Statistical Analysis Plan

Protocol Number/version	MPS-16IPVFT05 Amendment #1 (April 19, 2017)
Protocol Title	Statistical Analysis Plan for MPS-16IPVFT05: A RANDOMIZED, SINGLE-CENTER, CLINICAL PHASE II EVALUATION OF THE TIME-DEPENDENT ANTIMICROBIAL EFFECTIVENESS OF OCTENIDINE DIHYDROCHLORIDE IN ISOPROPYL ALCOHOL FOR PREOPERATIVE SKIN PREPARATION
SAP Date	August 2, 2017

Version History	Action
June 15, 2017	Original Release
August 2, 2017	Per Protocol Amendment #3, the order of the comparisons for non-inferiority was changed per communication with FDA dated June 6, 2017. Instead of using the wording of log reduction for the primary analysis, the wording of average treatment effect was used. For product expression volumes, inclusion of data points into Modified Intent to Treat (mITT) analysis will be based on whether the matching data point is included for efficacy for average treatment effect instead of for responder rate. Added a list of abbreviations and definitions of terms.

Approval

Signature below indicates approval of the statistical analysis plan as written.		
Function	**Name**	**Signature**
Statistician (Author)	Lingzhi Li	***This document is signed electronically in the eTMF system.***
Clinical Trial Manager	Carole Schatz	***This document is signed electronically in the eTMF system.***
Statistics Manager	Julie Bérubé	***This document is signed electronically in the eTMF system.***

Contents

1 List of Abbreviations and Definitions of Terms

AC	Active Control
AE	Adverse Event
ATE	Average Treatment Effect
BD	Becton Dickinson and Company
CI	Confidence Interval
Conf Int	Confidence Interval
IP	Investigational Product
ITT	Intent To Treat
LSMean	Least Squares Mean
mITT	Modified Intent To Treat
NC	Negative Control
SD	Standard Deviation

2 Executive Summary

For average treatment effect / responder rate, the per-protocol data may be different from the modified intent to treat (mITT) data and analysis will be made separately based on both the per-protocol data and the modified intent to treat (mITT) data if two data sets are different. Summary statements will be made for each analysis to show if there are statistically significant differences found, or if Efficacy Standards are met for different comparisons. Links to tables showing analysis results will be provided.

3 Overall Study Description

3.1 Study Background

Pilot clinical studies have been completed by BD in healthy volunteers to assess the safety and efficacy of topical application at various concentrations of octenidine dihydrochloride in isopropyl alcohol as summarized in the Investigator's Brochure (IB). No drug-related adverse events (AEs) have been observed. Additionally, a clinical study was conducted to measure the systemic absorption of octenidine after topical administration of octenidine dihydrochloride in isopropyl alcohol under simulated maximal clinical usage to intact and abraded skin in healthy volunteers. No measurable octenidine in plasma was observed (lower limit of quantification was 1 ng/mL).

3.2 Objectives

The primary objective of this study is to compare the immediate antimicrobial properties of octenidine dihydrochloride [OCT] in isopropyl alcohol - clear in a single-use applicator and octenidine dihydrochloride [OCT] in isopropyl alcohol - tinted in a single-use applicator to a Negative Control and an FDA NDA-approved Active Control. Single use applicators filled with 0.9% saline will serve as the Negative Control and ChloraPrep® - Hi-Lite Orange® applicator will serve as the Active Control. In agreement with the Food and Drug Administration Tentative Final Monograph (TFM) for Effectiveness Testing of a Patient Preoperative Skin Preparation, a $\log_{10}$ reduction study (i.e., average treatment effect) will be used to determine antimicrobial efficacy based on the statistical analyses outlined in published letters from the FDA on January 19, 2017.

The secondary objective is to determine the persistent antimicrobial activity of the two Investigative Products, Active Control and Negative Control by responder rate. Log_{10} bacterial counts at 6 hours for each treatment site will be converted to a binary (yes or no) success measure with success defined as having skin flora counts that are less than or equal to the baseline skin flora counts.

3.3 Study Design

This is a single-center and randomized design where each subject receives two different treatments out of 4 treatments, one on the left side of the body and one on the right side. This means there are 6 possible combinations of treatments (assuming A, B and C are either Investigational Product #1, Investigational Product #2 or Active Control, D is the Negative Control): A & B, A & C, A & D, B & C, B & D, and C & D.

A minimum of 72 evaluable subjects will be required to achieve a total of at least 144 test sites for groin and abdominal regions. To achieve 36 evaluable test sites from Investigational Product 1, Investigational Product 2, Active Control and Negative Control arms for groin and abdominal regions, of the minimum 72 evaluable subjects:

- 12 subjects will receive treatments from both A and B.
- 12 subjects will receive treatments from both A and C.
- 12 subjects will receive treatments from both A and D.
- 12 subjects will receive treatments from both B and C.
- 12 subjects will receive treatments from both B and D.
- 12 subjects will receive treatments from both C and D.

The two investigational products and the active control will be randomly assigned codes of A, B or C. However, due to visual differences between each of the test materials, the products cannot be blinded to all study staff. In order to guard against any bias of the study outcome, technicians who participate in product application or sample collection from subjects during treatment day procedures will not participate in the processing of samples. The technicians processing samples and/or counting the resultant plates will be blinded to the study randomization during the data gathering processes. Plate counts will be reviewed and recorded by separate BSLI staff prior to being entered into a data spreadsheet where product information (unblinding) would occur.

Subjects will be sampled for treatment day baseline microbial populations on randomly assigned sites of the abdominal and inguinal areas on both sides. Subjects will have samples collected at baseline, 30 seconds ($\pm$ 5 seconds), 10 minutes ($\pm$ 30 seconds) and at 6 hours ($\pm$ 30 minutes) post-treatment application on both abdomen and inguinal sites. All sampling times will be calculated from the completion of the dry time of each product following application.

There is a neutralization test before the main treatment study. This test is to assure that the neutralizers used in the recovery medium quench the antimicrobial activity of the three test materials (The Investigational Products and Active Control), and that the neutralizers are not toxic to the bacteria. Note: this statistical analysis plan is only for the Main Treatment study and does not include the analysis for the Neutralization study. The analysis for the Neutralization study will be conducted by the clinical site and included in the final clinical report.

The group treatments and number of treatment sites for each group for the main treatment study are shown in Table 1.

Table 1: Group treatments and number of treatment sites for the main treatment part

Product Code	Arm	Treatment	Number of evaluable treatment sites per anatomical site
A, B or C	Investigational Product #1 (IP1)	Octenidine Dihydrochloride in Isopropyl Alcohol in a single-use applicator - clear	36
A, B or C	Investigational Product #2 (IP2)	Octenidine Dihydrochloride in Isopropyl Alcohol in a single-use applicator -tinted	36
A, B or C	Active Control (AC)	ChloraPrep® - Hi-Lite Orange® applicator	36
D	Negative Control (NC)	Sterile 0.9% saline applied with single use applicator	36

3.4 Endpoints

- Primary Endpoints
  - Log CFU/cm^2 of skin for each treatment per body site per time point.
- Secondary Endpoints
  - Responder rates will be calculated by converting each $\log_{10}$ CFU/cm^2 reduction at 6 hours post-treatment time point to a binary (yes or no) success measure. Success at 6 hours post-treatment time point is defined as having skin flora counts that are less than or equal to the baseline skin flora counts on both body sites.
- Exploratory Endpoints
  - Log_{10} CFU/cm^2 reductions from baseline for each treatment per body site per post-application time point.
  - Product expression volumes: the weight of the treatment applied to a test site will be estimated by subtracting the weight measurement (g) of the treatment after the application from the weight of the treatment (g) before application.
- Safety Endpoints
  - Skin irritation scores (Erythema, Edema, Rash and Dryness scores for each time point)

4 Sample Size

A sufficient number of volunteers will be enrolled in the screening phase such that the total numbers of evaluable samples collected from groin and abdominal regions are not less than 36 evaluable test sites from each treatment arm (minimum of 144 each for groin and abdominal regions). A minimum of 72 evaluable subjects is required to achieve a total of at least 144 test sites for groin and abdominal regions. Subjects must satisfy all Screening Day and Treatment Day Inclusion/Exclusion Criteria prior to screening sample collection and Treatment Day procedures. If the required numbers of subjects do not qualify from the initial screening group, additional volunteers will be recruited.

This is a Phase II study to assess the immediate and persistent antimicrobial effect of the Investigational Products relative to the Negative Control and the Active Control using a $\log_{10}$ reduction study (i.e., average treatment effect) according to the procedures outlined in the Food and Drug Administration Tentative Final Monograph (TFM) for Effectiveness Testing of a Patient Preoperative Skin Preparation (FR 59:116, 17 June 94, pp. 31450-31452) and the test criteria outlined in the deferal letters from the FDA published on January 19, 2017. The purpose is to inform design of future efficacy studies, therefore no sample size calculation was performed to provide statistically-powered evidence of efficacy as part of this study. The sample size is deemed adequate for these purposes.

5 Intended Statistical Software

R 3.2.0 (2015-04-16) or later version if updated.

6 Data

6.1 Data Sets Analyzed

Descriptions of the modified intent to treat (mITT) data, per-protocol data and intent to treat (ITT) data are shown below. Inclusion for the per-protocol data / mITT data set is evaluated for each body area (left and right for the groin and abdomen).

- Modified Intent to Treat (mITT) data: data collected will be included if the treatment-day baseline counts are within the acceptable range ($\geq$ 3.00 - $\leq$ 5.50 $\log_{10}$ CFU/cm^2 on abdominal area, and $\geq$ 5.00 - $\leq$ 7.50 $\log_{10}$ CFU/cm^2 on inguinal area). Data collected will be excluded from mITT data if the treatment-day baseline counts are outside the acceptable range.
  - For responder rate at 6 hours post-treatment time point, missing data at either baseline or a certain time interval will be imputed as non-responders for that time interval in the mITT data.
  - For average treatment effect (ATE) and $\log_{10}$ reduction, missing data will not be imputed for mITT data.
- Per-Protocol data: data collected will be included if the treatment-day baseline counts are within the acceptable range ($\geq$ 3.00 - $\leq$ 5.50 $\log_{10}$ CFU/cm^2 on abdominal area, and $\geq$ 5.00 - $\leq$ 7.50 $\log_{10}$ CFU/cm^2 on inguinal area). Data collected will be excluded from per-protocol data if the treatment-day baseline counts are outside the acceptable range.
  - Missing data at any time interval will not be imputed for ATE, responder rate and log reduction. Missing data will be excluded from per-protocol ATE, responder rate and log reduction analysis.
- Intent to Treat (ITT) data: all randomized subjects with available data will be included and used for the ITT analysis. The full intent to treat (ITT) data set (all randomized subjects) will be used for the safety analysis.

For ATE, responder rate and log reduction, efficacy analyses will be first made on the modified intent to treat (mITT) data set. Efficacy analyses will be also conducted on the per-protocol data set as supportive analyses when per-protocol data are different from mITT data.

For product expression volumes, inclusion of data points will be based on whether the matching data point is included for efficacy for ATE. Analysis will be performed based on the per-protocol data as supportive analyses when per-protocol data are different from mITT data for product expression volumes.

For skin irritation scores, analysis will be performed based on the intent to treat (ITT) data (all randomized subjects).

6.2 Analysis Population Set(s)

For each treatment per body site, the number of subjects randomized, the number of subjects available for mITT analysis [average treatment effect (ATE), responder rate and product expression volumes] and / or for per-protocol analysis [average treatment effect (ATE), responder rate and product expression volumes] at each post-treatment time point (30 seconds, 10 minutes and 6 hours) will be provided in tables below (cf. Table 2, Table 3 and Table 4). If the number of subjects randomized is different from the number of subjects randomized with treatments received, the number of subjects randomized with treatment received will be also summarized.

Table 2: Number of subjects randomized

Treatment	Body Site	N Subjects Randomized with Treatment Received
A	Abdomen	
B	Abdomen	
C	Abdomen	
D	Abdomen	
A	Groin	
B	Groin	
C	Groin	
D	Groin	

Table 3: Number of subjects available for mITT analysis

Treatment	Body Site	N Subjects for mITT ATE at 30 sec	N Subjects for mITT ATE at 10 min	N Subjects for mITT ATE at 6 hr	N Subjects for mITT Responder Rate at 6 hr	N Subjects for mITT Product Expression Volumes
A	Abdomen					
B	Abdomen					
C	Abdomen					
D	Abdomen					
A	Groin					
B	Groin					
C	Groin					
D	Groin					

Table 4: Number of subjects available for per-protocol analysis

Treatment	Body Site	N Subjects for Per-Protocol ATE at 30 sec	N Subjects for Per-Protocol ATE at 10 min	N Subjects for Per-Protocol ATE at 6 hr	N Subjects for Per-Protocol Responder Rate at 6 hr	N Subjects for Per-Protocol Product Expression Volumes
A	Abdomen					
B	Abdomen					
C	Abdomen					
D	Abdomen					
A	Groin					
B	Groin					
C	Groin					
D	Groin					

7 Statistical Analysis/Calculations

7.1 Derived Variables

7.1.1 Log CFU/cm^2 of skin

The estimated $\log_{10}$ number of viable microorganisms per cm^2 recovered from each sample site will be designated the "R-value". To convert the volumetric measure of the sample into the number of colony-forming units per square centimeter (cm^2), the following formula will be employed:

$$R = log_{10}\left[\frac{F\left(\frac{\sum_{i=1}^{2} C_{\mathrm{i}}}{n}\right) 10^{-\mathrm{D}}}{A}\right] \quad (1)$$

Where:
R = the average colony-forming unit count in $\log_{10}$ scale per cm^2 of sampling surface
F = Total number of mL of sampling solution added to the sampling cylinder (in this study, F = 6 mL for all samples)
$\frac{\sum_{i=1}^{2} C_{\mathrm{i}}}{n}$ = average of the duplicate colony counts used for each sample collected (n = 2)
D = Dilution factor of the plates counted
A = Inside area of the sampling cylinder (A= 3.46 cm^2 in this study)

Note: The reason that a $\log_{10}$ transformation will be performed on the collected data is to obtain a linear scale. A linear scale, more appropriately a $\log_{10}$ linear scale, is a basic requirement of the statistical models used in this study.

If colonies on one of the plates are uncountable, the count from the remaining plate will be used.

In order to avoid potential calculation problems due to taking the logarithm of zero, counts of less than 1 CFU/cm^2 are treated as 1 CFU/cm^2, such that the $\log_{10}$ transformation is no less than zero.

7.1.2 Log Reduction

Log_{10} CFU/cm^2 reductions from baseline will be calculated separately for each subject, each post-application time point, and each of the treated anatomical sites by taking the baseline $\log_{10}$ CFU/cm^2 values and then subtracting the $\log_{10}$ CFU/cm^2 values for the samples taken after the baseline.

7.1.3 Responder Rate

Responder statuses for each subject will be evaluated separately for each of the four sites and separately at 6 hours post-application time point based on the $\log_{10}$ CFU/cm^2 changes from baseline. A subject is considered a responder at the 6-hour sample time for the abdomen or groin if the $\log_{10}$ CFU/cm^2 reduction is equal to or greater than 0.

7.1.4 Product Expression Volumes

The weight of drug product applied to a test site will be estimated by subtracting the weight measurement of the test material after the application from the weight of the test material before application.

7.2 Handling of Missing Data

For mITT responder rate analysis, subjects with missing data at either baseline or a certain time interval will be included as non-responders for that time interval. Responder rate analysis will be performed based on both

per-protocol data and mITT data if there are any missing data identified and if per-protocol data and mITT data are different.

For average treatment effect (ATE), log reduction and product expression volumes, missing data will not be imputed for mITT data. Analysis will be performed based on the per-protocol data as supportive analyses when per-protocol data are different from mITT data.

7.3 Analysis Methods

For each body site, when per-protocol data are different from mITT data for average treatment effect (ATE), responder rate, log reduction or product expression volumes, per-protocol analysis and mITT analysis will be performed separately using the same statistical methods. The following analysis methods will be applied to both per-protocol analysis and mITT analysis, both abdomen and groin.

An alpha level of 5% (two-sided) is used for all analyses.

7.3.1 Primary Analysis

The primary purpose of the study is to compare the immediate antimicrobial activity of the Investigational Products to the Negative Control and Active Control according to the methods described in the FDA TFM with the statistical analyses outlined in a deferral letter published January 19, 2017. There are two test criteria that drive the hypotheses.

- Evaluate the average treatment effect (ATE) of the Active Control (AC) compared to each Investigational Product (IP) at 30 seconds and 10 minutes for non-inferiority on both abdomen and groin. ATE of the AC compared to either IP is defined as the contrast of the treatment effect of the IP minus the treatment effect of the AC (IP - AC) in the linear regression. Statistical hypotheses are:
  - Ho: The upper bound 95% confidence interval of the ATE of IP - AC ≥ 0.5 $\log_{10}$
  - Ha: The upper bound 95% confidence interval of the ATE of IP - AC < 0.5 $\log_{10}$
- Evaluate the Average Treatment effect (ATE) of each Investigational Product (IP) compared to the Negative Control (NC) at 30 seconds and 10 minutes on both abdomen and inguen anatomical area. ATE of the IP compared to the NC is defined as the contrast of the treatment effect of the NC minus the treatment effect of the IP (NC - IP) in the linear regression. Statistical hypotheses are:
  - Ho: The lower bound 95% confidence interval of the ATE of NC - IP ≤ 1.2 $\log_{10}$
  - Ha: The lower bound 95% confidence interval of the ATE of NC - IP > 1.2 $\log_{10}$

Ho is the null hypothesis and Ha is the alternative hypothesis.

A linear regression model for each body site will be used for primary analysis. In the model, the response is the post-treatment $\log_{10}$ bacterial counts and predictors are the treatment effect as a fixed effect and the pre-treatment $\log_{10}$ bacterial counts as a covariate. The interaction between the treatment and pre-treatment $\log_{10}$ bacterial counts will be also explored. If there is no significant interaction detected between treatment and pre-treatment $\log_{10}$ bacterial counts, the interaction term will be removed from the model and the average difference between treatments will be estimated from the linear regression model after correcting for pre-treatment $\log_{10}$ bacterial counts. If there is a significant interaction detected between treatment and pre-treatment $\log_{10}$ bacterial counts, the average difference between treatments will be estimated from the linear regression model at the average of the covariate. Additional factors (e.g., technician and subject demographics) may be also included in the model and the effects will be investigated.

Since the study will declare success when and only when both non-inferiority and superiority are demonstrated for both abdomen and groin, no multiplicity adjustment for multiple primary objectives is required.

An example for results of comparisons above can be found in Tables 5 and 6. All results that do not meet the targets are highlighted.

Table 5: Example: ATE for IP1/IP2 - AC for Groin

Comparison	Time Point	N Pairs	Difference in ATE	Conf Int	Meet Non-inferiority Criteria
IP1 - AC	30 Seconds	36	-0.50	-1.31, 0.30	Yes
IP2 - AC	30 Seconds	36	-0.35	-1.00, 0.31	Yes
IP1 - AC	10 Minutes	36	0.32	-0.52, 1.15	No
IP2 - AC	10 Minutes	36	0.50	-0.39, 1.39	No

Table 6: Example: ATE for NC - IP1/IP2 for Groin

Comparison	Time Point	N Pairs	Difference in ATE	Conf Int	Meet Superiority Criteria
NC - IP1	30 Seconds	36	1.35	1.01, 1.59	No
NC - IP2	30 Seconds	36	1.23	0.90, 1.50	No
NC - IP1	10 Minutes	36	0.4	0.02, 0.82	No
NC - IP2	10 Minutes	36	1.6	1.28, 1.91	Yes

7.3.2 Secondary Analysis

The secondary purpose of the study is to determine the persistent antimicrobial activity of the two Investigational Products, Active Control and Negative Control by calculating the responder rate along with 95% lower confidence bound at 6-hour post-treatment time point for each body site using an Exact method (one-sided). The goal is to demonstrate a success proportion $\geq$ 95% on both the abdomen and the inguen with 95% confidence. However, since this is a Phase II study to inform design of future efficacy studies and is not designed to provide statistically powered evidence of efficacy, the lower confidence bounds cannot lie entirely above 95%.

An example of comparing responder rates to 95% for each treatment is provided below.

Table 7: Example: Analysis results for responder rate for Groin

Treatment	Time Point	N Non-responder	N Responder	N Total	Responder Rate	95% Lower Confidence Bound
A	6 Hours	1	35	36	97.2%	87.5%
B	6 Hours	0	36	36	100.0%	92.0%
C	6 Hours	2	34	36	94.4%	83.5%
D	6 Hours	1	35	36	97.2%	87.5%

7.3.3 Exploratory Analyses

7.3.3.1 Summary Statistics

Summary statistics will be provided for $\log_{10}$ CFU/cm^2 of skin, $\log_{10}$ CFU/cm^2 reduction and product expression volumes based on mITT data and/or per-protocol data. For continuous variables, data will be summarized with the following descriptive statistics: number of observations, mean (LSMean), median, standard deviation and range (minimum - maximum). See examples for summary statistics Tables 8 and 9 below for $\log_{10}$ CFU/cm^2 and $\log_{10}$ CFU/cm^2 reduction.

Table 8: Example: summary statistics for $\log_{10}$ CFU/cm^2 for Groin

$\log_{10}$ CFU/cm^2	Treatment	Time Point	N	Mean	LSMean	Median	SD	Range
$\log_{10}$ CFU/cm^2	A	Baseline	36	5.92	5.92	5.82	0.53	5.01 - 7.23
$\log_{10}$ CFU/cm^2	A	30 Seconds	36	3.29	3.29	3.44	1.10	0.89 - 5.89
$\log_{10}$ CFU/cm^2	A	10 Minutes	36	3.05	3.05	3.09	0.72	1.22 - 4.09
$\log_{10}$ CFU/cm^2	A	6 Hours	36	5.78	5.78	5.66	0.54	5.05 - 7.04
$\log_{10}$ CFU/cm^2	B	Baseline	36	3.27	3.27	3.43	1.09	0.00 - 4.95
$\log_{10}$ CFU/cm^2	B	30 Seconds	36	3.26	3.26	3.39	0.89	0.00 - 4.47
$\log_{10}$ CFU/cm^2	B	10 Minutes	36	5.92	5.92	5.82	0.53	5.01 - 7.23
$\log_{10}$ CFU/cm^2	B	6 Hours	36	3.27	3.27	3.43	1.09	0.00 - 4.95
$\log_{10}$ CFU/cm^2	C	Baseline	36	3.26	3.26	3.39	0.89	0.00 - 4.47
$\log_{10}$ CFU/cm^2	C	30 Seconds	36	3.05	3.05	3.09	0.72	1.22 - 4.09
$\log_{10}$ CFU/cm^2	C	10 Minutes	36	3.29	3.29	3.44	1.10	0.89 - 5.89
$\log_{10}$ CFU/cm^2	C	6 Hours	36	3.05	3.05	3.09	0.72	1.22 - 4.09
$\log_{10}$ CFU/cm^2	D	Baseline	36	5.78	5.78	5.66	0.54	5.05 - 7.04
$\log_{10}$ CFU/cm^2	D	30 Seconds	36	3.27	3.27	3.43	1.09	0.00 - 4.95
$\log_{10}$ CFU/cm^2	D	10 Minutes	36	5.92	5.92	5.82	0.53	5.01 - 7.23
$\log_{10}$ CFU/cm^2	D	6 Hours	36	3.29	3.29	3.44	1.10	0.89 - 5.89

Table 9: Example: summary statistics for $\log_{10}$ CFU/cm^2 reduction for Groin

$\log_{10}$ CFU/cm^2 reduction	Treatment	N	Mean	LSMean	Median	SD	Range
$\log_{10}$ CFU/cm^2 reduction	A Baseline - 30 Seconds	36	2.63	2.63	2.57	1.18	0.20 - 4.80
$\log_{10}$ CFU/cm^2 reduction	A Baseline - 10 Minutes	36	2.83	2.83	2.83	0.89	1.24 - 4.86
$\log_{10}$ CFU/cm^2 reduction	A Baseline - 6 Hours	36	2.52	2.52	2.33	1.09	0.83 - 5.50
$\log_{10}$ CFU/cm^2 reduction	B Baseline - 30 Seconds	36	2.53	2.53	2.30	1.11	0.90 - 5.50
$\log_{10}$ CFU/cm^2 reduction	B Baseline - 10 Minutes	36	2.63	2.63	2.57	1.18	0.20 - 4.80
$\log_{10}$ CFU/cm^2 reduction	B Baseline - 6 Hours	36	2.53	2.53	2.30	1.11	0.90 - 5.50
$\log_{10}$ CFU/cm^2 reduction	C Baseline - 30 Seconds	36	2.83	2.83	2.83	0.89	1.24 - 4.86
$\log_{10}$ CFU/cm^2 reduction	C Baseline - 10 Minutes	36	2.52	2.52	2.33	1.09	0.83 - 5.50
$\log_{10}$ CFU/cm^2 reduction	C Baseline - 6 Hours	36	2.63	2.63	2.57	1.18	0.20 - 4.80
$\log_{10}$ CFU/cm^2 reduction	D Baseline - 30 Seconds	36	2.52	2.52	2.33	1.09	0.83 - 5.50
$\log_{10}$ CFU/cm^2 reduction	D Baseline - 10 Minutes	36	2.83	2.83	2.83	0.89	1.24 - 4.86
$\log_{10}$ CFU/cm^2 reduction	D Baseline - 6 Hours	36	2.53	2.53	2.30	1.11	0.90 - 5.50

7.3.3.2 Log Reduction

Log_{10} reductions will be compared between test materials by an Analysis of Variance (ANOVA) for each body site based on mITT data and/or per-protocol data. In the ANOVA model, the response is the $\log_{10}$ reduction and the predictors are treatment and time point as well as the interaction between treatment and time point. The average difference in log reduction between treatments per time point will be estimated. However, all conclusions for the difference in the average treatment effect between treatments will be based on the primary analysis only.

An example for results of comparisons above can be found in Tables 10.

Table 10: Example: results for product comparisons within time interval (10 minutes) for Groin

Comparison	N Pairs	Difference	Conf Int	Conclusion
IP1 Baseline-A 10 min vs. AC Baseline-AC 10 min	36	0.13	-0.44, 0.70	no sign. diff.
IP2 Baseline-A 10 min vs. AC Baseline-AC 10 min	36	0.11	-0.65, 0.87	no sign. diff.
IP1 Baseline-A 10 min vs. VC Baseline-VC 10 min	36	0.33	-0.44, 1.09	no sign. diff.
IP2 Baseline-A 10 min vs. VC Baseline-AC 10 min	36	0.33	-0.44, 1.09	no sign. diff.

7.3.3.3 Product Expression Volumes/Weight

An analysis of variance (ANOVA) of the applied volumes/weight will be performed separately for each body site based on mITT data and/or per-protocol data. In the model, there will be a random subject effect, a fixed body side (left or right) effect, a fixed treatment effect and the interaction between fixed effects. ANOVA table will be checked to see whether the treatment has a significant effect on product expression volumes as well as the interaction. If a significant difference is found, subgroup comparisons with Tukey's method will be made to determine which pair of comparisons shows significant difference.

7.3.4 Safety Analysis

The ITT data set (all randomized subjects) will be considered evaluable for safety. Skin irritation scores will be reported for any subject who is scored with a 1 or more at any observation [baseline, post-application/prior to 30 seconds, 10 minutes, and 6 hours sampling procedures], in any category for any site.

Adverse Events (including post treatment skin irritation scores of 3), will also be summarized. Summary tables will present incidence rates of adverse events by treatment group for all subjects who enter the treatment period. Listings of Adverse Events will be provided.

The statistical significance of differences in skin irritation between the Investigational Products, Active Control and Negative Control will be evaluated using Fisher's exact test on skin irritation data summarized as follows: any reaction above zero (no reaction) on the skin irritation rating scale for any category (erythema, edema, rash, and dryness) will be considered a positive signal for that substance. If Fisher's exact test shows there is an overall significant treatment effect, Fisher's exact test or two-proportion test with Score method with multiplicity adjustment on the alpha level may be used for multiple sub-group comparisons to find out which comparisons show the difference. Odds ratio (odds of having positive signals for one treatment / odds of having positive signals for the other treatment) using Fisher's exact test or the difference in percentage of positive signals between treatments using Two-proportion test may be calculated along with the confidence interval for each category at the certain time interval. The analysis for skin irritation scores will be for informational purposes only.

Examples for Fisher's exact test and two-proportion test with Score method are shown below.

```
> # overall treatment effect
> irr_tab <- matrix(c(2, 1, 3, 1, 34, 35, 33, 35),
+        nrow = 4,
+        dimnames =
+        list(c("A", "B", "C", "D"),
+             c("Positive (Score > 0)", "Non-positive (Score = 0)")))
> irr_tab
```

```
  Positive (Score > 0) Non-positive (Score = 0)
A                    2                       34
B                    1                       35
C                    3                       33
D                    1                       35
```

```
> irr <- fisher.test(irr_tab, alternative = "two.sided",
+                 conf.level = 0.95)
> print(irr)
```

```
        Fisher's Exact Test for Count Data

data:  irr_tab
p-value = 0.8361
alternative hypothesis: two.sided
```

```
> # subgroup comparisons
> library(PropCIs)
> print(diffscoreci(2, 34, 1, 35, conf.level = 0.972))
```

```
data:

97.2 percent confidence interval:
 -0.1176006  0.1904725
```

7.4 Graphs

7.4.1 ATE Per Primary Analysis

The graph for the ATE of IP1/IP2 - AC with 95% intervals estimated from the regression model (as in Tables 5 and 6) will be provided (see examples in Figure 1 and 2).

Groin

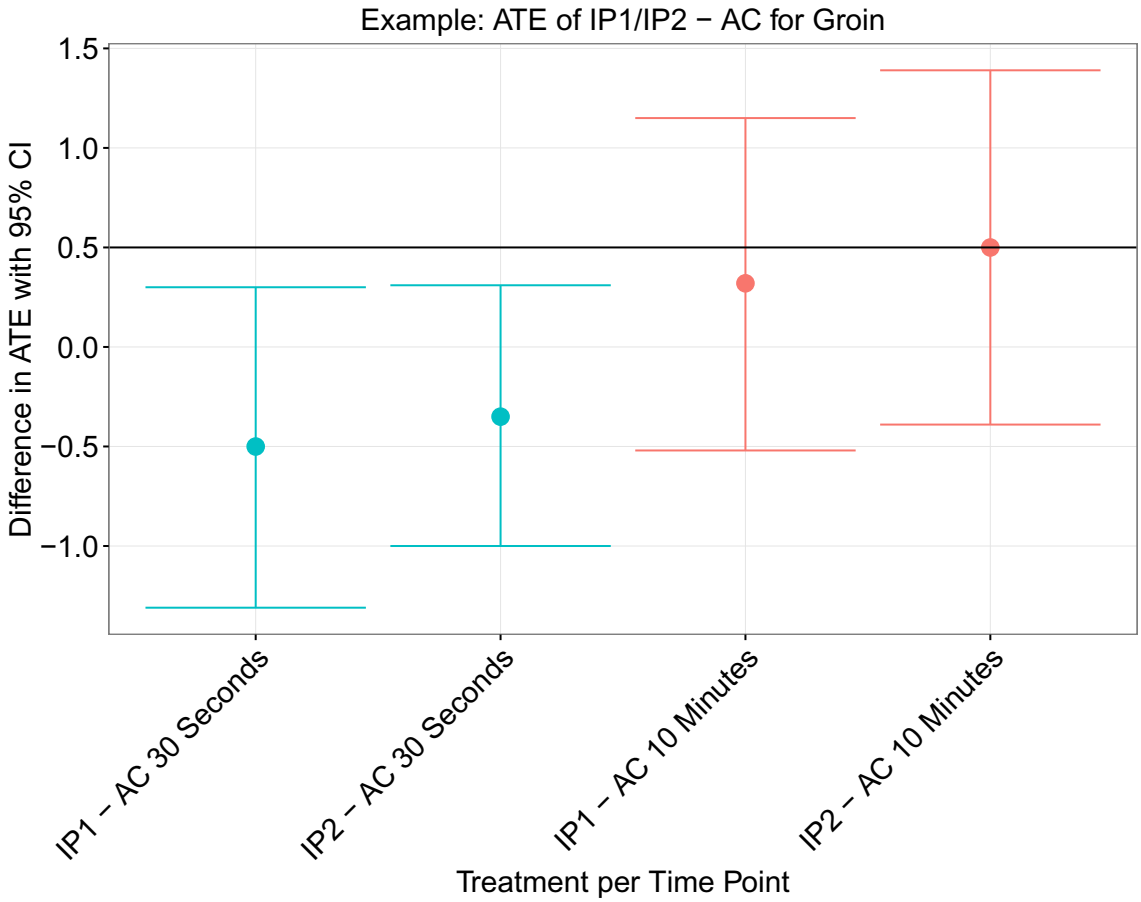

Figure 1: ATE of IP1/IP2 - AC for Groin

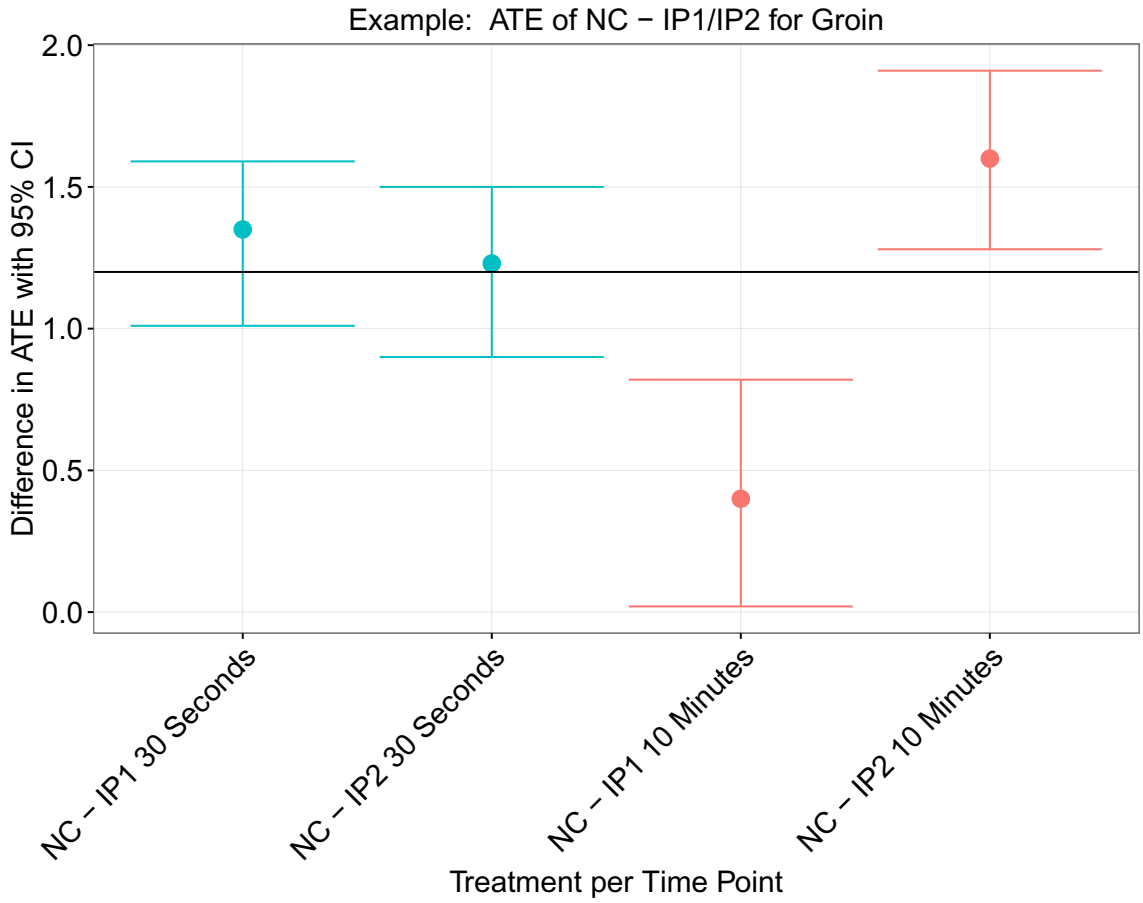

Figure 2: ATE of NC - IP1/IP2 for Groin

8 Appendix

Subjects not passing the treatment-day baseline criteria will be listed.

Subjects included in mITT analysis, but excluded from per-protocol analysis, if any, will be listed.

Signature Page for MPS-16IPVFT05_Statistical_Analysis_Plan_V2

Approved	Lingzhi Li Statistician, Corporate/Shared Services Statistician 14-Aug-2017 20:13:12 GMT+0000

Approved	Carole Schatz Study Manager 14-Aug-2017 20:45:59 GMT+0000

Approved	Julie Berube Manager Statistics, Corporate/Shared Services Functional Management 15-Aug-2017 12:28:16 GMT+0000

Signature Page for MPS-16IPVFT05_Statistical_Analysis_Plan_V2 - System Version N