CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

209483Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 209483
Drug Name:	Clobetasol propionate cream, 0.025%
Indication(s):	Moderate to severe plaque psoriasis
Applicant:	Promius Pharma, LLC
Date(s):	Letter Date: January 30, 2017
	PDUFA Date: November 30, 2017
Review Priority:	Standard
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Rebecca Hager, PhD
Concurring Reviewers:	Mohamed Alosh, PhD
Medical Division:	Division of Dermatology and Dental Products
Clinical Team:	Hamid Tabatabai, MD / Snezana Trajkovic, MD
Project Manager:	Angela Brown, MPH

Keywords: plaque psoriasis, superiority trial**,** multiple imputation, randomization

Table of Contents

LIST OF TABLES

1 EXECUTIVE SUMMARY

Promius Pharma, LLC is seeking approval of clobetasol propionate cream, 0.025% (DFD-06 cream) for the topical treatment of moderate to severe plaque psoriasis in patients 18 years of age and older.

The applicant submitted data from two identically-designed, randomized, double-blind, multicenter, vehicle-controlled, parallel-group, pivotal Phase 3 trials (Trials 004 and 005). The trials enrolled subjects 18 years of age and older with a clinical diagnosis of stable (at least 3 months) plaque-type psoriasis; psoriasis involving at least 3% body surface area (BSA), not including the face, scalp, groin, axillae, or other intertriginous areas; and an Investigator's Global Assessment (IGA) score of 3 (moderate) or 4 (severe) at baseline. In both trials, subjects were instructed to apply study product twice daily for 14 days.

The primary endpoint was the proportion of subjects with treatment success at Day 15, where treatment success was defined as an IGA score of 0 (clear) or 1 (almost clear) with at least a 2-grade reduction from baseline. The pre-specified secondary endpoints were the percent change from baseline in BSA at Day 15 and the proportion of subjects with treatment success at Day 8. The results from these endpoints are presented in Table 1. In both trials, DFD-06 cream was statistically superior to vehicle cream for all endpoints.

Table 1: Results of the Primary and Secondary Endpoints for Trials 004 and 005

	Trial 004			Trial 005		
ITT Population[1]	**DFD-06 N=178**	**Vehicle N=89**	**p-value**	**DFD-06 N=176**	**Vehicle N=89**	**p-value**
Primary Endpoint Treatment success at Day 15[2]	30.2%	9.0%	<0.001	30.1%	9.7%	<0.001
Secondary Endpoints % change from baseline in BSA at Day 15[3]	-28.9 (34.0)	-6.1 (32.7)	<0.001	-25.1 (36.6)	-7.2 (19.7)	<0.001
Treatment success at Day 8[2]	15.7%	5.6%	0.006	14.2%	1.6%	0.001

Source: Reviewer's analysis

(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

(2) Treatment success defined as an IGA score of 0 (clear) or 1 (almost clear) with at least a 2-grade reduction from baseline. Missing values were handled using multiple imputation and results were averaged over the 5 imputed data sets. The p-value was calculated from a CMH test for general association adjusted for analysis center and baseline IGA score.

(3) Results presented as mean (standard deviation). Missing values were handled using last observation carried forward. The p-value was calculated from a two-way ANOVA with fixed factors for treatment, analysis center, and baseline IGA score.

The applicant changed the conduct of randomization midway through the trials without informing the Agency. Sensitivity analyses evaluating the impact of this change support the superiority of DFD-06 cream compared to vehicle cream.

2 INTRODUCTION

2.1 Overview

Promius Pharma, LLC is seeking approval of clobetasol propionate cream, 0.025%, hereinafter referred to as DFD-06 cream, for the topical treatment of moderate to severe plaque psoriasis in patients 18 years of age and older. Clobetasol propionate is a corticosteroid available in different formulations such as ointments, creams, gels, foams, shampoos, and solutions. The applicant's product is half the strength of currently available products containing clobetasol propionate, and the applicant states that this is intended to reduce the incidence of adverse events while maintaining efficacy. The applicant submitted data from two pivotal Phase 3 trials (Trials 004 and 005) conducted in the United States (US). An overview of the trials is presented in Table 2.

Table 2: Clinical Study Overview

Trial	Location	Study Population	Treatment Arms	# of Subjects	Dates
DFD-06-CD-004 (N=267)	27 sites in the US	Age ≥ 18 years; diagnosis of stable (≥ 3 months) plaque-type psoriasis; BSA ≥ 3% not including the face, scalp, groin, axillae, or other intertriginous areas; IGA=3 or 4	DFD-06 Cream	178	December 17, 2015 - May 23, 2016
			Vehicle Cream	89	
DFD-06-CD-005 (N=265)	27 sites in the US		DFD-06 Cream	176	November 16, 2015 – May 17, 2016
			Vehicle Cream	89	

2.2 Regulatory History

The applicant opened the IND on April 11, 2011. The Agency sent the applicant Advice/Information Request letters dated April 15, 2011; April 19, 2011; August 8, 2013; and April 24, 2015 and a Study May Proceed letter dated October 18, 2013. The comments in these letters were not relevant to biostatistics. The applicant sent an inactivation request dated July 31, 2012, and a reactivation request dated June 3, 2013.

(b) (4)

The applicant's proposed primary endpoint was the proportion of subjects with treatment success defined as an IGA score of 0 or 1 with at least a 2-grade reduction from baseline at the Day 15 visit.

(b) (4)

(b) (4)

(b) (4)

(b) (4)

On September 9, 2015, the applicant submitted an amended Phase 3 protocol and Statistical Analysis Plan (SAP) for a Special Protocol Assessment (SPA). In a Special Protocol No Agreement letter dated October 16, 2015, the Agency agreed with the applicant on their revised IGA scale along with the following proposals relevant to biostatistics:

(b) (4)

The Agency did not agree with the applicant's plan to (b) (4) (b) (4) The Agency also did not agree that (b) (4) (b) (4)

In response to these comments from the Agency, the applicant revised their Phase 3 protocol and SAP to (b) (4) (b) (4) (b) (4) (b) (4) (b) (4)

(b) (4)

The applicant submitted the NDA on January 30, 2017. There was still lack of clarity about the conduct of the randomization and some of the variables after a preliminary check of the data. The Agency sent an information request (IR) to the applicant on March 3, 2017 asking for clarification on these issues, and the applicant provided responses to the IR on March 10, 2017. The Agency did not find the responses to be adequate. On March 15, 2017, the Agency provided the applicant with additional information and questions that were discussed during a teleconference on March 17, 2017. On March 20, 2017, the applicant provided written responses to the questions discussed during the teleconference and other questions that there was insufficient time to discuss.

During the teleconference and in the written responses that followed, the applicant acknowledged that the conduct of the randomization was changed midway through the studies without informing the Agency. In addition, there were other issues pertaining to randomization which are detailed in Section 3.2.5 of this review.

2.3 Data Sources

This reviewer evaluated the applicant's clinical study reports, datasets, clinical summaries, and proposed labeling. This submission was submitted in eCTD format and was entirely electronic. Both SDTM and analysis datasets were submitted. The datasets used in this review are archived at \\cdsesub1\evsprod\NDA209483\0002\m5\datasets.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The data pertaining to randomization was insufficient in the original submission of the NDA dated January 30, 2017. After several IRs and a teleconference, the applicant submitted modified datasets on March 20, 2017. There were several coding errors in the analysis datasets, so this reviewer recoded most of the analyses from the SDTM data.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

The applicant conducted two identically-designed, randomized, double-blind, multicenter, vehicle-controlled, parallel-group pivotal Phase 3 studies, Trials 004 and 005. Subjects were enrolled into the studies who met the following key inclusion criteria:

- At least 18 years old
- Clinical diagnosis of stable (at least 3 months) plaque-type psoriasis
- Psoriasis involving at least 3% BSA not including the face, scalp, groin, axillae, or other intertriginous areas. The protocol stated, "BSA should be determined using the area of the subject's hand print (palm plus extended fingers) as an estimate of 1% BSA."
- IGA score of 3 (moderate) or 4 (severe) at baseline (Table 3 presents the IGA scale)

Table 3: Investigator's Global Assessment (IGA) of Disease Severity

Score	Grade	Definition
0	Clear	No signs of psoriasis Post-inflammatory hyperpigmentation may be present
1	Almost Clear	No thickening to minimal plaque elevation Normal to slight pink coloration/faint erythema Focal to minimal scaling
2	Mild	Slight elevation/thickening Pink to light red coloration Predominantly fine scaling partially or mostly covering lesions
3	Moderate	Clearly distinguishable/distinct thickening Definite red coloration Coarse scaling covering most plaques
4	Severe	Marked thickening with hard/sharp edges Bright to deep dark red coloration Thick/coarse scaling covering almost all or all lesions

Source: Applicant's Table 9.2 on page 28 of Study Report for Trial 004

Subjects were randomized to DFD-06 cream or vehicle cream in a 2:1 ratio using an interactive web-based response system (IWRS). Prior to March 3, 2016 at 5 p.m. Eastern Time (ET), randomization was stratified by site and baseline IGA score (3, 4). After this time, randomization was stratified only by baseline IGA score. The applicant stated that this change in the conduct of randomization was due to low enrollment in sites of subjects with a baseline IGA score of 4, causing an imbalance of treatment assignments in this stratum. The applicant did not discuss this change with the Agency prior to its implementation. Subjects were instructed to apply the study product twice daily for 14 days, and had visits at baseline (Day 1), Day 4 (±1 day), Day 8 (±2 days), and Day 15 (±3 days).

The primary efficacy endpoint was the proportion of subjects with treatment success at Day 15, where treatment success was defined as an IGA score of 0 (clear) or 1 (almost clear) with at least a 2-grade reduction from baseline. The secondary endpoints were the percent change from baseline in BSA at Day 15 and the proportion of subjects with treatment success at Day 8.

3.2.2 Statistical Methodologies

The protocol stated that the population used for the primary efficacy evaluation was the intent-to-treat (ITT) population, defined as all subjects who were randomized and dispensed study product. The per protocol (PP) population was defined as all subjects in the ITT population who met the following criteria:

- Met all inclusion/exclusion criteria
- Did not take prohibited concomitant medication during the evaluation period
- Completed the Day 15 visit within the allotted window
- Applied 80%-120% of the expected applications of study product during the evaluation period

Subjects who discontinued the study due to documented lack of efficacy, worsening condition, or a treatment-related adverse event (AE) were included in the PP population and considered treatment failures on the IGA for all subsequent visits in the analysis. The applicant stated that the safety population included all subjects who received at least one confirmed dose of study product and provided post-baseline safety information.

The SAP stated that multiple imputation (MI) was used to replace missing IGA scores at each visit. The applicant stated that the variables treatment, site, and "key baseline characteristics such as age, gender, and race" would go into the imputation model. However, the applicant's code showed that none of these variables besides treatment were used in the imputation models. The SAP specified that data missing intermittently was imputed first by the Markov Chain Monte Carlo (MCMC) method, which results in monotone missing data. After the first imputation step, the SAP stated that the logistic regression method was used to produce 5 imputed datasets with the missing IGA scores filled in. However, the applicant's code shows that a linear regression was performed for this imputation step using treatment and the IGA scores from each of the previous visits as covariates. The imputed values were then rounded and trimmed to reflect the domain of IGA scores (i.e. 0, 1, 2, 3, and 4). Last observation carried forward (LOCF) was specified as an alternative method for handling missing IGA scores.

The SAP stated that missing data for BSA was replaced using LOCF. BSA was only evaluated at baseline and the Day 15 visit, so the LOCF method replaces missing BSA data at Day 15 with the baseline evaluation. Thus, a 0% change from baseline in BSA is imputed for subjects with missing BSA data at Day 15. No alternative method for handling missing BSA data was specified.

Analysis centers were defined as adequately large original sites and pooled smaller sites. The protocol stated that a site was considered adequately large if there were at least 8 ITT subjects in the DFD-06 arm and 4 ITT subjects in the vehicle arm. Smaller sites were pooled from biggest to smallest until the pooled center was adequately large. If the last few small sites were pooled and failed to be adequately large, then they were pooled with the smallest analysis center.

The SAP specified that the primary endpoint, the proportion of subjects with treatment success at Day 15, was analyzed with a Cochran-Mantel-Haenszel (CMH) test for general association

stratified by analysis center and baseline IGA score. The Breslow-Day test was conducted to assess homogeneity of the treatment effect across analysis centers.

The proportion of subjects with treatment success at Day 8 was also analyzed with a CMH test for general association stratified by analysis center and baseline IGA score. The percent change from baseline in BSA at Day 15 was analyzed using a two-way analysis of variance (ANOVA) model with factors for treatment group, analysis center, and baseline IGA score.

The primary endpoint was tested at an alpha level of 0.05. The SAP stated, "If the primary efficacy endpoint analysis at Day 15 achieves $p < 0.05$, statistical significance testing for the secondary efficacy endpoints will use a step-down approach, according to a pre-specified order: percent change in BSA at Day 15, then treatment success at Day 8."

3.2.3 Patient Disposition, Demographic, and Baseline Characteristics

Trial 004 enrolled and randomized a total of 267 subjects, 178 to DFD-06 cream and 89 to vehicle cream, from 27 sites in the US. Trial 005 enrolled and randomized a total of 265 subjects, 176 to DFD-06 cream and 89 to vehicle cream, from 27 sites in the US. Table 4 presents the reasons for discontinuation from the studies. There were very few discontinued subjects in both trials.

Table 4: Disposition of Subjects

	Trial 004		Trial 005	
ITT Population[(1)]	**DFD-06 N=178**	**Vehicle N=89**	**DFD-06 N=176**	**Vehicle N=89**
Discontinued	2 (1.1%)	1 (1.1%)	1 (0.6%)	4 (4.5%)
Subject decision/withdrawal of consent	0 (0.0%)	0 (0.0%)	1 (0.6%)	2 (2.3%)
A treatment-related AE occurred	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.1%)
Non-treatment-related AE occurred	1 (0.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lost to follow-up	1 (0.6%)	1 (1.1%)	0 (0.0%)	1 (1.1%)

Source: Reviewer's analysis (same as applicant's analysis)
(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

The demographics of the trial subjects are presented in Table 5. The demographics were generally balanced across the treatment arms within each trial and across both trials; however, there was a higher proportion of males in the DFD-06 arms in both trials, with Trial 005 having a higher proportion of males than Trial 004.

The baseline disease characteristics are presented in Table 6. Baseline IGA scores were similar across treatment arms in the two trials; however, Trial 005 had a higher proportion of subjects with a severe (4) baseline IGA score than in Trial 004. In both trials, the mean baseline BSA was slightly higher in the vehicle arm than in the DFD-06 arm.

Table 5: Subject Demographics

ITT Population[1]	Trial 004 DFD-06 N=178	Trial 004 Vehicle N=89	Trial 005 DFD-06 N=176	Trial 005 Vehicle N=89
Age (years)				
Mean (SD)	49.5 (14.8)	49.9 (14.3)	49.5 (13.6)	50.6 (15.9)
Median	50	53	51	50
Range	(18, 79)	(18, 82)	(20, 78)	(19, 79)
Sex				
Male	100 (56.2%)	44 (49.4%)	111 (63.1%)	51 (57.3%)
Female	78 (43.8%)	45 (50.6%)	65 (36.9%)	38 (42.7%)
Race				
White or Caucasian	149 (83.7%)	74 (83.2%)	146 (83.0%)	79 (88.8%)
Black or African American	18 (10.1%)	9 (10.1%)	15 (8.5%)	2 (2.3%)
Asian	7 (3.9%)	3 (3.4%)	2 (1.1%)	2 (2.3%)
American Indian or Alaskan Native	1 (0.6%)	1 (1.1%)	5 (2.8%)	1 (1.1%)
Native Hawaiian or Other Pacific Islander	1 (0.6%)	1 (1.1%)	0 (0.0%)	1 (1.1%)
Other	2 (1.1%)	1 (1.1%)	8 (4.6%)	4 (4.5%)
Ethnicity				
Hispanic or Latino	39 (21.9%)	25 (28.1%)	41 (23.3%)	21 (23.6%)
Not Hispanic or Latino	139 (78.1%)	64 (71.9%)	134 (76.1%)	67 (75.3%)
Unwilling to Provide	0 (0.0%)	0 (0.0%)	1 (0.6%)	1 (1.1%)

Source: Reviewer's analysis (similar to applicant's analysis)
(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

Table 6: Baseline Disease Characteristics

ITT Population[1]	Trial 004 DFD-06 N=178	Trial 004 Vehicle N=89	Trial 005 DFD-06 N=176	Trial 005 Vehicle N=89
IGA				
Moderate (3)	154 (86.5%)	77 (86.5%)	142 (80.7%)	72 (80.9%)
Severe (4)	24 (13.5%)	12 (13.5%)	34 (19.3%)	17 (19.1%)
BSA (%)				
Mean (SD)	6.8 (8.1)	8.8 (11.5)	8.7 (10.2)	9.2 (11.3)
Median	5.0	5.0	5.0	5.0
Range	(3.0, 80.0)	(3.0, 80.0)	(3.0, 80.0)	(3.0, 80.0)

Source: Reviewer's analysis (same as applicant's analysis)
(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

3.2.4 Efficacy Results

The primary endpoint was treatment success at Day 15, where treatment success was defined as an IGA score of 0 or 1 with at least a 2-grade reduction from baseline. The secondary endpoints were the percent change in BSA from baseline to Day 15 and treatment success at Day 8. The primary method for handling missing data for IGA scores was MI, and the primary method for handling missing data for BSA was LOCF. Table 7 presents the results from the primary analyses of the efficacy endpoints on the ITT population.

Percent change in BSA is related to the magnitude of BSA at baseline, i.e., when the baseline BSA is small, a minor absolute change can translate to a large percent change. The mean absolute change (standard deviation) from baseline in BSA at Day 15 was -1.8 (3.3) in the DFD-06 arm and -0.5 (2.4) in the vehicle arm of Trial 004, and similarly, -2.1 (5.0) and -0.4 (1.7) in

Trial 005 with a statistically significant p-value in both trials. It is unclear whether an approximate difference in reduction of absolute BSA of 1.5 is clinically meaningful, as this difference may be less than the investigator's precision in the evaluation of BSA.

Table 7: Efficacy Results for Trials 004 and 005

	Trial 004			Trial 005		
ITT Population[1]	**DFD-06 N=178**	**Vehicle N=89**	**p-value**	**DFD-06 N=176**	**Vehicle N=89**	**p-value**
Primary Endpoint Treatment success at Day 15[2]	30.2%	9.0%	<0.001	30.1%	9.7%	<0.001
Secondary Endpoints % change from baseline in BSA at Day 15[3]	-28.9 (34.0)	-6.1 (32.7)	<0.001	-25.1 (36.6)	-7.2 (19.7)	<0.001
Treatment success at Day 8[2]	15.7%	5.6%	0.006	14.2%	1.6%	0.001

Source: Reviewer's analysis
(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.
(2) Treatment success defined as an IGA score of 0 (clear) or 1 (almost clear) with at least a 2-grade reduction from baseline. Missing values were handled using multiple imputation and results were averaged over the 5 imputed data sets. The p-value was calculated from a CMH test for general association adjusted for analysis center and baseline IGA score.
(3) Results presented as mean (standard deviation). Missing values were handled using last observation carried forward. The p-value was calculated from a two-way ANOVA with fixed factors for treatment, analysis center, and baseline IGA score.

The results in Table 7 differ slightly from the results the applicant submitted. The applicant coded the baseline BSA flag incorrectly for several subjects in Trial 005, so the vehicle response rate when this is coded correctly differs from the applicants results (-7.4 [20.3]). The p-values for treatment success at Day 8 also differ from the applicant's results as the applicant did not properly transform the CMH test statistics prior combining the p-values from the multiply imputed datasets.

Table 8 presents the number of subjects who had a missing IGA score within the pre-specified window of each visit. There were not many subjects with a missing IGA score at the Day 15 or Day 8 visits at which the primary and secondary endpoints were evaluated.

Table 8: Number of Missing IGA Scores by Visit

	Trial 004		Trial 005	
ITT Population[1]	**DFD-06 N=178**	**Vehicle N=89**	**DFD-06 N=176**	**Vehicle N=89**
Baseline	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Day 4 (±1 day)	19 (10.7%)	4 (4.5%)	13 (7.4%)	4 (4.5%)
Day 8 (±2 days)	3 (1.7%)	1 (1.1%)	3 (1.7%)	5 (5.6%)
Day 15 (±3 days)	6 (3.4%)	3 (3.4%)	5 (2.8%)	6 (6.7%)

Source: Reviewer's analysis
(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

As a sensitivity analysis, the applicant used LOCF to handle missing IGA scores. The applicant did not conduct MI in the manner pre-specified in the SAP, so this reviewer also conducted an additional analysis where all missing IGA scores at Day 15 were assumed to be a treatment failure, called missing value treated as failure (MVTF). As a supportive analysis, the applicant analyzed the primary endpoint for the PP population. The PP population in Trial 005 included one subject who had a treatment-related AE and was therefore considered a treatment failure

after Day 4 according to the plan pre-specified by the applicant. The results of the sensitivity analyses for the primary endpoint are presented in Table 9, and the results are consistent with the primary analysis.

Table 9: Sensitivity Analyses for the Primary Endpoint

	Trial 004			Trial 005		
	DFD-06	**Vehicle**	**p-value**(1)	**DFD-06**	**Vehicle**	**p-value**(1)
LOCF (ITT)(2)	53/178 (29.8%)	8/89 (9.0%)	<0.001	53/176 (30.1%)	8/89 (9.0%)	<0.001
MVTF (ITT)(3)	52/178 (29.2%)	8/89 (9.0%)	<0.001	53/176 (30.1%)	8/89 (9.0%)	<0.001
PP Population(4)	48/153 (31.4%)	8/81 (9.9%)	<0.001	50/163 (30.7%)	8/77 (10.4%)	<0.001

Source: Reviewer's analysis
(1) The p-value was calculated from a CMH test for general association adjusted for analysis center and baseline IGA score.
(2) Missing data was handled using last observation carried forward (LOCF). The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.
(3) Missing data was handled using missing value treated as failure (MVTF). The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.
(4) The per protocol (PP) population was analyzed, and no method for handling missing data was used.

Table 10 presents the results of the sensitivity analyses for the secondary endpoints. The same sensitivity analyses were conducted for treatment success at Day 8 as were performed for the primary endpoint. If the IGA score was missing at Day 8 for a subject in the PP population, the applicant states that this was analyzed as a treatment failure. The percent change from baseline in BSA at Day 15 was analyzed using the PP population. No other sensitivity analyses for that endpoint were pre-specified by the applicant. The mean absolute change (standard deviation) from baseline in BSA at Day 15 was -1.9 (3.5) in the DFD-06 arm and -0.6 (2.3) in the vehicle arm of Trial 004, and similarly, -1.8 (4.4) and -0.5 (1.9) in Trial 005. The results of the sensitivity analyses are consistent with the primary analyses.

Table 10: Sensitivity Analyses for the Secondary Endpoints

	Trial 004			Trial 005		
	DFD-06	**Vehicle**	**p-value**	**DFD-06**	**Vehicle**	**p-value**
Treatment success at Day 8(1)						
LOCF (ITT) (2)	30/178 (16.9%)	5/89 (5.6%)	0.002	25/176 (14.2%)	1/89 (1.1%)	<0.001
MVTF (ITT)(3)	29/178 (16.3%)	5/89 (5.6%)	0.003	23/176 (13.1%)	1/89 (1.1%)	<0.001
PP Population (4)	27/153 (17.7%)	5/81 (6.2%)	0.006	22/163 (13.5%)	1/77(1.3%)	0.001
% change in BSA at Day 15(5)	**N=153**	**N=81**		**N=163**	**N=77**	
PP Population	-31.1 (35.0)	-9.3 (25.0)	<0.001	-25.2 (37.3)	-8.0 (20.9)	<0.001

Source: Reviewer's analysis
(1) The p-value was calculated from a CMH test for general association adjusted for analysis center and baseline IGA score.
(2) Missing data was handled using last observation carried forward (LOCF). The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.
(3) Missing data was handled using missing value treated as failure (MVTF). The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.
(4) The per protocol (PP) population was analyzed, and missing IGA scores at Day 8 were treated as failures.
(5) Results presented as mean (standard deviation). The p-value was calculated from a two-way ANOVA with fixed factors for treatment, analysis center, and baseline IGA score.

The LOCF results for the DFD-06 arm in Trial 004 in Tables 9 and 10 differ slightly from the results submitted by the applicant. If a subject had 2 visits in the same visit window, the applicant used the data from the earlier visit. The results in this review use the data from the later visit in the window, as this is more reasonable when performing LOCF. The applicant only

included 152 subjects from the DFD-06 arm in the analysis of the percent change in BSA endpoint for the PP population in Trial 004. The definition of the PP population is the same regardless of what endpoint is being analyzed, so this review includes all 153 subjects in the analyses of the PP population.

3.2.5 Randomization Issue

There was lack of clarity about the conduct of randomization in the IND stage and after the initial NDA submission. During a teleconference with the Agency on March 17, 2017, the applicant acknowledged that the conduct of the randomization was changed midway through the trials without informing the Agency. This section will describe how the randomization was conducted and present sensitivity analyses to support the findings from the primary analysis.

During the teleconference on March 17, 2017, the applicant stated that randomization was originally stratified by both site and baseline IGA score (3, 4). Prior to any changes in randomization, about 80% of subjects had a baseline IGA score of 3 and 20% had a baseline IGA score of 4. Due to the small number of subjects enrolled with a baseline IGA score of 4, low enrollment in sites, and the 2:1 randomization ratio, the applicant stated that there was unbalanced treatment allocation in the stratum where baseline IGA equals 4. Therefore, on March 3, 2016 at 5 p.m. ET, the randomization was changed to only be stratified by baseline IGA score (3, 4) and no longer stratified by site.

This reviewer understands the randomization as follows based on the randomization list and other written materials submitted by the applicant to the Agency on March 20, 2017:

1. Prior to March 3, 2016 at 5 p.m. ET, subjects were randomized in a 2:1 ratio to DFD-06 cream or vehicle cream stratified by both site and baseline IGA score (3, 4). Within each stratum, treatment was assigned in blocks of size 3 with 2 subjects receiving DFD-06 cream and 1 receiving vehicle cream. Any ordering of treatments within the block was possible.
2. On March 3, 2016 at 5 p.m. ET, the randomization scheme was changed. There was a "catch up" period when treatment assignment was balanced within the baseline IGA score strata so that the 2:1 allocation ratio was met study-wide. This resulted in long strings of entering subjects who were assigned the same treatment.
3. After the 2:1 ratio was met within a baseline IGA stratum study-wide, randomization continued in a 2:1 ratio stratified only by baseline IGA score. The 2:1 ratio was met first in the baseline IGA 3 stratum, and for this stratum, randomization proceeded in blocks of 3 and any ordering of treatments within the block was possible.
4. During the "catch up" period in the baseline IGA 4 stratum, the applicant's submitted randomization lists state that there was an issue identified with rebalance and the IWRS system was updated. This occurred on March 17, 2016 at 4:04 p.m. ET in Trial 004 and on March 14, 2016 at 12:24 p.m. ET in Trial 005. After this change and once the 2:1 ratio was met within the both baseline IGA strata study-wide, randomization still occurred in blocks of 3; however, it appears that not all treatment orderings were possible. The ordering (DFD-06, vehicle, DFD-06) and (vehicle, DFD-06, DFD-06) occurred, but not (DFD-06, DFD-06, vehicle). This was due to the coding of the randomization algorithm which the applicant submitted to the Agency on March 20, 2017.

As there are multiple issues with the conduct of the randomization in both studies, this reviewer performed several sensitivity analyses. First, subjects who were randomized prior to and after March 3, 2016 at 5 p.m. ET were examined separately. Baseline disease characteristics for these subjects are presented in Table 11. In Trial 004, there was a larger proportion of subjects after the randomization change that had a baseline IGA score of 4 in the treatment arm, and the opposite trend was observed in the vehicle arm. Trends in Trial 005 went in the opposite direction as those in Trial 004.

Table 11: Sample Sizes by Randomization Period and Baseline IGA Score

	Trial 004		Trial 005	
	DFD-06	**Vehicle**	**DFD-06**	**Vehicle**
ITT Population[1]	**N=178**	**N=89**	**N=176**	**N=89**
Baseline IGA 3	154 (86.5%)	77 (86.5%)	142 (80.7%)	72 (80.9%)
Baseline IGA 4	24 (13.5%)	12 (13.5%)	34 (19.3%)	17 (19.1%)
Prior to Change[2]	**N=81**	**N=43**	**N=119**	**N=56**
Baseline IGA 3	74 (91.4%)	36 (83.7%)	90 (75.6%)	47 (83.9%)
Baseline IGA 4	7 (8.6%)	7 (16.3%)	29 (24.4%)	9 (16.1%)
After Change[3]	**N=97**	**N=46**	**N=57**	**N=33**
Baseline IGA 3	82 (84.5%)	42 (91.3%)	52 (91.2%)	25 (75.8%)
Baseline IGA 4	15 (15.5%)	4 (8.7%)	5 (8.8%)	8 (24.2%)

Source: Reviewer's analysis
(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.
(2) Prior to Change includes subjects randomized prior to March 3, 2016 at 5 p.m. ET.
(3) After Change includes subjects randomized after March 3, 2016 at 5 p.m. ET.

The efficacy endpoints were re-analyzed for subjects who were randomized before the initial randomization change on March 3, 2016 at 5 p.m. ET. These subjects were randomized into the studies prior to any randomization changes, but statistical tests applied to this subset have lower power due to the reduced sample size. Results of the analyses are presented in Table 12 along with efficacy summary statistics for subjects who were randomized after March 3, 2016 at 5 p.m. ET. As there were multiple changes that occurred after the initial randomization change, statistical testing was not performed on this subset as the treatment assignment scheme was not consistent.

The results in Table 12 have the same trends as the primary analyses in Table 7. When examining only subjects who were enrolled prior to any randomization changes, the endpoints are all still statistically significant except for the secondary endpoint of treatment success at Day 8 in Trial 004. This could be due to the decreased power resulting from the smaller sample size.

The treatment success rates at Day 15 and percent change from baseline in BSA at Day 15 were lower for the subjects who entered the trials after the randomization change compared to those subjects who entered before the change. In Trial 004, the effect sizes (vehicle response rate subtracted from the DFD-06 response rate) for treatment success at Day 8 and Day 15 are approximately the same in both parts of the study; however, the effect sizes in Trial 005 differ for both of these endpoints with a higher treatment effect in the second part of the study. This could be the result of analyzing smaller sample sizes which results in more unreliable estimates of the treatment effect.

Table 12: Efficacy Results by Randomization Period

	Trial 004			Trial 005		
	DFD-06	**Vehicle**	**p-value**	**DFD-06**	**Vehicle**	**p-value**
Treatment success at Day 15[(1)]						
ITT Population[(2)]	30.2%	9.0%	<0.001	30.1%	9.7%	<0.001
Prior to change[(3)]	35.6%	14.0%	0.006	31.1%	15.0%	0.006
After change[(4)]	25.8%	4.4%	-	28.1%	0.6%	-
% change from baseline in BSA at Day 15[(5)]						
ITT Population[(2)]	-28.9 (34.0)	-6.1 (32.7)	<0.001	-25.1 (36.6)	-7.2 (19.7)	<0.001
Prior to change[(3)]	-30.9 (34.0)	-13.2 (27.9)	<0.001	-27.1 (31.4)	-8.7 (20.6)	<0.001
After change[(4)]	-27.3 (34.1)	-0.5 (35.6)	-	-20.8 (45.6)	-4.7 (17.9)	-
Treatment success at Day 8[(1)]						
ITT Population[(2)]	15.7%	5.6%	0.006	14.2%	1.6%	0.001
Prior to change[(3)]	16.1%	7.0%	0.095	11.8%	2.5%	0.025
After change[(4)]	15.5%	4.4%	-	19.3%	0%	-

Source: Reviewer's analysis

(1) Missing values were handled using multiple imputation and results were averaged over the 5 imputed data sets. The p-value was calculated from a CMH test for general association adjusted for analysis center and baseline IGA score.

(2) Sample sizes in Trial 004 were (N_D, N_V)=(178, 89), and sample sizes in Trial 005 were (N_D, N_V)=(176, 89) where N_D = subgroup sample size in the DFD-06 arm and N_V = subgroup sample size in the vehicle arm. The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

(3) Sample sizes in Trial 004 were (N_D, N_V)=(81, 43), and sample sizes in Trial 005 were (N_D, N_V)=(119, 56). Prior to Change includes subjects randomized prior to March 3, 2016 at 5 p.m. ET.

(4) Sample sizes in Trial 004 were (N_D, N_V)=(97, 46), and sample sizes in Trial 005 were (N_D, N_V)=(57, 33). After Change includes subjects randomized after March 3, 2016 at 5 p.m. ET.

(5) Results presented as mean (standard deviation). Missing values were handled using last observation carried forward. The p-value was calculated from a two-way ANOVA with fixed factors of treatment, analysis center, and baseline IGA score.

Table 13 further breaks down the results of the primary endpoint by both randomization period and baseline IGA score. In Trial 004, the treatment effect for subjects with baseline IGA equal to 3 appears to be similar both before and after the randomization change, though the response rates decreased by approximately 10% in both treatment arms after the randomization change. It is difficult to make conclusions for subjects with a baseline IGA score of 4 due to the small sample sizes, though none of those 12 subjects in the vehicle arm obtained treatment success at Day 15 in Trial 004.

In Trial 005, the response rates in the DFD-06 arm for subjects with baseline IGA equal to 3 appeared to be similar both before and after the randomization change; however, the treatment effect size doubled after the change. This may be partially due to smaller sample sizes in the vehicle arm after the randomization change. It is again difficult to make conclusions about subjects with baseline IGA equal to 4 in Trial 005 as there are very few subjects in this group that were randomized after the change. Another sensitivity analysis of the primary endpoint in the form of a simulation study is described in the Appendix.

Table 13: Treatment Success at Day 15[(1)] by Randomization Period and Baseline IGA Score

	Trial 004				Trial 005			
	(N_D, N_V)	DFD-06	Vehicle	Trt Effect[(5)]	(N_D, N_V)	DFD-06	Vehicle	Trt Effect[(5)]
ITT Population[(2)]	**(178, 89)**	**30.2%**	**9.0%**	**21.2%**	**(176, 89)**	**30.1%**	**9.7%**	**20.4%**
Baseline IGA 3	(154, 77)	33.6%	10.4%	23.2%	(142, 72)	31.0%	10.6%	20.4%
Baseline IGA 4	(24, 12)	8.3%	0%	8.3%	(34, 17)	26.5%	5.9%	20.6%
Prior to Change[(3)]	**(81, 43)**	**35.6%**	**14.0%**	**21.6%**	**(119, 56)**	**31.1%**	**15.0%**	**16.1%**
Baseline IGA 3	(74, 36)	38.1%	16.7%	21.4%	(90, 47)	31.1%	15.7%	15.4%
Baseline IGA 4	(7, 7)	12.5%	0%	12.5%	(29, 9)	31.0%	11.1%	19.9%
After Change[(4)]	**(97, 46)**	**25.8%**	**4.4%**	**21.4%**	**(57, 33)**	**28.1%**	**0.6%**	**27.5%**
Baseline IGA 3	(82, 42)	29.6%	4.9%	24.7%	(52, 25)	30.8%	0.8%	30.0%
Baseline IGA 4	(15, 4)	6.3%	0%	6.3%	(5, 8)	0%	0%	0%

Source: Reviewer's analysis
(1) Missing values were handled using multiple imputation and results were averaged over the 5 imputed data sets
(2) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.
(3) Prior to Change includes subjects randomized prior to March 3, 2016 at 5 p.m. ET.
(4) After Change includes subjects randomized after March 3, 2016 at 5 p.m. ET.
(5) Trt Effect equals the vehicle response rate subtracted from the DFD-06 response rate

3.3 Evaluation of Safety

3.3.1 Extent of Exposure

The extent of exposure to the study product is presented in Table 14. The number of applications, days of exposure, and compliance with dosing were similar across treatment arms and trials.

Table 14: Extent of Exposure to Study Product

	Trial 004		Trial 005	
Safety Population[(1)]	**DFD-06 N=175**	**Vehicle N=87**	**DFD-06 N=176**	**Vehicle N=85**
Number of Applications				
Mean (SD)	28.6 (2.9)	27.7 (3.1)	28.2 (3.0)	28.1 (3.5)
Median	29	28	28	28
Range	(19, 38)	(14, 36)	(5, 36)	(6, 38)
Days of Exposure				
<11 days	1 (0.6%)	1 (1.2%)	1 (0.6%)	1 (1.2%)
11 to <18 days	162 (92.6%)	83 (95.4%)	168 (95.5%)	79 (92.9%)
≥ 18 days	12 (6.9%)	3 (3.5%)	7 (4.0%)	5 (5.9%)
Compliant with Dosing Regimen[(2)]				
Yes	154 (88.0%)	81 (93.1%)	165 (93.8%)	79 (92.9%)
No	21 (12.0%)	6 (6.9%)	11 (6.3%)	6 (7.1%)

Source: Reviewer's analysis (same results as applicant's analysis)
(1) The safety population included all subjects who received at least one confirmed dose of study product and provided post-baseline safety information. The applicant stated that subjects with missing treatment end date were not analyzed.
(2) A subject was compliant with the dosing regimen if the subject applied between 80%-120% of the expected applications during the evaluation period.

3.3.2 Adverse Events

Table 15 presents the treatment-emergent adverse events (TEAE) that were reported by at least 1% of subjects within any treatment group in either trial, and for which the incidence was higher in the DFD-06 arm in at least one of the trials.

Table 15: Treatment-Emergent Adverse Events (TEAE) Reported by ≥1% of Subjects with a Higher Incidence in the Active Arm Within Any Treatment Group and Trial

	Trial 004		Trial 005	
Safety Population[(1)]	**DFD-06 N=178**	**Vehicle N=89**	**DFD-06 N=176**	**Vehicle N=89**
Gastrointestinal disorders				
Vomiting	0	0	2 (1.1%)	0
General disorders and administration site conditions				
Application site atrophy	2 (1.1%)	0	0	0
Application site discoloration	3 (1.7%)	2 (2.2%)	4 (2.3%)	0
Application site telangiectasia	2 (1.1%)	0	0	0
Respiratory, thoracic, and mediastinal disorders				
Oropharyngeal pain	0	0	2 (1.1%)	0
Vascular disorder				
Hypertension	0	0	2 (1.1%)	0

Source: Applicant's Study Reports

(1) The safety population included all subjects who received at least one confirmed dose of study product and provided post-baseline safety information.

One or more TEAEs were reported for 14.4% of subjects in the DFD-06 arms and 24.2% of subjects in the vehicle arms across both Phase 3 trials.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Age, Race, Ethnicity, and Baseline Disease Severity

Table 16 presents the results for the primary endpoint by gender, age (<65, ≥65), race (white, non-white), ethnicity, and baseline IGA score (3, 4). Any possible differences in the response rate could not be determined as observed differences could be attributed to small sample sizes.

Table 16: Treatment Success at Day 15 by Gender, Age, Race, Ethnicity, and Baseline Disease Severity

	Trial 004			Trial 005		
ITT Population[1]	**(N_D, N_V)[2] (178, 89)**	**DFD-06**	**Vehicle**	**(N_D, N_V)[2] (176, 89)**	**DFD-06**	**Vehicle**
Gender						
Male	(100, 44)	30.8%	9.1%	(111, 51)	30.6%	12.2%
Female	(78, 45)	29.5%	8.9%	(65, 38)	29.2%	6.3%
Age						
<65 years of age	(151, 75)	28.9%	10.7%	(152, 70)	27.6%	6.6%
≥65 years of age	(27, 14)	37.8%	0.0%	(24, 19)	45.8%	21.1%
Race						
White	(149, 74)	31.3%	6.8%	(146, 79)	30.8%	8.4%
Non-White	(29, 15)	24.8%	20.0%	(30 ,10)	26.7%	20.0%
Ethnicity						
Hispanic or Latino	(39, 25)	41.0%	12.0%	(41, 21)	29.3%	12.4%
Not Hispanic or Latino[3]	(139, 64)	27.2%	7.8%	(135, 68)	30.4%	8.8%
Baseline IGA						
Moderate (3)	(154, 77)	33.6%	10.4%	(142, 72)	31.0%	10.6%
Severe (4)	(24, 12)	8.3%	0.0%	(34, 17)	26.5%	5.9%

Source: Reviewer's analysis

(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product. Missing values were handled using multiple imputation and results were averaged over the 5 imputed data sets.

(2) N_D = subgroup sample size in the DFD-06 arm and N_V = subgroup sample size in the vehicle arm.

(3) Includes 1 "Unwilling to Provide" patient in Trial 005.

4.2 Center

The applicant performed a Breslow-Day test to assess homogeneity of the treatment effect for the primary endpoint across analysis centers. This test was significant at the 0.10 level for Trial 005, but not significant for Trial 004. As pooling could mask differences among sites, Tables 17 and 18 present the sample sizes and treatment success rates at Day 15 averaged over the 5 multiple imputation data sets for the original sites which enrolled more than 10 subjects. None of the sites appear to drive the efficacy results.

Table 17: Treatment Success at Day 15 by Site for Trial 004

	Sample Size			Treatment Success[1]		Missing Data	
Site ID	Total	DFD-06	Vehicle	DFD-06	Vehicle	DFD-06	Vehicle
103	24	15	9	48.0%	0.0%	1	
101	20	17	3	47.1%	0.0%		1
112	20	13	7	69.2%	57.1%		
113	19	16	3	18.8%	0.0%		
118	17	12	5	35.0%	0.0%	2	1
122	17	11	6	9.1%	0.0%	1	
117	14	8	6	25.0%	0.0%		
124	14	9	5	0.0%	0.0%		
129	14	9	5	22.2%	40.0%		
121	13	9	4	22.2%	0.0%		1
123	12	7	5	57.1%	20.0%		
ITT Population[2]	**267**	**178**	**89**	**30.2%**	**9.0%**	**6**	**3**

Source: Reviewer's analysis
(1) Missing values were handled using multiple imputation and results were averaged over the 5 imputed data sets.
(2) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

Table 18: Treatment Success at Day 15 by Site for Trial 005

	Sample Size			Treatment Success [1]		Missing Data	
Site ID	Total	DFD-06	Vehicle	DFD-06	Vehicle	DFD-06	Vehicle
106	24	15	9	46.7%	0.0%		1
120	22	16	6	25.0%	3.3%		2
123	20	15	5	6.7%	8.0%	1	1
115	18	12	6	0.0%	16.7%	1	
113	14	10	4	50.0%	0.0%		
110	12	8	4	0.0%	0.0%		1
124	12	8	4	50.0%	75.0%		
128	12	7	5	42.9%	20.0%	1	
104	11	7	4	28.6%	25.0%		
130	11	8	3	87.5%	66.7%		
102	10	8	2	50.0%	0.0%		
109	10	7	3	14.3%	0.0%		
118	10	8	2	25.0%	0.0%		
119	10	8	2	37.5%	0.0%		
129	10	4	6	25.0%	0.0%	1	
ITT Population[2]	**265**	**176**	**89**	**30.1%**	**9.7%**	**5**	**6**

Source: Reviewer's analysis
(1) Missing values were handled using multiple imputation and results were averaged over the 5 imputed data sets.
(2) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

The applicant changed the conduct of randomization midway through the trials without informing the Agency. This brought into question whether the change in the randomization scheme affected the efficacy results. When examining the efficacy results of the trials prior to the change, treatment success at Day 15 and percent change from baseline in BSA at Day 15 remained significant in both trials. The secondary endpoint treatment success at Day 8 remained significant in only one trial, but this may be attributed to lower power to detect a treatment difference due to a smaller sample size.

Treatment success at Day 15 was further examined by both randomization period and baseline IGA score. In subjects with a baseline IGA score of 3 in Trial 004, the treatment effect was similar across the randomization periods, but the response rates decreased by 10% after the randomization change. In subjects with a baseline IGA score of 3 in Trial 005, the treatment effect size almost doubled after the randomization change, but the response rates in the DFD-06 arm were similar across the randomization periods. Conclusions about subjects enrolled with a baseline IGA score of 4 are limited due to small sample sizes. Additional sensitivity analyses presented in the Appendix support the overall primary efficacy endpoint results.

5.2 Collective Evidence

The applicant submitted results from two identically-designed, randomized, double-blind, multicenter, vehicle-controlled, parallel-group pivotal Phase 3 trials (Trials 004 and 005). The trials enrolled subjects 18 years of age and older with a clinical diagnosis of stable (at least 3 months) plaque-type psoriasis; psoriasis involving at least 3% BSA, not including the face, scalp, groin, axillae, or other intertriginous areas; and an IGA score of 3 (moderate) or 4 (severe) at baseline.

The primary endpoint was treatment success at Day 15, where treatment success was defined as an IGA score of 0 (clear) or 1 (almost clear) with at least a 2-grade reduction from baseline. The pre-specified secondary endpoints were the percent change in BSA from baseline to Day 15 and treatment success at Day 8. Table 19 presents the results of the primary and secondary endpoints. In both trials, DFD-06 cream was statistically superior to vehicle cream for all endpoints.

Table 19: Collective Efficacy Results from Trials 004 and 005

	Trial 004			Trial 005		
ITT Population[1]	**DFD-06 N=178**	**Vehicle N=89**	**p-value**	**DFD-06 N=176**	**Vehicle N=89**	**p-value**
Primary Endpoint Treatment success at Day 15[2]	30.2%	9.0%	<0.001	30.1%	9.7%	<0.001
Secondary Endpoints % change from baseline in BSA at Day 15[3]	-28.9 (34.0)	-6.1 (32.7)	<0.001	-25.1 (36.6)	-7.2 (19.7)	<0.001
Treatment success at Day 8[2]	15.7%	5.6%	0.006	14.2%	1.6%	0.001

Source: Reviewer's analysis

(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.

(2) Treatment success defined as an IGA score of 0 (clear) or 1 (almost clear) with at least a 2-grade reduction from baseline. Missing values were handled using multiple imputation and results were averaged over the 5 imputed data sets. The p-value was calculated from a CMH test for general association adjusted for analysis center and baseline IGA score.

(3) Results presented as mean (standard deviation). Missing values were handled using last observation carried forward. The p-value was calculated from a two-way ANOVA with fixed factors for treatment, analysis center, and baseline IGA score.

5.3 Conclusions and Recommendations

Findings from the two pivotal Phase 3 studies (Trials 004 and 005) support the efficacy of clobetasol propionate cream, 0.025% (DFD-06 cream) for the topical treatment of moderate to severe plaque psoriasis in patients 18 years of age and older.

APPENDIX

A.1 Randomization Sensitivity Simulation

This section will present a simulation designed to explore what would have happened if randomization had been conducted properly after the change occurred. Randomization in the trials prior to March 3, 2016 at 5 p.m. ET appears to have been properly conducted. Therefore, these subjects' responses will not be changed in the simulation. The simulation will examine the "catch up" period where strings of subject were assigned the same treatment to meet the 2:1 treatment allocation ratio study-wide and the period after the IWRS was updated when only 2 of the 3 possible blocks of size 3 were observed. See Section 3.2.5 for a description of these periods.

The simulation investigates what would have happened if randomization had been stratified only by baseline IGA (3, 4) after March 3, 2016 at 5 p.m. ET and all possible blocks of size 3 could have been allocated for treatment assignment. To explain this further, let A represent the active treatment (DFD-06), and V represent vehicle.

After March 17, 2016 at 4:04 p.m. ET in Study 004 and March 14, 2016 at 12:24 p.m. ET in Study 005, the blocks AVA and VAA are observed, but not AAV. It is expected under block randomization that each of the three possible blocks would be allocated with probability 1/3. In the applicant's studies, the block AVA is observed approximately 2/3 of the time, and the block VAA is observed approximately 1/3 of the time. Therefore, the simulation focuses on addressing

the AVA blocks. Instead of only allowing AVA as was observed in the study, the simulation explores what the outcome would have been if the block AAV had been allocated as well.

For ease of computation, LOCF was used to impute any missing data at Day 15 prior to beginning this simulation. The observed probability of obtaining each IGA score at Day 15 given a subject's baseline IGA and treatment assignment was calculated as presented in Table 20. If some observed probabilities were 0, 0.5 was added to the count of each entry in the row, and the probabilities were recalculated. This was done so that no possible outcome would have probability 0 in the simulation.

Table 20: Probabilities(1) of IGA Scores at Day 15 by Baseline IGA Score and Treatment

	Trial 004					Trial 005				
Day 15 IGA(2)	**0**	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**
Baseline IGA 3										
Active	0.09	0.24	0.44	0.22	0.003	0.06	0.25	0.37	0.31	0.01
Vehicle	0.03	0.08	0.29	0.60	0.01	0.01	0.10	0.26	0.62	0.01
Baseline IGA 4										
Active	0.02	0.09	0.36	0.40	0.13	0.01	0.26	0.32	0.32	0.10
Vehicle	0.03	0.03	0.38	0.24	0.31	0.03	0.08	0.18	0.33	0.38

(1) If some observed probabilities were 0, 0.5 was added to the count of each entry in the row, and the probabilities were recalculated. Probabilities are rounded for presentation.
(2) Missing values were handled using last observation carried forward (LOCF).

After calculating the probabilities in Table 20, the simulation performs the following steps:

1. Identify all AVA blocks of subjects randomized after March 17, 2016 at 4:04 p.m. ET in Trial 004 and after March 14, 2016 at 12:24 p.m. ET in Trial 005. These are the only subjects whose outcomes may be altered in the simulation. For each AVA block, follow Steps 2-5.
2. The first subject is randomized correctly, so their data remains unchanged.
3. The second subject's treatment assignment, V, is unchanged with probability 0.5, and changed to A with probability 0.5. If the treatment assignment is changed, go to Step 4. If the treatment assignment is unchanged, the third subject is also unchanged and both subjects' final IGA score remains unchanged from the observed data; stop here for this block of subjects.
4. The third subject's treatment assignment is changed from A to V.
5. Impute the second and third subjects' IGA score at Day 15 based on the probabilities in Table 20 given their baseline IGA score and new treatment assignment.
6. Combine the subjects in the identified blocks with the rest of the subjects in the study and analyze the results as the SAP specified.
7. Repeat 10,000 times and average the results.

Table 21 presents the simulation results from analyzing all ITT subjects in the study and just those randomized after the randomization change on March 3, 2016. The table presents the average response rates and p-values across the 10,000 simulations along with the Monte Carlo standard deviations of these averages. The IQR of the treatment effects and the percent of nonsignificant p-values at the 0.05 level across all 10,000 simulations are also presented.

Table 21: Randomization Sensitivity Simulation Results for Treatment Success at Day 15

	Trial 004		Trial 005	
	DFD-06	**Vehicle**	**DFD-06**	**Vehicle**
ITT Population[1]	**N=178**	**N=89**	**N=176**	**N=89**
Treatment success at Day 15[2]	31.1% (1.3%)	9.5% (1.5%)	30.8% (1.0%)	10.7% (1.2%)
p-value[2, 3]	<0.001 (<0.001)		<0.001 (<0.001)	
Non-significant[4]	0%		0%	
IQR of treatment effect[5]	(20.2-23.0%)		(19.1-21.1%)	
After change	**N=97**	**N=46**	**N=57**	**N=33**
Treatment success at Day 15[2]	28.2% (2.3%)	5.5% (2.9%)	30.2% (3.1%)	3.9% (3.5%)
p-value[2, 3]	0.005 (0.01)		0.01 (0.02)	
Non-significant[4]	1.4%		3.9%	
IQR of treatment effect[5]	(20.1-25.3%)		(23.3-29.5%)	

Source: Reviewer's simulation

(1) The intent-to-treat (ITT) population was defined as all subjects who were randomized and dispensed study product.
(2) Based on 10,000 simulations where missing values were handled using last observation carried forward. Presented as mean (sd) where mean is the Monte Carlo average and sd is the Monte Carlo standard deviation of the simulation results.
(3) The p-value was calculated from a CMH test for general association adjusted for analysis center and baseline IGA score.
(4) Defined as the percentage of the 10,000 simulations that produced p-values greater than 0.05.
(5) The interquartile range of the estimated treatment effects from the 10,000 simulations.

The simulation results are consistent with the conclusion from the primary analysis that DFD-06 cream is statistically superior to vehicle cream for the primary endpoint with a treatment effect size of approximately 20%.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

REBECCA S HAGER
09/29/2017

MOHAMED A ALOSH
09/29/2017