.2.13.3).
* Missing data are treated as treatment failures in all analyses.

The summary is:

- IT is statistically superior to vehicle with respect to each of the co-primary efficacy endpoints in the primary ITT analysis (p-value ≤ 0.0075). The success rates of IT vs. vehicle in Evaluator Wrinkle Severity Assessment were 33% vs. 7% for study 005; and 19% vs. 7% for study 006.

The success rates in Subject Wrinkle Assessment were 57% vs. 30% for study 005; and 45% vs. 18% for study 006. The observed treatment effect in success rate of Evaluator Wrinkle Severity Assessment is 26% and 12% for studies 005 and 006, respectively; while the observed effect in success rate of Subject Wrinkle Assessment is 27% for each of the two studies.

- Analyses based on EE and MITT populations are in agreement with the ITT analyses that IT is statistically superior to vehicle with respect to each of the co-primary efficacy endpoints.

3.1.4.2 Discussion of Results of Co-Primary Efficacy Endpoints

The robustness of results of the co-primary efficacy endpoints is assessed in terms of different methodologies for data analyses, different methods of handling missing data, and impact of some investigators who had participated in earlier studies, had financial relation with the sponsor; and study sites that had extreme outcomes.

1. Sensitivity Analyses

- Analyses based on different statistical methodology:
Repeated measures analysis based on generalized estimating equation (GEE) method to incorporate data over visits with the assumption that data are missing at random (MAR) has been performed. Results are presented in Table A.6 of the Appendix. Repeated measures analysis supports the superiority of IT to vehicle with respect to at least 2-point improvement from baseline to 6-month post-treatment in Evaluator Wrinkle Severity Assessment and in Subject Wrinkle Assessment.

- Analyses based on different missing data handling:
The pre-specified primary analysis imputes missing with treatment failures and demonstrates that IT is statistically superior to vehicle with respect to the co-primary efficacy endpoints in each study. The missing data rates in IT and vehicle group were 20% vs. 14.6%, respectively, in study 005; while they were 15.5% vs. 9.3% in study 006. The difference is not statistically significant ($p = 0.3548$ in study 005 and $p = 0.2174$ in study 006 based on Fisher's exact test). Because the missing data rates are not small, sensitivity analyses were performed to evaluate the impact of various missing data handling on study results. The pre-specified sensitivity analyses included:

- (1) Treating missing as successes.
- (2) Worst case – treating missing in vehicle as successes and missing in IT as failures.
- (3) LOCF method (i.e., carried forward to 6-month if there were data at 2- and 4-month post-treatment; otherwise, treatment failure at 6-month post-treatment).

Results are presented in Table A.7 of the Appendix. Efficacy conclusion based on different missing data handling is generally in agreement with the primary analysis. Results based on LOCF are similar to the ITT analyses with treatment failures as missing data handling. On the other hand, results based on the worst case scenario fail to demonstrate the superiority of IT to vehicle statistically in some comparisons (see the shaded area in Table A.7). The efficacy trend however is consistent that IT is better than vehicle.

2. Impact of Study Sites on Efficacy Evaluations

- Primary Investigators Participated in Other IT Studies

As stated earlier, the primary investigators at sites 5100, 5300 and 5600 in study 005 and at site 6400 in study 006 had participated in other IT non-pivotal studies under the same IND. The enrollment at sites 5100, 5300 and 5600 accounted for 65.5% of study population in study 005;

while site 6400 accounted for 16% of study population in study 006. To examine whether these sites have impact on the efficacy evaluation, particularly Evaluator Wrinkle Severity Assessment, results of the co-primary efficacy endpoints by site are presented in Table 5.

**Table 5: Results* of Co-Primary Efficacy Endpoints by Site
Studies 005 and 006 (ITT)**

| Study 005 Site | Subject Wrinkle Assessment | | Evaluator Wrinkle Severity Assessment | |
|--|----------------------------|----------------|---------------------------------------|--------------|
| | IT | Vehicle | IT | Vehicle |
| 5100 | 15/25 (60%) | 9/24 (37.5%) | 5/25 (20%) | 1/24 (4.2%) |
| 5200 | 6/10 (60%) | 4/11 (36.4%) | 2/10 (20%) | 1/11 (9.1%) |
| 5300 | 9/21 (42.9%) | 5/20 (25%) | 8/21 (38.1%) | 1/20 (5%) |
| 5400 | 7/7 (100%) | 3/8 (37.5%) | 4/7 (57.1%) | 0/8 |
| 5500 | 5/9 (55.6%) | 3/10 (30%) | 3/9 (33.3%) | 1/10 (10%) |
| 5600 | 12/21 (57.1%) | 6/22 (27.3%) | 7/21 (33.3%) | 1/22 (4.5%) |
| 5700 | 3/7 (42.9%) | 1/8 (12.5%) | 4/7 (57.1%) | 2/8 (25%) |
| Total | 57/100 (57%) | 31/103 (30.1%) | 33/100 (33%) | 7/103 (6.8%) |
| Study 006 Site | IT | Vehicle | IT | Vehicle |
| 6100 | 11/19 (57.9%) | 4/17 (23.5%) | 1/19 (5.3%) | 2/17 (11.8%) |
| 6200 | 7/13 (53.8%) | 2/16 (12.5%) | 5/13 (38.5%) | 2/16 (12.5%) |
| 6300 | 13/22 (59.1%) | 1/22 (4.5%) | 1/22 (4.5%) | 0/22 |
| 6400 | 6/18 (33.3%) | 5/17 (29.4%) | 5/18 (27.8%) | 2/17 (11.8%) |
| 6500 | 7/18 (38.9%) | 3/17 (17.6%) | 7/18 (38.9%) | 2/17 (11.8%) |
| 6600 | 6/20 (30%) | 4/19 (21.1%) | 2/20 (10%) | 0/19 |
| Total | 50/110 (45.5%) | 19/108 (17.6%) | 21/110 (19.1%) | 8/108 (7.4%) |
| Source: Module 5 – Vol.43 (Table 14.2.1.2.2 and Table 14.2.3.3.2) and Vol.52 (Table 14.2.1.2.2 and Table 14.2.3.3.2). | | | | |
| * Missing data are treated as treatment failures. | | | | |

The findings are summarized in the following:

- a. Though the success rates of Evaluator Wrinkle Severity Assessment in vehicle group are similar between studies (6.8% for study 005 and 7.4% for study 006), sites in study 005 generally have higher success rates in IT group as compared to those in study 006. As a result, the overall success rate for IT treatment is 33% in study 005 in contrast to 19.1% in study 006.
- b. For study 006, the success rate of Evaluator Wrinkle Severity Assessment in IT group at site 6400 is not the greatest; but it is better than that at sites 6100, 6300, and 6600. Because study 006 has a relatively lower success rate in Evaluator Wrinkle Severity Assessment as compared to study 005 (19% vs. 33%), a sensitivity analysis excluding site 6400 is performed. IT is statistically superior to vehicle. Results are:
 - o Success rate in Evaluator Wrinkle Severity Assessment: 17% vs. 6% (IT vs. vehicle) with $p = 0.0159$.
 - o Success rate in Subject Wrinkle Assessment: 48% vs. 15% (IT vs. vehicle) with $p < 0.0001$.
- c. For study 005, the impact of the three sites (5100, 5300 and 5600) on efficacy results is not pronounced for the following reasons:

- Success rates of Evaluator Wrinkle Severity Assessment at sites 5100, 5300 and 5600 combined are 30% (= 20/67) and 4.5% (= 3/66) for IT and vehicle group, respectively; while the success rates at other sites are 39.4% (= 13/33) and 10.8% (= 4/37) for the respective group. The success rate in IT group at sites 5100, 5300 and 5600 is not larger than that at other sites (30% vs. 39.4%).
- But success rate of Evaluator Wrinkle Severity Assessment in vehicle group however is smaller in the three sites as compared to others (4.5% vs. 10.8%). A sensitivity analysis excluding the 3 sites is performed. IT is statistically superior to vehicle. Results are:
 - i. Success rates in Evaluator Wrinkle Severity Assessment are 39% vs. 11% (IT vs. vehicle) with $p = 0.0054$.
 - ii. Success rates in Subject Wrinkle Assessment are 64% vs. 30% (IT vs. vehicle) with $p = 0.0042$.

- Investigators had Financial Relation with the Sponsor

Based on the section of Financial Disclosures in sponsor's BLA submission, the primary investigators at sites 5300 and 5600 in study 005; and at site 6600 in study 006 appeared to have financial relation with the sponsor. Analyses, particularly the success rate in Evaluator Wrinkle Severity Assessment, were performed to assess the impact on the efficacy evaluation.

For study 005, the success rate in Evaluator Wrinkle Severity Assessment for sites 5300 and 5600 combined is 35.7% for IT which is greater than 31.0% for the remaining sites combined; while vehicle has a lower success rate 4.8% in sites 5300+5600 than 8.2% for the remaining sites combined. Comparison between IT and vehicle was performed for the remaining sites. Analyses excluding sites 5300 and 5600 demonstrate that IT is statistically superior to vehicle with p -value = 0.0012. The success rates are 31% for IT and 8.2% for vehicle.

For study 006, though the success rate in Evaluator Wrinkle Severity Assessment for site 6600 is 10% for IT which is smaller than 21.1% for the remaining sites combined; the success rate in vehicle group is 0 as compared to 9.0% for the remaining sites. Analyses excluding site 6600 give $p = 0.018$ that IT is statistically superior to vehicle with success rates of 21.1% for IT and 9% for vehicle.

- Study Sites with Extreme Outcomes

Extreme outcomes occurred in some study sites. Sensitivity analyses are conducted to investigate the impact of these study sites. They are summarized in the following:

- a. For study 005, site 5400 had 100% success rate in Subject Wrinkle Assessment for IT group as compared to 37.5% for vehicle group; while the success rates in Evaluator Wrinkle Severity Assessment for IT vs. vehicle were 57.1% vs. 0 (See Table 5). Since the site size is small (7 and 8 in IT and vehicle), the impact of the site is not expected to be great.

A sensitivity analysis excluding site 5400 gives $p = 0.0008$ for success rate in Subject Wrinkle Assessment and $p < 0.0001$ for success rate in Evaluator Wrinkle Severity Assessment. IT is statistically superior to vehicle in the co-primary efficacy endpoints. The success rates of Subject Wrinkle Assessment are 53.8% (=50/93) and 29.5% (=28/95) in the respective group; while the success rates of Evaluator Wrinkle Severity Assessment are 31.2% (=29/93) and 7.4% (=7/95).

- b. For study 006, sites 6100, 6300 and 6600 had considerably lower success rate of Evaluator Wrinkle Severity Assessment in IT group as compared to other sites (see Table 5). The success rates were 5.3%, 4.5% and 10%. Since the three sites represent more than 1/2 of the study enrollment (about 55%), the patient population has been examined in terms of patient age, baseline wrinkle severity, missing data rate, and product injection volume. No outstanding discrepancies as compared to others are noted. Consequently, evaluator assessment could be a potential factor.

To examine the impact of the three sites on the efficacy results, outcome of the co-primary efficacy endpoints are presented in the table below for the 3 sites as compared to the remaining sites in study 006.

| Sites | Evaluator Wrinkle Severity Assessment | | Subject Wrinkle Assessment | |
|--------------------|---------------------------------------|------------|----------------------------|-------------|
| | IT | Vehicle | IT | Vehicle |
| 6100 + 6300 + 6600 | 4/61 (7%) | 2/58 (3%) | 30/61 (49%) | 9/58 (16%) |
| Remaining sites | 17/49 (35%) | 6/50 (12%) | 20/49 (41%) | 10/50 (20%) |

The summary is:

- The observed treatment effect of Evaluator Wrinkle Severity Assessment is 4% for the 3 sites as compared to 23% for remaining sites. The observed treatment effect of Subject Wrinkle Assessment is 33% for the 3 sites as compared to 21% for the remaining sites.
- Sensitivity analyses are performed. Results show that IT is statistically superior to vehicle in success rates of Subject Wrinkle Assessment for the 3 sites as well as for the remaining sites. For the success rate in Evaluator Wrinkle Severity Assessment, IT is superior to vehicle for the remaining sites, but not for the 3 sites alone.

3. Distribution of the Number of Points Changed in Evaluator Wrinkle Severity Assessment and Subject Wrinkle Assessment

Since the co-primary efficacy endpoints are binary, it is of interest to understand the distribution of patient improvement in terms of number of points changed from baseline to 6-month in Evaluator Wrinkle Severity Assessment and Subject Wrinkle Assessment. As Evaluator Wrinkle Severity Assessment evaluated both sides of the face for each patient, results are displayed in a 2-dimensional table. That is, number of points changed on the right side of face (vertical) vs. number of points changed on the left side of face (horizontal). Results are summarized in Tables A.8-A.9 of the Appendix. Note that results on shaded area are outcomes of the co-primary efficacy endpoints. The summary of Tables A.8-A.9 is:

- Most patients who met the primary efficacy endpoint in Subject Wrinkle Assessment had exactly 2-point change from baseline to 6-month (Table A.8). They accounted for about 65% (=37/57) vs. 71% (=22/31) of successful cases for IT vs. vehicle in study 005; while they were 58% (=29/50) vs. 84% (=16/19) in study 006.
- For Evaluator Wrinkle Severity Assessment (Table A.9(a) and (b)),
 - Study 005: Most patients who were successes had exactly 2-point improvement on each side of face. They accounted for about 61% (= 20/33) and 71% (= 5/7) in IT and vehicle, respectively.
 - Study 006: Eleven (11) patients who were successes had exactly 2-point or 3-point improvement on each side of face in IT group; while 5 patients in vehicle group had exactly 2-point improvement on each side of face.

- c. For Evaluator Wrinkle Severity Assessment (Table A.9(a) and (b)), a total of 13 patients in each of IT and vehicle groups had at least 2-point improvement on either side of face and not on both sides for study 005; while there were 13 and 8 patients in the respective group for study 006. These patients however are not successes with respect to the primary efficacy endpoint of Evaluator Wrinkle Severity Assessment.
4. Distribution of Successes in Evaluator Wrinkle Severity Assessment and Subject Wrinkle Assessment

To evaluate the responder to both Evaluator Wrinkle Severity Assessment and Subject Wrinkle Assessment (i.e., successes for both Assessments), results are summarized and presented in Table A.10 of the Appendix. The shaded areas are the ITT results of the co-primary efficacy endpoints.

It is observed from Table A.10 that 30 subjects treated with IT in study 005 had at least 2-point change from baseline to 6-month in Subject Wrinkle Assessment but were evaluated as failures in Evaluator Wrinkle Severity Assessment; while there were 28 such subjects in vehicle group. For study 006, there were 37 and 15 subjects in IT and vehicle group, respectively, had at least 2-point change in Subject Wrinkle Assessment but were evaluated as failures in Evaluator Wrinkle Severity Assessment.

As a result, totals of 27 (27%) and 3 (3%) subjects in IT and vehicle group, respectively, in study 005 were responders in both Evaluator Wrinkle Severity Assessment and Subject Wrinkle Assessment. There were 13 (12%) and 4 (4%) such subjects in the respective group in study 006. The treatment group comparison yields a p-value of < 0.0001 for study 005; while it yields a p-value of 0.0206 for study 006. IT is statistically superior to vehicle with respect to the responder analysis in both the Evaluator Wrinkle Severity Assessment and the Subject Wrinkle Assessment.

3.1.4.3 Secondary Endpoints

Analyses of the secondary efficacy endpoints are presented in Tables A.11-A.12 of the Appendix. The listed p-values are unadjusted p-values and for reference purposes only. The efficacy trend of the secondary efficacy endpoints is consistent with that of the co-primary efficacy endpoints. The summary is:

- IT is better than vehicle with respect to the proportion of patients with at least 2-point improvement from baseline in Evaluator Wrinkle Severity Assessment over Visit 3, 4, and 5 in studies 005 and 006 (Table A.11).
- IT is better than vehicle with respect to the proportion of patients with at least 2-point improvement from baseline in Subject Wrinkle Assessment over Visit 3, 4, and 5 in studies 005 and 006 (Table A.11).
- A higher rate of patients in IT group felt their appearance at 6-month is better (+1) or much better (+2) than before based on Subject Improvement Assessment in both studies (Table A.12). The response rates were 61% vs. 28.2% for IT vs. vehicle in study 005; they were 54.5% vs. 18.5% in study 006.
- A higher rate of patients in IT group were evaluated as better (+1) or much better (+2) than before in appearance of each side of face at 6-month based on Evaluator Improvement Assessment in both studies (Table A.12). The response rates for right side of face were 52% vs. 18.4% for IT vs. vehicle in study 005; they were 50% vs. 20.4% in study 006. The response rates for left side of face were 52% vs. 18.4% in study 005; they were 47.3% vs. 18.5% in study 006.

3.1.4.4 Other Results

In addition to the analyses of the co-primary efficacy endpoints and secondary efficacy endpoints, several analyses were included in the BLA: (a) Time-to-success analysis; and (b) Responder analysis with at least 1-point improvement from baseline to Visit 3, 4, 5 and 6 in Subject Wrinkle Assessment and Evaluator Wrinkle Severity Assessment.

It should be noted that the responder analyses with at least 1-point improvement in Subject Wrinkle Assessment and in Evaluator Wrinkle Severity Assessment were not pre-specified in the protocols and the SAP. Though the sponsor states that it is clinically relevant to the proposed indication and results show the superiority of IT to vehicle, they are not included in this review.

Time-to-success analysis evaluates the time to sustained response of at least 2-point improvement from baseline. For an example, if a patient had at least 2-point improvement at 2-month post-treatment (Visit 4) and sustained till 6-month post-treatment (Visit 6), the patient's time-to-success is 2-month post-treatment. On the other hand, if a patient had a 2-point improvement at 2-month, had less than 2-point improvement or no evaluation at 4-month, and had a 2-point improvement at 6-month post-treatment, the patient's time-to-success is 6-month post-treatment.

Numerical results of time-to-success analyses are presented in Table A.13 of the Appendix. Treatment comparison was carried out based on log-rank test as pre-specified in the protocols. Kaplan-Meier curves of sustained success rates are presented in Figures 1-4 in the Appendix for studies 005 and 006. Note that the clinical evaluation of Subject Wrinkle Assessment and Evaluator Wrinkle Severity Assessment started at Visit 3 (or Treatment 3). As a result, the sustained success rates of the two treatment groups do not start from 0 at Visit 3. In summary, the time-to-success analysis supports the outcomes of the co-primary efficacy endpoints that IT is superior to vehicle:

- IT is statistically superior to vehicle based on time-to-success analyses (Table A.13).
- For study 005, Figures 1-2 show two separate and parallel success rate curves between Visit 3 and Visit 6 (6-month post-treatment) for Subject Wrinkle Assessment and Evaluator Wrinkle Severity Assessment, respectively.
- Figures 3-4 for study 006 give the same conclusion as that of study 005.

3.1.4.5 Conclusions

IT is statistically superior to vehicle with respect to each of the co-primary efficacy endpoints for studies 005 and 006. The overall success rates in Evaluator Wrinkle Severity Assessment based on the ITT analysis with missing data treated as failures are 33% vs. 7% (IT vs. vehicle) in study 005; and 19% vs. 7% (IT vs. vehicle) in study 006. The overall success rates in Subject Wrinkle Assessment are 57% vs. 30% (IT vs. vehicle) in study 005; and 45% vs. 18% (IT vs. vehicle) in study 006.

Results of the co-primary efficacy endpoints are generally robust over different statistical methods for data analyses. Since missing data rates ranged from 9% to 20% for treatment groups in the two studies, results were impacted in terms of statistical significance in particularly under the worst case scenario, where the worst case scenario imputes missing in IT as failures and missing in vehicle as successes. However, the efficacy trend favors IT treatment numerically in this scenario. Results of the secondary efficacy endpoints and time-to-sustained success analyses support outcomes of the co-primary efficacy endpoints.

3.2 Evaluation of Safety

The pre-specified safety population included all randomized patients who received at least one treatment of either IT or vehicle. This results in 83 (83%) and 92 (89%) patients in IT and vehicle group, respectively, for study 005; and 98 (89%) and 99 (92%) in the respective group for study 006. Results of treatment exposure and incidence of adverse events (AEs) are summarized in Tables 6-7, respectively, below. The summary is:

- a. Treatment exposure:
 - Majority patients in the safety population received all three treatments – 92% vs. 98% for IT vs. vehicle in study 005; and 88% vs. 98% for IT vs. vehicle in study 006.
 - The total treatment dose is comparable between treatment groups within each study with mean 3.0 mL vs. 3.1 mL for IT vs. vehicle in study 005; and 2.7 mL vs. 2.7 mL for IT vs. vehicle in study 006.
 - The time interval between treatments appears to be longer generally in IT group as compared to vehicle group.

- b. Incidence of adverse events:
 - Two groups are generally comparable with respect to the overall incidence rate of treatment emergent adverse events (TEAEs) regardless of relationship to study treatments for each study. The incidence rates (per patient basis) were 61% vs. 62% for IT vs. vehicle in study 005; and 63% vs. 57% in study 006.
 - Totals of 112 (58%) and 81 (37%) TEAEs in IT and vehicle group, respectively, were considered by investigators to be probably or definitely related to study treatment for study 005; while there were 67 (42%) and 80 (46.5%) events in the respective group for study 006.
 - Most TEAEs were mild to moderate in severity. One death occurred in vehicle group for each study.
 - For TEAEs considered to be probably or definitely related to study treatments, all 112 events (100%) in IT group and 76 events (94%) in vehicle group were injection site reactions for study 005 (Sponsor’s BLA, Table 54, Vol.43). Totals of 63 events (94%) in IT group and 80 events (100%) in vehicle group were injection site reactions for study 006 (Sponsor’s BLA, Table 54, Vol.52).

Table 6: Exposure to Treatments – Studies 005 and 006

| Variable | Study 005 | | Study 006 | |
|---|--------------|-------------------|--------------|-------------------|
| | IT (n=83) | Vehicle (n=92) | IT (n=98) | Vehicle (n=99) |
| Total Number of Treatment, n (%) | | | | |
| None | 0 | 0 | 0 | 0 |
| 1 | 3 (4%) | 2 (2%) | 6 (6%) | 0 |
| 2 | 4 (5%) | 0 | 6 (6%) | 2 (2%) |
| 3 | 76 (92%) | 90 (98%) | 86 (88%) | 97 (98%) |
| Total Dose (mL) | | | | |
| Mean (SD) | 3.0 (0.74) | 3.1 (0.66) | 2.7 (0.81) | 2.7 (0.62) |
| Range | 0.9 – 4.8 | 1.0 – 5.4 | 0.7 – 5.0 | 1.2 – 5.1 |
| Days between 1st and 2nd Treatments | | | | |
| Mean (SD) | 41.6 (40.20) | 32.3 (6.24) | 32.9 (8.00) | 34.3 (12.32) |
| Range | 24 – 259 | 25 – 58 | 23 – 72 | 26 – 133 |
| Days between 2nd and 3rd Treatments | | | | |
| Mean (SD) | 38.0 (26.41) | 34.7 (8.97) | 41.4 (30.49) | 33.2 (12.41) |
| Range | 22 – 239 | 21 – 77 | 23 – 196 | 21 – 140 |
| Source: Module 5 – Table 14.3.1 in Vol.44; and Table 14.3.1 in Vol.53. | | | | |

Table 7: Incidence of Adverse Events – Studies 005 and 006

| Variable | Study 005 | | | | Study 006 | | | |
|---------------------------------------|--------------|--------|-------------------|--------|--------------|--------|-------------------|--------|
| | IT (n=83) | | Vehicle (n=92) | | IT (n=98) | | Vehicle (n=99) | |
| | n (%) | Events | n (%) | Events | n (%) | Events | n (%) | Events |
| All Treatment-Emergent Adverse Events | 51 (61%) | 193 | 57 (62%) | 219 | 62 (63%) | 161 | 56 (57%) | 172 |
| Severity ¹ | | | | | | | | |
| Mild | 30 (36%) | 165 | 30 (33%) | 160 | 35 (36%) | 109 | 32 (32%) | 131 |
| Moderate | 19 (23%) | 24 | 20 (22%) | 51 | 20 (20%) | 43 | 19 (19%) | 36 |
| Severe | 2 (2%) | 4 | 5 (5%) | 5 | 7 (7%) | 9 | 4 (4%) | 4 |
| Deaths | 0 | 0 | 1 (1%) | 1 | 0 | 0 | 1 (1%) | 1 |
| Relationship ² | | | | | | | | |
| Unrelated | 20 (24%) | 69 | 25 (27%) | 120 | 23 (24%) | 78 | 19 (19%) | 79 |
| Unlikely | 2 (2%) | 5 | 4 (4%) | 10 | 7 (7%) | 11 | 6 (6%) | 12 |
| Possible | 1 (1%) | 7 | 2 (2%) | 7 | 1 (1%) | 5 | 0 | 0 |
| Probable | 4 (5%) | 10 | 3 (3%) | 8 | 5 (5%) | 8 | 2 (2%) | 4 |
| Definitely | 24 (29%) | 102 | 22 (24%) | 73 | 26 (27%) | 59 | 29 (29%) | 76 |
| TEAE leading to early termination | 1 (1%) | 1 | 1 (1%) | 1 | 1 (1%) | 1 | 0 | 0 |
| Serious TEAE | 1 (1%) | 1 | 3 (3%) | 4 | 6 (6%) | 7 | 3 (3%) | 4 |

Source: Module 5 – Vol.43 (Tables 51-54, p.119-125); Vol.52 (Tables 51-54, p.117-122)).

¹ Subjects -(b)(6)- and -(b)(6)- in vehicle group for study 005 had TEAEs with missing severity.

² Subject -(b)(6)- in vehicle group for study 005 had TEAEs with missing relationship to study treatment.

3.3 Gender, Race, Age and Other Special/Subgroup Populations

Subgroup efficacy results by gender, race, age, and baseline characteristics for studies 005 and 006 are examined. Results are presented in Tables A.14 – A.15 of the Appendix. Subgroup results are intended to be used to observe efficacy trend over subgroups. Inferential statistical comparisons between treatment groups within subgroup are inappropriate as studies were not designed for such a purpose.

3.3.1 Gender

As the majority of the enrolled patients are female (about 90%), the efficacy results of female subjects are similar to those of the overall ITT results. The success rates for female in Subject Wrinkle Assessment were 58% vs. 29% (IT vs. vehicle) in study 005; and 45% vs. 19% (IT vs. vehicle) in study 006. The success rates for female subjects in Evaluator Wrinkle Severity Assessment were 35% vs. 6% (IT vs. vehicle) in study 005; and 19% vs. 7% (IT vs. vehicle) in study 006.

The overall efficacy trend favors IT within male subjects for each study. The male enrollment accounted for about 9.7% of study population and was under-represented.

3.3.2 Race

The enrollment in both studies was pre-dominated by white subjects with an overall 92% of the study population. As a result, the success rates in Subject Wrinkle Assessment and Evaluator Wrinkle Severity Assessment for white subjects are similar to those of the overall ITT results. The success rates for white subjects in Subject Wrinkle Assessment were 55% vs. 29% (IT vs. vehicle)

in study 005; and 46% vs. 17% (IT vs. vehicle) in study 006. The success rates for white subjects in Evaluator Wrinkle Severity Assessment were 32% vs. 7% (IT vs. vehicle) in study 005; and 17% vs. 6% (IT vs. vehicle) in study 006.

The evidence of IT efficacy in non-white subjects is limited due to the small size (8%).

3.3.3 Age

The efficacy trend favors IT for all age groups except the group of 65 years and older in Evaluator Wrinkle Severity Assessment for study 005; and in Subject Wrinkle Assessment for study 006. That is, the success rates in Evaluator Wrinkle Severity Assessment were 5% vs. 13% (IT vs. vehicle) in study 005; and the success rates in Subject Wrinkle Assessment were 13% vs. 16% (IT vs. vehicle) in study 006 for age group of 65 years and older. This age group accounted for about 17% of study population.

Additionally, patients aged 40 and younger accounted for 6% (= 25/421) of study population with 1% (= 5/421) of patients aged below 35. Though the efficacy trend favors IT treatment in this age group, the proposed lower limit age of 18 years old for the product may be an issue due to under-represented subgroup.

3.3.4 Baseline Evaluator Wrinkle Severity Assessment

Since Evaluator Wrinkle Severity Assessment evaluated both sides of face of an individual, subgroup results of the co-primary efficacy endpoints by Evaluator Wrinkle Severity Assessment on each side of face are summarized.

The efficacy trend of IT vs. vehicle is generally consistent within each baseline wrinkle severity grade regardless of face side. The success rate of Subject Wrinkle Assessment in IT group appears to be decreasing as the baseline Evaluator Wrinkle Severity grade increases. Such a trend however is not observed in the success rate of Evaluator Wrinkle Severity Assessment.

3.3.5 Baseline Subject Wrinkle Assessment

Patients had a score of -2 (very dissatisfied) or -1 (dissatisfied) in Subject Wrinkle Assessment when they entered the studies. Results of the co-primary efficacy endpoints by baseline Subject Wrinkle Assessment generally indicate that IT is better than vehicle within each subgroup in both studies. Interestingly, the success rates in Subject Wrinkle Assessment and Evaluator Wrinkle Severity Assessment increases as dissatisfaction increases in study 005 regardless of treatment received. Study 006 appears to have the opposite trend as compared to study 005.

3.3.6 Baseline Wrinkle Length

Results in Tables A.14-A.15 of the Appendix show that generally IT is better than vehicle with respect to the co-primary efficacy endpoints within each subgroup of baseline wrinkle length. Generally, the shorter the wrinkle length was at baseline, the higher the success rates in Subject Wrinkle Assessment and in Evaluator Wrinkle Severity Assessment were in the IT group.

4. SUMMARY AND CONCLUSIONS

4.1 Statistical Issues and Collective Evidence

Though the overall efficacy results of IT are superior to vehicle with respect to the two co-primary efficacy endpoints, the following issues are observed:

- In study IT-R-006, sites 6100, 6300 and 6600 had considerably lower success rate for IT treatment in Evaluator Wrinkle Severity Assessment than other sites. The success rates were 5.3% (1/19), 4.5% (1/22) and 10% (2/20) in the respective sites. Following the examination of relevant information, evaluator assessment may be a potential factor.
- The evidence of IT efficacy is limited for male and non-white subjects due to under-represented subgroups.
- Only a total of 5 subjects (1%) aged ≤ 34 years old; and a total of 25 subjects (6%) aged ≤ 40 years old were randomized in the pivotal trials. Although the efficacy trend favors IT, the evidence of IT efficacy is limited in this age group. Consequently, the proposed lower limit of age of 18 years old for the product may be an issue.
- Observed numerically reverse efficacy trend in different co-primary endpoints for the two studies in subjects aged 65 and older. These subjects accounted for 17% of study population.
- As only efficacy data up to 6-month post-treatment are evaluated, long-term efficacy beyond 6-month post-treatment has not been established.

4.2 Conclusions and Recommendations

Results of two pivotal studies (IT-R-005 and IT-R-006) were submitted in support of the efficacy and safety claim of IT Therapy (IT) for the treatment of moderate to severe nasolabial fold wrinkles in adults 18 years of age or older.

In agreement with the FDA, the co-primary efficacy endpoints are:

- a. Proportion of patients with at least 2-point improvement from baseline to 6-month post-treatment on both sides of face in Evaluator Wrinkle Severity Assessment.
- b. Proportion of patients with at least 2-point improvement from baseline to 6-month post-treatment in Subject Wrinkle Assessment.

Each study is declared as a success if IT is shown to be superior to vehicle with respect to each of the co-primary efficacy endpoints. Results of the co-primary efficacy endpoints based on the intent-to-treat (ITT) population (subjects with missing data were treated as treatment failure) are summarized in the following:

| No. of subjects (%) | Study IT-R-005 | | | Study IT-R-006 | | |
|---------------------------------|-----------------|----------------------|----------|-----------------|----------------------|----------|
| | IT (n = 100) | Vehicle (n = 103) | p-value | IT (n = 110) | Vehicle (n = 108) | p-value |
| Evaluator Wrinkle Assessment | 33 (33%) | 7 (7%) | < 0.0001 | 21 (19%) | 8 (7%) | 0.0075 |
| Subject Wrinkle Assessment | 57 (57%) | 31 (30%) | 0.0001 | 50 (45%) | 19 (18%) | < 0.0001 |

In summary, IT is statistically superior to vehicle with respect to each co-primary efficacy endpoints.

- The success rates in Evaluator Wrinkle Severity Assessment were 33% for IT vs. 7% for vehicle in study IT-R-005; and 19% for IT vs. 7% for vehicle in study IT-R-006.
- The success rates in Subject Wrinkle Assessment were 57% for IT vs. 30% for vehicle in study IT-R-005; and 45% for IT vs. 18% for vehicle in study IT-R-006.

The overall results of the co-primary efficacy endpoints are generally robust over different statistical methods for data analyses. Since missing data rates ranged from 9% to 20% for treatment groups in the two studies, results were impacted in terms of statistical significance in particular under the worst case scenario for missing data handling, where the worst case scenario imputes missing in IT as failures and missing in vehicle as successes. However, the efficacy trend favors IT numerically in this scenario. Additionally, results of the secondary efficacy endpoints support outcomes of the co-primary efficacy endpoints.

Though the overall results of IT are superior to vehicle with respect to the co-primary efficacy endpoints, the following issues are observed:

- a. In study IT-R-006, sites 6100, 6300 and 6600 had considerably lower success rate for IT treatment in Evaluator Wrinkle Severity Assessment than other sites. The success rates were 5.3% (1/19), 4.5% (1/22) and 10% (2/20) in the respective sites. Following the examination of relevant information, evaluator assessment may be a potential factor.
- b. The evidence of IT efficacy is limited for male and non-white subjects due to under-represented subgroups.
- c. Only a total of 5 subjects (1%) aged ≤ 34 years old; and a total of 25 subjects (6%) aged ≤ 40 years old were randomized in the pivotal trials. Although the efficacy trend favors IT, the evidence of IT efficacy is limited in this age group. Consequently, the proposed lower limit of age of 18 years old for the product may be an issue.
- d. Observed numerically reverse efficacy trend in different co-primary endpoints for the two studies in subjects aged 65 and older. These subjects accounted for 17% of study population.
- e. As only efficacy data up to 6-month post-treatment are evaluated, long-term efficacy beyond 6-month post-treatment has not been established.

For possible labeling inclusion, though the sponsor's pre-specified hierarchical testing of Subject Improvement Assessment, Evaluator Improvement Assessment and time-to-success on the co-primary endpoints in the Statistical Analysis Plan (dated 6/17/2008) satisfied statistical principle and was reviewed by the FDA; the endpoints to be included in the labeling however should be clinically relevant to the proposed indication at a minimum. Medical Division's input is essential regarding labeling inclusion.

For safety evaluation, two treatment groups are generally comparable with respect to the overall incidence rate (per patient basis) of treatment emergent adverse events (TEAEs) in both trials regardless of relationship to the study treatments. The overall incidence rates were 61% vs. 62% for IT vs. vehicle in study IT-R-005; and 63% vs. 57% in study IT-R-006. Most TEAEs were injection site reactions and were mild to moderate in severity. Medical Division should provide more in-depth safety review from clinical perspective.

APPENDICES

Table A.1: Subjects with Re-Biopsy Prior to Randomization – Studies 005 and 006

| Subject ID | Treatment | Reason for Re-biopsy | 1 st Biopsy Date | Re-Biopsy Date | Injection Date | | |
|--|-----------|----------------------|-----------------------------|----------------|----------------|---------|---------|
| | | | | | 1 | 2 | 3 |
| Pre-Randomization* | | | | | | | |
| -(b)(6)- | IT | Shipping error | 11/21/06 | 11/28/06 | 2/16/07 | 3/20/07 | 4/19/07 |
| -(b)(6)- | Vehicle | Shipping error | 11/21/06 | 11/28/06 | 2/9/07 | 3/6/07 | 4/10/07 |
| -(b)(6)- | Vehicle | Shipping error | 11/21/06 | 11/28/06 | 2/1/07 | 3/2/07 | 4/5/07 |
| -(b)(6)- | Vehicle | Shipping error | 12/21/06 | 1/2/07 | NA | NA | NA |
| -(b)(6)- | IT | Shipping error | 12/21/06 | 1/4/07 | 6/27/07 | 8/3/07 | 8/29/07 |
| -(b)(6)- | Vehicle | Shipping error | 12/21/06 | 1/16/07 | 4/10/07 | NA | NA |
| -(b)(6)- | IT | Vial not labeled | 11/7/06 | 11/14/06 | 4/26/07 | 6/21/07 | 8/8/07 |
| -(b)(6)- | IT | Shipping error | 1/25/07 | 1/31/07 | 5/24/07 | 6/21/07 | 7/18/07 |
| <p>Source: Module 5 – Tables 1 and 2 in Appendix 16.1.13 Summary IT Manufacture, Vol. 46 and SAS electronic datasets.</p> <p>* At the pre-randomization stage, subjects have not been randomized. The column of Treatment is for reference purpose.</p> <p>-(b)(6)- had re-biopsy but lost to follow-up.</p> <p>-(b)(6)- was not shown for the 2nd injection and thereafter.</p> | | | | | | | |

Table A.2: Subjects with Re-Biopsy Post-Randomization – Studies 005 and 006

| Subject ID | Treatment | Reason for Re-biopsy | 1 st Biopsy Date | Re-Biopsy Date | Injection Date | | |
|---|-----------|---------------------------------------|-----------------------------|----------------|----------------|---------|----------|
| | | | | | 1 | 2 | 3 |
| Post-Randomization | | | | | | | |
| -(b)(6)- | IT | Procedural | 1/9/07 | 1/29/07 | 6/19/07 | 7/26/07 | 8/28/07 |
| -(b)(6)- | IT | Insufficient material | 11/15/06 | 9/20/07 | 1/24/07 | 3/2/07 | NA |
| -(b)(6)- | IT | Slow growth of 1 st biopsy | 12/20/06 | 8/27/07 | 4/11/07 | 5/17/07 | 11/29/07 |
| -(b)(6)- | Vehicle | Sham biopsy for -(b)(6)- | 12/13/06 | 8/15/07 | 3/14/07 | 7/25/07 | 12/12/07 |
| -(b)(6)- | IT | Culture contamination | 12/13/06 | 1/24/07 | 7/10/07 | 8/8/07 | 12/13/07 |
| <p>Source: Module 5 – Tables 1 and 2 in Appendix 16.1.13 Summary IT Manufacture, Vol. 46 and SAS electronic datasets.</p> <p>-(b)(6)- had re-biopsy due to delay in start of processing of the 1st biopsy.</p> <p>-(b)(6)- had re-biopsy in order to provide Treatment 3. However, the re-biopsy was rejected. Subject did not want to have another biopsy.</p> | | | | | | | |

**Table A.3: Patient Enrollment and Treatment by Site
Studies 005 and 006 (ITT)**

| Study 005 Site ID | IT | Vehicle | Total |
|--|-----------|----------------|--------------|
| 5100 | 25 (25%) | 24 (23%) | 49 (24%) |
| 5200 | 10 (10%) | 11(11%) | 21 (10%) |
| 5300 | 21 (21%) | 20 (19%) | 41 (20%) |
| 5400 | 7 (7%) | 8 (8%) | 15 (7%) |
| 5500 | 9 (9%) | 10 (10%) | 19 (9%) |
| 5600 | 21 (21%) | 22 (21%) | 43 (21%) |
| 5700 | 7 (7%) | 8 (8%) | 15 (7%) |
| Total | 100 | 103 | 203 |
| Study 006 Site ID | IT | Vehicle | Total |
| 6100 | 19 (17%) | 17 (16%) | 36 (17%) |
| 6200 | 13 (12%) | 16 (15%) | 29 (13%) |
| 6300 | 22 (20%) | 22 (20%) | 44 (20%) |
| 6400 | 18 (16%) | 17 (16%) | 35 (16%) |
| 6500 | 18 (16%) | 17 (16%) | 35 (16%) |
| 6600 | 20 (18%) | 19 (18%) | 39 (18%) |
| Total | 110 | 108 | 218 |
| Source: Table 9 on p.58, Vol.43; Table 9 on p.57, Vol.52; and Sponsor's SAS electronic data sets. | | | |

Table A.4: Patient Demographic and Baseline Characteristics – Study 005 (ITT)

| Demographic/Baseline | ITT (n = 100) | Vehicle (n = 103) | Group Comparability |
|--|--------------------------|------------------------------|--------------------------------|
| Age (years) | | | |
| Mean (SD) | 57.5 (8.32) | 55.9 (7.87) | 0.1650 |
| Min – Max | 38 – 75 | 35 – 78 | |
| Age Group, n(%) | | | 0.3311 |
| <= 40 years | 4 (4%) | 3 (2.9%) | |
| 41 – 50 years | 15 (15%) | 22 (21.4%) | |
| 51 – 64 years | 60 (60%) | 62 (60.2%) | |
| >= 65 years | 21 (21%) | 16 (15.5%) | |
| Gender | | | |
| Male | 12 (12%) | 9 (8.7%) | 0.4953 |
| Female | 88 (88%) | 94 (91.3%) | |
| Race | | | |
| American Indian or Alaska Native | 0 | 1 (1%) | 0.3796 |
| Asian | 2 (2%) | 0 | |
| Black or African American | 1 (1%) | 2 (1.9%) | |
| White | 94 (94%) | 99 (96.1%) | |
| Other | 3 (3%) | 1 (1%) | |
| Total wrinkle length (cm) | | | |
| Mean (SD) | 10.6 (1.98) | 10.5 (2.0) | 0.4720 |
| Min – Max | 6.2 – 16.0 | 6.5 – 18.0 | |
| Subject Wrinkle Assessment | | | |
| Very dissatisfied (-2) | 65 (65%) | 61 (59.2%) | 0.3976 |
| Dissatisfied (-1) | 35 (35%) | 42 (40.8%) | |
| Somewhat satisfied (0) | 0 | 0 | |
| Satisfied (+1) | 0 | 0 | |
| Very Satisfied (+2) | 0 | 0 | |
| Evaluator Wrinkle Severity Assessment, right | | | |
| Grade 0 (No visible wrinkle) | 0 | 0 | 0.9394 |
| Grade 1 (Just perceptible) | 0 | 0 | |
| Grade 2 (Shallow wrinkle) | 0 | 0 | |
| Grade 3 (Moderate deep) | 49 (49%) | 53 (51.5%) | |
| Grade 4 (Deep wrinkle) | 39 (39%) | 38 (36.9%) | |
| Grade 5 (Very deep wrinkle) | 12 (12%) | 12 (11.7%) | |
| Evaluator Wrinkle Severity Assessment, left | | | |
| Grade 0 (No visible wrinkle) | 0 | 0 | 0.6686 |
| Grade 1 (Just perceptible) | 0 | 0 | |
| Grade 2 (Shallow wrinkle) | 0 | 0 | |
| Grade 3 (Moderate deep) | 44 (44%) | 51 (49.5%) | |
| Grade 4 (Deep wrinkle) | 45 (45%) | 40 (38.8%) | |
| Grade 5 (Very deep wrinkle) | 11 (11%) | 12 (11.7%) | |
| Source: Table 14.1.6.1, Table 14.1.6.2, and Table 14.6.3.1, Vol.43; and sponsor’s SAS electronic data sets. | | | |

Table A.5: Patient Demographic and Baseline Characteristics – Study 006 (ITT)

| Demographic/Baseline | IT (n = 110) | Vehicle (n = 108) | Group Comparability |
|--|-------------------------|------------------------------|--------------------------------|
| Age (years) | | | |
| Mean (SD) | 53.9 (10.38) | 55.4 (9.42) | 0.4578 |
| Min – Max | 23 – 75 | 26 – 81 | |
| Age Group, n(%) | | | 0.1272 |
| <= 40 years | 14 (12.7%) | 4 (3.7%) | |
| 41 – 50 years | 25 (22.7%) | 30 (27.8%) | |
| 51 – 64 years | 56 (50.9%) | 55 (50.9%) | |
| >= 65 years | 15 (13.6%) | 19 (17.6%) | |
| Gender | | | |
| Male | 7 (6.4%) | 13 (12%) | 0.1655 |
| Female | 103 (93.6%) | 95 (88%) | |
| Race | | | |
| American Indian or Alaska Native | 0 | 0 | 0.9138 |
| Asian | 0 | 0 | |
| Black or African American | 1 (0.9%) | 1 (0.9%) | |
| White | 98 (89.1%) | 95 (88.0%) | |
| Other | 11 (10.0%) | 12 (11.1%) | |
| Total wrinkle length (cm) | | | |
| Mean (SD) | 9.5 (2.34) | 9.2 (1.93) | 0.5818 |
| Min – Max | 4.0 – 16.5 | 4.0 – 17.0 | |
| Subject Wrinkle Assessment | | | |
| Very dissatisfied (-2) | 69 (62.7%) | 62 (57.4%) | 0.4237 |
| Dissatisfied (-1) | 41 (37.3%) | 46 (42.6%) | |
| Somewhat satisfied (0) | 0 | 0 | |
| Satisfied (+1) | 0 | 0 | |
| Very Satisfied (+2) | 0 | 0 | |
| Evaluator Wrinkle Severity Assessment, right | | | |
| Grade 0 (No visible wrinkle) | 0 | 0 | 0.2054 |
| Grade 1 (Just perceptible) | 0 | 0 | |
| Grade 2 (Shallow wrinkle) | 0 | 0 | |
| Grade 3 (Moderate deep) | 55 (50.0%) | 45 (41.7%) | |
| Grade 4 (Deep wrinkle) | 41 (37.3%) | 53 (49.1%) | |
| Grade 5 (Very deep wrinkle) | 14 (12.7%) | 10 (9.3%) | |
| Evaluator Wrinkle Severity Assessment, left | | | |
| Grade 0 (No visible wrinkle) | 0 | 0 | 0.5535 |
| Grade 1 (Just perceptible) | 0 | 0 | |
| Grade 2 (Shallow wrinkle) | 0 | 0 | |
| Grade 3 (Moderate deep) | 49 (44.5%) | 43 (39.8%) | |
| Grade 4 (Deep wrinkle) | 45 (40.9%) | 52 (48.1%) | |
| Grade 5 (Very deep wrinkle) | 16 (14.5%) | 13 (12.0%) | |
| Source: Table 14.1.6.1, Table 14.1.6.2, and Table 14.6.3.1, Vol.52; and sponsor’s SAS electronic data sets. | | | |

**Table A.6: Repeat Measures Analysis of the Co-Primary Efficacy Endpoints
Studies 005 and 006 (ITT)**

| | Repeated Measures Analysis² | | | |
|---|---|----------------|--|----------------|
| Study 005 | Subject Wrinkle Assessment | | Evaluator Wrinkle Severity Assessment | |
| Least Squares Estimate of Successes¹ | IT | Vehicle | IT | Vehicle |
| Visit 3 | 39% | 22% | 15% | 5% |
| Visit 4 | 57% | 26% | 33% | 11% |
| Visit 5 | 56% | 27% | 33% | 10% |
| Visit 6 | 67% | 33% | 41% | 8% |
| Comparison (GEE ²) | | | | |
| Treatment | < 0.0001 | | < 0.0001 | |
| Visit | < 0.0001 | | 0.0157 | |
| Treat/Visit interaction | 0.3016 | | 0.4134 | |
| Study 006 | Subject Wrinkle Assessment | | Evaluator Wrinkle Severity Assessment | |
| Least Squares Estimate of Successes¹ | IT | Vehicle | IT | Vehicle |
| Visit 3 | 33% | 19% | 16% | 4% |
| Visit 4 | 55% | 25% | 24% | 8% |
| Visit 5 | 51% | 24% | 27% | 7% |
| Visit 6 | 51% | 19% | 21% | 8% |
| Comparison (GEE ²) | | | | |
| Treatment | < 0.0001 | | 0.0001 | |
| Visit | 0.0031 | | 0.0451 | |
| Treat/Visit interaction | 0.2842 | | 0.7693 | |
| Source: Module 5 – Vol.43 (Table 22 on p.83, Table 23 on p.84, Table 25 on p.87); Vol.52 (Table 22 on p.83, Table 23 on p.84, Table 25 on p.86). | | | | |
| ¹ Success is defined as at least 2-point improvement from baseline. | | | | |
| ² GEE repeated measures model with unstructured covariance matrix. | | | | |

**Table A.7: Sensitivity Analyses of the Co-Primary Efficacy Endpoints
due to Missing Data Handling – Studies 005 and 006 (ITT)**

| Sensitivity Analyses | Subject Wrinkle Assessment | | | Evaluator Wrinkle Severity Assessment | | |
|--|----------------------------|------------|----------------------|---------------------------------------|------------|----------------------|
| | IT | Vehicle | p-value ¹ | IT | Vehicle | p-value ¹ |
| Study 005 (100, 103) | | | | | | |
| Missing as Success | 77 (77%) | 46 (44.7%) | < 0.0001 | 53 (53%) | 22 (21.4%) | < 0.0001 |
| Worst Case Scenario | 57 (57%) | 46 (44.7%) | 0.0783 | 33 (33%) | 22 (21.4%) | 0.0572 |
| LOCF | 57 (57%) | 31 (30.1%) | 0.0001 | 33 (33%) | 7 (6.8%) | < 0.0001 |
| Study 006 (110, 108) | | | | | | |
| Missing as Success | 67 (60.9%) | 29 (26.9%) | < 0.0001 | 38 (34.5%) | 18 (16.7%) | 0.0016 |
| Worst Case Scenario | 50 (45.5%) | 29 (26.9%) | 0.0049 | 21 (19.1%) | 18 (16.7%) | 0.5941 |
| LOCF | 51 (46.4%) | 19 (17.6%) | < 0.0001 | 21 (19.1%) | 8 (7.4%) | 0.0075 |
| Source: Module 5 – Vol.43 (Table 14.2.1.2, Table 14.2.3.3) and Vol.52 (Table 14.2.1.2, Table 14.2.3.3). | | | | | | |
| ¹ p-value is based on CMH test stratified by site. | | | | | | |

Table A.8: Distribution of No. of Points Change from Baseline to Visit 6 in Subject Wrinkle Assessment – Studies 005 and 006

| No. of points improvement | Study 005 | | Study 006 | |
|---------------------------|---------------|--------------------|---------------|--------------------|
| | IT (n=100) | Vehicle (n=103) | IT (n=110) | Vehicle (n=108) |
| Missing | 20 (20%) | 15 (14.6%) | 17 (15.5%) | 10 (9.3%) |
| -1 (worse) | 2 (2%) | 5 (4.8%) | 3 (2.7%) | 7 (6.5%) |
| 0 (no change) | 7 (7%) | 38 (36.9%) | 19 (17.3%) | 45 (41.7%) |
| 1 | 14 (14%) | 14 (13.6%) | 21 (19.1%) | 27 (25%) |
| 2 | 37 (37%) | 22 (21.4%) | 29 (26.4%) | 16 (14.8%) |
| 3 | 11 (11%) | 6 (5.8%) | 15 (13.6%) | 3 (2.8%) |
| 4 | 9 (9%) | 3 (2.9%) | 6 (5.5%) | 0 |

Source: Sponsor's SAS electronic data sets.

Table A.9(a): Distribution of No. of Points Change from Baseline to Visit 6 in Evaluator Wrinkle Severity Assessment – Study 005

| Group | Right side face No. points Improv. | Left side face No. of points Improvement | | | | | | |
|-----------------|------------------------------------|--|------------|------------|------------|------------|----------|--------|
| | | Missing | -2 (worse) | -1 (worse) | 0 | 1 | 2 | 3 |
| IT (n=100) | Missing | 20 (20%) | 0 | 0 | 0 | 0 | 0 | 0 |
| | -1 (worse) | 0 | 0 | 1 (1%) | 0 | 0 | 0 | 0 |
| | 0 | 0 | 0 | 0 | 11 (11%) | 0 | 1 (1%) | 0 |
| | 1 | 0 | 0 | 0 | 5 (5%) | 17 (17%) | 4 (4%) | 1 (1%) |
| | 2 | 0 | 0 | 0 | 0 | 7 (7%) | 20 (20%) | 4 (4%) |
| | 3 | 0 | 0 | 0 | 0 | 0 | 4 (4%) | 3 (3%) |
| | 4 | 0 | 0 | 0 | 0 | 0 | 1 (1%) | 1 (1%) |
| Vehicle (n=103) | Missing | 15 (14.6%) | 0 | 0 | 0 | 0 | 0 | 0 |
| | -1 (worse) | 0 | 0 | 2 (1.9%) | 4 (3.9%) | 1 (1%) | 0 | 0 |
| | 0 | 0 | 1 (1%) | 4 (3.9%) | 25 (24.3%) | 7 (6.8%) | 0 | 0 |
| | 1 | 0 | 0 | 0 | 8 (7.8%) | 16 (15.5%) | 5 (4.9%) | 0 |
| | 2 | 0 | 0 | 0 | 1 (1%) | 7 (6.8%) | 5 (4.9%) | 0 |
| | 3 | 0 | 0 | 0 | 0 | 0 | 1 (1%) | 1 (1%) |

Source: Sponsor's SAS electronic data sets.

Table A.9(b): Distribution of No. of Points Change from Baseline to Visit 6 in Evaluator Wrinkle Severity Assessment – Study 006

| Group | Right side face No. points Improv. | Left side face No. of points Improvement | | | | | | |
|-----------------|------------------------------------|--|----------|----------|------------|----------|----------|----------|
| | | Missing | -1 | 0 | 1 | 2 | 3 | 4 |
| IT (n=110) | Missing | 17 (15.5%) | 0 | 0 | 0 | 0 | 0 | 0 |
| | -1 | 0 | 0 | 1 (0.9%) | 0 | 0 | 0 | 0 |
| | 0 | 0 | 0 | 22 (20%) | 4 (3.6%) | 0 | 0 | 0 |
| | 1 | 0 | 0 | 11 (10%) | 21 (19%) | 5 (4.5%) | 0 | 0 |
| | 2 | 0 | 0 | 1 (0.9%) | 7 (6.4%) | 4 (3.6%) | 3 (2.7%) | 1 (0.9%) |
| | 3 | 0 | 0 | 0 | 0 | 5 (4.5%) | 7 (6.4%) | 0 |
| | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 (0.9%) |
| Vehicle (n=108) | Missing | 10 (15.5%) | 0 | 0 | 0 | 0 | 0 | 0 |
| | -1 | 0 | 3 (2.7%) | 3 (2.7%) | 1 (0.9%) | 0 | 0 | 0 |
| | 0 | 0 | 3 (2.7%) | 37 (34%) | 5 (4.6%) | 0 | 0 | 0 |
| | 1 | 0 | 0 | 12 (11%) | 18 (16.7%) | 4 (3.7%) | 0 | 0 |
| | 2 | 0 | 0 | 1 (0.9%) | 3 (2.7%) | 5 (4.6%) | 1 (0.9%) | 1 (0.9%) |
| | 3 | 0 | 0 | 0 | 0 | 0 | 1 (0.9%) | 0 |

Source: Sponsor's SAS electronic data sets.

Table A.10: Distribution of Successes (or Responders) in Evaluator Wrinkle Severity Assessment and Subject Wrinkle Assessment Studies 005 and 006 (ITT) – Reviewer’s Analysis

| Study Size (IT, vehicle) | | IT | | | Vehicle | | |
|---|----------------|---------------------------------------|----------------|--------------|---------------------------------------|----------------|--------------|
| | | Evaluator Wrinkle Severity Assessment | | | Evaluator Wrinkle Severity Assessment | | |
| Study 005 (100, 103) | | Failure | Success | Total | Failure | Success | Total |
| Subject Wrinkle Assessment | Failure | 37 (37%) | 6 (6%) | 43 (43%) | 68 (66%) | 4 (3.9%) | 72 (69.9%) |
| | Success | 30 (30%) | 27 (27%) | 57 (57%) | 28 (27.2%) | 3 (2.9%) | 31 (30.1%) |
| | Total | 67 (67%) | 33 (33%) | 100 | 96 (93.2%) | 7 (6.8%) | 103 |
| Study 006 (110, 108) | | Failure | Success | Total | Failure | Success | Total |
| Subject Wrinkle Assessment | Failure | 52 (47.3%) | 8 (7.3%) | 60 (54.5%) | 85 (78.7%) | 4 (3.7%) | 89 (82.4%) |
| | Success | 37 (33.6%) | 13 (11.8%) | 50 (45.5%) | 15 (13.9%) | 4 (3.7%) | 19 (17.6%) |
| | Total | 89 (80.9%) | 21 (19.1%) | 110 | 100 (92.6%) | 8 (7.4%) | 108 |
| Source: Sponsor’s electronic SAS data sets. Missing data are treated as treatment failures. | | | | | | | |

Table A.11: Secondary Endpoints – Proportion of Patients with ≥ 2 Pts Improvement in Subject Wrinkle Assessment and Evaluator Wrinkle Severity Assessment over Visits Studies 005 and 006 (ITT)

| No. of subjects (%) | | Study 005 | | | Study 006 | | |
|--|---------|-----------------|----------------------|----------------------|-----------------|----------------------|----------------------|
| ITT Analysis | Visit | IT (n = 100) | Vehicle (n = 103) | p-value ¹ | IT (n = 110) | Vehicle (n = 108) | p-value ¹ |
| Subject Wrinkle Assessment | Visit 3 | 38 (38%) | 23 (22.3%) | 0.0133 | 34 (30.9%) | 20 (18.5%) | 0.0311 |
| | Visit 4 | 49 (49%) | 25 (24.3%) | 0.0001 | 51 (46.4%) | 25 (23.1%) | 0.0003 |
| | Visit 5 | 48 (48%) | 26 (25.2%) | 0.0006 | 47 (42.7%) | 24 (22.2%) | 0.0012 |
| Evaluator Wrinkle Severity Assessment | Visit 3 | 14 (14%) | 5 (4.9%) | 0.0211 | 17 (15.5%) | 4 (3.7%) | 0.0022 |
| | Visit 4 | 28 (28%) | 10 (9.7%) | 0.0005 | 22 (20.0%) | 8 (7.4%) | 0.0039 |
| | Visit 5 | 27 (27%) | 9 (8.7%) | 0.0003 | 26 (23.6%) | 7 (6.5%) | 0.0002 |

Source: Module 5 – Vol.43 (Table 14.2.1.2, Table 14.2.3.3) and Vol.52 (Table 14.2.1.2, Table 14.2.3.3).
^{*}Missing data are treated as treatment failures in all analyses.
¹p-values are unadjusted and are based on CMH test stratified by study site. They are listed for reference purpose.

Table A.12: Secondary Endpoints – Subject Improvement Assessment and Evaluator Improvement Assessment at Visit 6 – Studies 005 and 006 (ITT)

| No. of subjects (%) | Study 005 | | | Study 006 | | |
|---|-----------------|----------------------|----------------------|-----------------|----------------------|----------------------|
| | IT (n = 100) | Vehicle (n = 103) | p-value ¹ | IT (n = 110) | Vehicle (n = 108) | p-value ¹ |
| Available-Data Analysis | | | | | | |
| Subject Improvement Assessment | | | | | | |
| No Info. | 22 (22%) | 15 (14.6%) | | 17 (15.5%) | 10 (9.3%) | |
| No improved or worsening | 17 (17%) | 59 (57.3%) | < 0.0001 | 33 (30%) | 78 (72.2%) | < 0.0001 |
| ≥ 1 pt. better than before | 61 (61%) | 29 (28.2%) | | 60 (54.5%) | 20 (18.5%) | |
| Evaluator Improvement Assessment | | | | | | |
| Right Side | | | | | | |
| No Info. | 22 (22%) | 15 (14.6%) | | 17 (15.5%) | 11 (10.2%) | |
| No improved or worsening | 26 (26%) | 69 (67.0%) | < 0.0001 | 38 (34.5%) | 75 (69.4%) | < 0.0001 |
| ≥ 1 pt. better than before | 52 (52%) | 19 (18.4%) | | 55 (50%) | 22 (20.4%) | |
| Evaluator Improvement Assessment | | | | | | |
| Left Side | | | | | | |
| No Info. | 22 (22%) | 15 (14.6%) | | 17 (15.5%) | 10 (9.3%) | |
| No improved or worsening | 26 (26%) | 69 (67.0%) | < 0.0001 | 41 (37.3%) | 78 (72.2%) | < 0.0001 |
| ≥ 1 pt. better than before | 52 (52%) | 19 (18.4%) | | 52 (47.3%) | 20 (18.5%) | |
| Source: Module 5 – Vol.43 (Table 14.2.2, Table 14.2.4.1 and Table 14.2.4.2); and Vol.52 (Table 14.2.2, Table 14.2.4.1 and Table 14.2.4.2). | | | | | | |
| ¹ Sponsor’s p-values are based on patients who had data at Visit 6, and were derived using CMH stratified by site for ≥ 1 pt. better than before vs. no improved or worsening. | | | | | | |

**Table A.13: Time-To-Success Analysis of the Co-Primary Efficacy Endpoints
Studies 005 and 006 (ITT)**

| | Time-to-Success Analysis | | | |
|--|-----------------------------------|----------------|--|----------------|
| Study 005 | Subject Wrinkle Assessment | | Evaluator Wrinkle Severity Assessment | |
| Time to success¹ Cumulative Successes (%) | IT | Vehicle | IT | Vehicle |
| Termination from study ² | 20 | 15 | 20 | 15 |
| Baseline – Visit 3 | 27 (33.8%) | 15 (16.7%) | 8 (10.0%) | 0 |
| – Visit 4 | 40 (50.0%) | 20 (22.2%) | 18 (22.5%) | 4 (4.4%) |
| – Visit 5 | 42 (52.5%) | 22 (24.5%) | 22 (27.5%) | 4 (4.4%) |
| – Visit 6 | 57 (71.3%) | 31 (34.8%) | 33 (41.3%) | 7 (7.9%) |
| Comparison (log-rank test) Treatment | < 0.0001 | | < 0.0001 | |
| Study 006 | Subject Wrinkle Assessment | | Evaluator Wrinkle Severity Assessment | |
| Time-to-success¹ Cumulative Successes (%) | IT | Vehicle | IT | Vehicle |
| Termination from study ² | 17 | 10 | 17 | 10 |
| Baseline – Visit 3 | 24 (25.5%) | 7 (7.1%) | 11 (11.7%) | 1 (1.0%) |
| – Visit 4 | 37 (39.6%) | 13 (13.3%) | 13 (13.9%) | 1 (1.0%) |
| – Visit 5 | 41 (43.9%) | 15 (15.3%) | 18 (19.2%) | 4 (4.1%) |
| – Visit 6 | 50 (53.6%) | 19 (19.4%) | 21 (22.5%) | 8 (8.2%) |
| Comparison (log-rank test) Treatment | < 0.0001 | | 0.0043 | |
| Source: Summary is based on Module 5 – Vol.43 (Table 23 on p.84; Table 26 on p.88) and Vol.52 (Table 23 on p.84; Table 26 on p.87). | | | | |
| ¹ Subjects were censored upon termination from the study in time-to-success analyses. | | | | |
| ² For study 005, 13 were censored during Baseline – Visit 3; and 1 each was censored during Visit 4 – Visit 5 and Visit 5 – Visit 6 in vehicle group. All were censored during Baseline-Visit 3 in IT group. For study 006, 16 were censored during Baseline – Visit 3; and 1 was censored during Visit 3 – Visit 4 in IT group. On the other hand, all were censored during Baseline – Visit 3 in vehicle group. | | | | |

Figure 1: Kaplan-Meier Curves for Time-to-Success (Subject Wrinkle Assessment) – Study 005

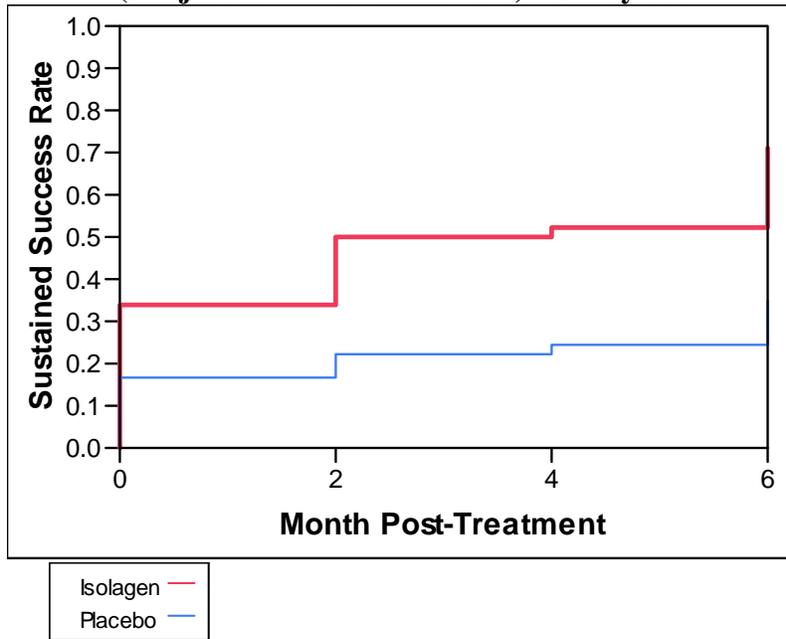


Figure 2: Kaplan-Meier Curves for Time-to-Success (Evaluator Wrinkle Severity Assessment) – Study 005

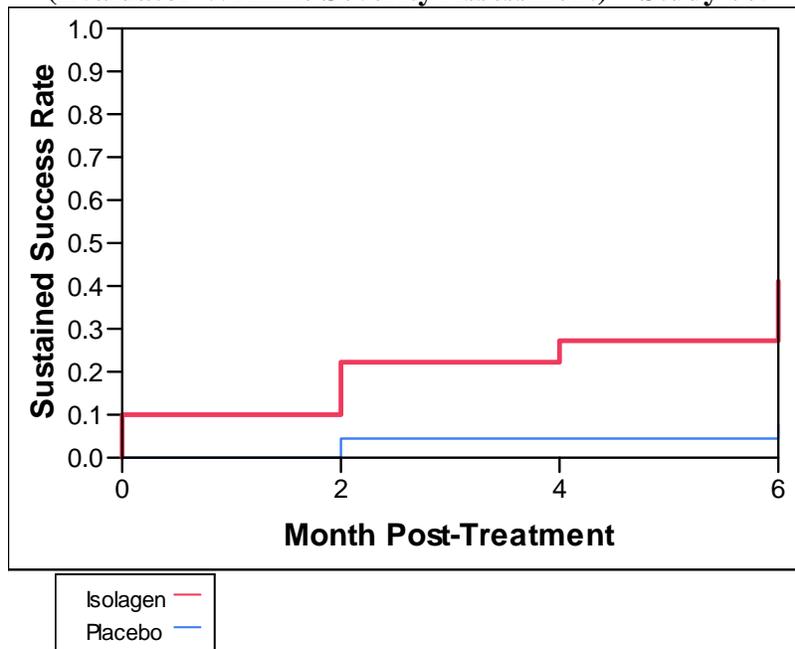


Figure 3: Kaplan-Meier Curves for Time-to-Success (Subject Wrinkle Assessment) – Study 006

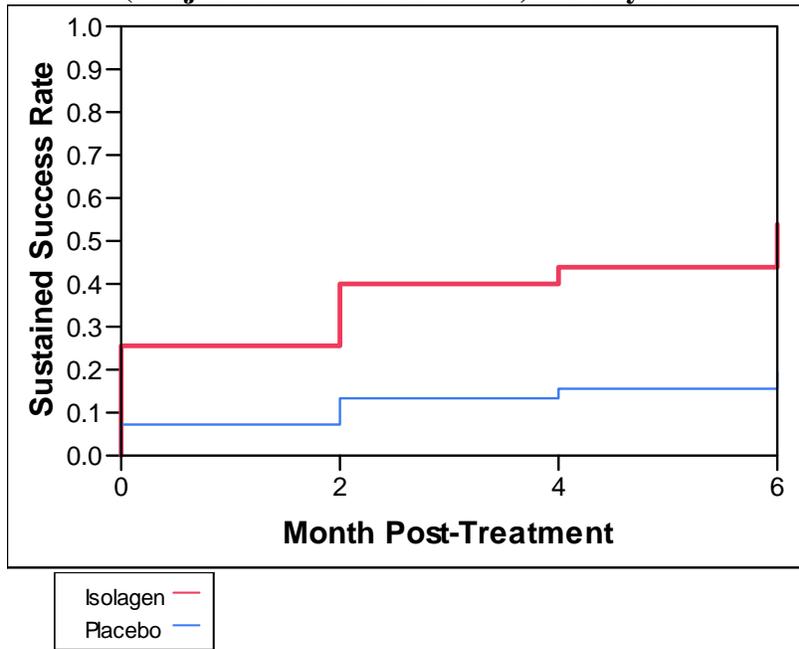


Figure 4: Kaplan-Meier Curves for Time-to-Success (Evaluator Wrinkle Severity Assessment) – Study 006

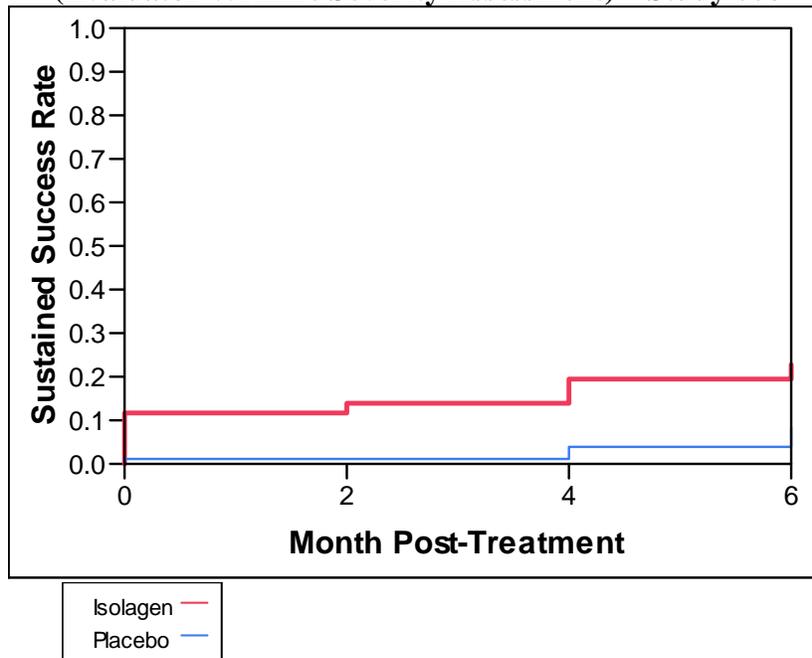


Table A.14: Results* of Co-Primary Efficacy Endpoints by Subgroups – Study 005 (ITT)

| Efficacy by Demographic/Baseline (IT, Vehicle) group size | Subject Wrinkle Assessment | | Evaluator Wrinkle Assessment | |
|--|----------------------------|-------------------|------------------------------|-------------------|
| | IT (n = 100) | Vehicle (n = 103) | IT (n = 100) | Vehicle (n = 103) |
| Overall | 57 (57%) | 31 (30.1%) | 33 (33%) | 7 (6.8%) |
| Age (years) | | | | |
| <= 40 years (4, 3) | 3 (75%) | 1 (33%) | 2 (50%) | 1 (33%) |
| 41 – 50 years (15, 22) | 9 (60%) | 8 (36%) | 5 (33%) | 1 (5%) |
| 51 – 64 years (60, 62) | 34 (57%) | 19 (31%) | 25 (42%) | 3 (5%) |
| >= 65 years (21, 16) | 11 (52%) | 3 (19%) | 1 (5%) | 2 (13%) |
| Gender | | | | |
| Male (12, 9) | 6 (50%) | 4 (44%) | 2 (17%) | 1 (11%) |
| Female (88, 94) | 51 (58%) | 27 (29%) | 31 (35%) | 6 (6%) |
| Race | | | | |
| American Indian/Alaska Native (0, 1) | na | 0 | na | 0 |
| Asian (2, 0) | 1 (50%) | na | 0 | na |
| Black or African American (1, 2) | 1 (100%) | 1 (50%) | 0 | 0 |
| White (94, 99) | 52 (55%) | 29 (29%) | 30 (32%) | 7 (7%) |
| Other (3, 1) | 3 (100%) | 1 (100%) | 3 (100%) | 0 |
| Total Wrinkle Length (cm) | | | | |
| <= 9.0 cm (26, 29) | 17 (65%) | 10 (34%) | 15 (58%) | 4 (14%) |
| > 9.0 to < 12.0 cm (44, 47) | 28 (64%) | 14 (30%) | 11 (25%) | 1 (2%) |
| >= 12.0 cm (30, 27) | 12 (40%) | 7 (26%) | 7 (23%) | 2 (7%) |
| Subject Wrinkle Assessment | | | | |
| Dissatisfied (35, 42) | 12 (34%) | 9 (21%) | 11 (31%) | 2 (5%) |
| Very dissatisfied (65, 61) | 45 (69%) | 22 (36%) | 22 (34%) | 5 (8%) |
| Evaluator Wrinkle Severity Assessment, right | | | | |
| Grade 3 (Moderate deep) (49, 53) | 33 (67%) | 19 (36%) | 17 (35%) | 4 (7.5%) |
| Grade 4 (Deep wrinkle) (39, 38) | 19 (49%) | 9 (24%) | 12 (31%) | 3 (8%) |
| Grade 5 (Very deep wrinkle) (12, 12) | 5 (42%) | 3 (25%) | 4 (33%) | 0 |
| Evaluator Wrinkle Severity Assessment, left | | | | |
| Grade 3 (Moderate deep) (44, 51) | 25 (57%) | 19 (37%) | 14 (32%) | 3 (6%) |
| Grade 4 (Deep wrinkle) (45, 40) | 27 (60%) | 9 (23%) | 16 (36%) | 3 (7.5%) |
| Grade 5 (Very deep wrinkle) (11, 12) | 5 (45%) | 3 (25%) | 3 (27%) | 1 (8%) |
| Source: Summary is based on sponsor’s electronic SAS data sets. | | | | |
| *Missing data are treated as failures. | | | | |

Table A.15: Results* of Co-Primary Efficacy Endpoints by Subgroups – Study 006 (ITT)

| Efficacy by Demographic/Baseline (IT, vehicle) group size | Subject Wrinkle Assessment | | Evaluator Wrinkle Assessment | |
|--|----------------------------|-------------------|------------------------------|-------------------|
| | IT (n = 110) | Vehicle (n = 108) | IT (n = 110) | Vehicle (n = 108) |
| Overall | 50 (45.5%) | 19 (17.6%) | 21 (19.1%) | 8 (7.4%) |
| Age (years) | | | | |
| <= 40 years (14, 4) | 10 (71%) | 1 (25%) | 5 (36%) | 0 |
| 41 – 50 years (25, 30) | 11 (44%) | 2 (7%) | 5 (20%) | 1 (3%) |
| 51 – 64 years (56, 55) | 27 (48%) | 13 (24%) | 8 (14%) | 6 (11%) |
| >= 65 years (15, 19) | 2 (13%) | 3 (16%) | 3 (20%) | 1 (5%) |
| Gender | | | | |
| Male (7, 13) | 4 (57%) | 1 (8%) | 1 (14%) | 1 (8%) |
| Female (103, 95) | 46 (45%) | 18 (19%) | 20 (19%) | 7 (7%) |
| Race | | | | |
| Black or African American (1, 1) | 0 | 0 | 0 | 1 (100%) |
| White (98, 95) | 45 (46%) | 16 (17%) | 17 (17%) | 6 (6%) |
| Other (11, 12) | 5 (45%) | 3 (25%) | 4 (36%) | 1 (8%) |
| Total wrinkle length (cm) | | | | |
| <= 9.0 cm (55, 61) | 27 (49%) | 10 (16%) | 12 (22%) | 7 (11%) |
| > 9.0 to < 12.0 cm (38, 36) | 17 (45%) | 5 (14%) | 6 (16%) | 1 (3%) |
| >= 12.0 cm (17, 11) | 6 (35%) | 4 (36%) | 3 (18%) | 0 |
| Subject Wrinkle Assessment | | | | |
| Dissatisfied (41, 46) | 21 (51%) | 7 (15%) | 10 (24%) | 5 (11%) |
| Very dissatisfied (69, 62) | 29 (42%) | 12 (19%) | 11 (16%) | 3 (5%) |
| Evaluator Wrinkle Severity Assessment, right | | | | |
| Grade 3 (Moderate deep) (55, 45) | 31 (56%) | 11 (24%) | 13 (24%) | 4 (9%) |
| Grade 4 (Deep wrinkle) (41, 53) | 16 (39%) | 6 (11%) | 3 (7%) | 3 (6%) |
| Grade 5 (Very deep wrinkle) (14, 10) | 3 (21%) | 2 (20%) | 5 (36%) | 1 (10%) |
| Evaluator Wrinkle Severity Assessment, left | | | | |
| Grade 3 (Moderate deep) (49, 43) | 28 (57%) | 7 (16%) | 7 (14%) | 3 (7%) |
| Grade 4 (Deep wrinkle) (45, 52) | 18 (40%) | 8 (15%) | 9 (20%) | 4 (8%) |
| Grade 5 (Very deep wrinkle) (16, 13) | 4 (25%) | 4 (31%) | 5 (31%) | 1 (8%) |
| Source: Summary is based on the sponsor's electronic SAS data sets. | | | | |
| *Missing data are treated as failures. | | | | |

