Statistical Analysis Plan

Protocol Number/version	MPS-17IPVAW10
Protocol Title	Statistical Analysis Plan for MPS-17IPVAW10: A Randomized, Single-Center, Blinded, Clinical Evaluation of the Antimicrobial Effectiveness of an Antimicrobial Cloth Compared to Vehicle and 0.9% Saline
SAP Date	June 23, 2017

Version History	Action
June 23, 2017	Original Release

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

AE	Adverse Events
ASTM	American Society for Testing And Materials
BD	Becton Dickinson and Company
CFU	Colony Forming Units
CI	Confidence Interval
cm^2	Square Centimeter
Conf Int	Confidence Interval
FDA	Food and Drug Administration
IP	Investigational Product
ITT	Intent to Treat
LSMean	Least Squares Mean
mITT	Modified Intent to Treat
mL	Milliliter
OCT	Octenidine Dihydrochloride
SC	Saline Control
SD	Standard Deviation
TFM	Tentative Final Monograph
VC	Vehicle Control

2 Executive Summary

For the average treatment effect correcting for pre-treatment $\log_{10}$ bacterial counts, 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 whether superiority of the Investigational Product as compared to the Vehicle Control and the Saline Control is satisfied. Links to tables showing analysis results will be provided.

3 Overall Study Description

3.1 Study Background

A Phase 2 study MPS-16IPVAW01 evaluated the antimicrobial efficacy of thermally treated polyester cloths impregnated with 0.4% w/v octenidine dihydrochloride in aqueous formulation as compared to the thermally treated polyester cloth impregnated with vehicle formulation and the FDA-approved active comparator SAGE 2% CHG cloth. Activity was assessed relative to a 70% response rate target based on the log reduction standards in the 1994 FDA TFM. The current study will evaluate the immediate antimicrobial efficacy of thermally treated polyester cloths impregnated with 0.4% w/v octenidine dihydrochloride in aqueous formulation as compared to two negative controls, the thermally treated polyester cloth impregnated with vehicle formulation and a polyester cloth containing 0.9% normal saline, according to the revised statistical analyses and expected performance criteria published by the FDA on January 19, 2017.

3.2 Objectives

3.2.1 Primary Objectives

The primary objective of this study is to compare the immediate antimicrobial activities of a thermally treated polyester cloth impregnated with 0.4% w/v octenidine dihydrochloride in aqueous solution to thermally treated polyester cloths impregnated with vehicle formulation and a polyester cloth containing 0.9% saline. In agreement with the Food and Drug Administration Tentative Final Monograph (TFM) for *Effectiveness Testing of a Patient Preoperative Skin Preparation*, a log reduction study will be used to determine antimicrobial efficacy based on the statistical analyses outlined in published letters from the FDA on January 19, 2017. Testing will be performed 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). The test methods for this evaluation will be based on ASTM Standard Test Method E1173-15, *Evaluation of Preoperative, Precatheterization, or Preinjection Skin Preparations* and the test criteria from the deferal letters published by the FDA on January 19, 2017.

At 10 minutes post-prep, the mean treatment effect of the test article vs. negative controls should be $>$ 1.2 $\log_{10}$ on both the abdomen and the groin with 95% confidence. In this protocol two negative controls will be used, the Vehicle Control and the Saline Control.

3.2.2 Secondary Objectives

Safety will be evaluated using skin irritation scores and the incidence of adverse events.

3.3 Study Design

This single site study is a randomized, blinded design employing a minimum of 36 healthy volunteers, where each subject receives two of the planned treatments on the abdomen and groin.

Treatment materials for the following treatment groups will be blinded with treatment codes A or B:

- Thermally treated cloth impregnated with 0.4% w/v octenidine dihydrochloride aqueous solution (IP: Investigational Product).

- Thermally treated cloth impregnated with vehicle formulation (VC: Vehicle Control).

The 0.9% Saline Control (SC) cannot be blinded due to product application requirements and will be assigned to treatment code C.

Out of the minimum 36 evaluable subjects, 12 subjects will receive treatments from both A and B, 12 subjects will receive treatments from both A and C, and 12 subjects will receive treatments from both B and C to achieve 24 test sites from A, B and C for groin and abdominal regions. The application will be randomized so that each treatment is applied on an equal number of left and right sides of the body. Due to baseline requirements the final data set may be imbalanced with regard to right and left sides upon achieving the minimum number of qualifying body sites.

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 and 10 minutes (± 30 seconds) 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. 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 planned for each group for the main treatment study are shown in Table 1.

Table 1: Group treatments and number of treatment sites planned

Arm	Treatment	Number of evaluable treatment sites per anatomical site
Investigational Product (IP)	Thermally treated polyester cloth impregnated with 0.4% w/v octenidine dihydrochloride	24
Vehicle Control (VC)	Thermally treated polyester cloth impregnated with vehicle formulation	24
Saline Control (SC)	0.9% normal saline applied with polyester cloth	24

3.4 Endpoints

- Primary Endpoints: The primary endpoints are the microbial counts for each product at baseline and 10 minutes post-application time point for both the abdomen and groin. Log_{10} CFU/cm^2 calculated based on the microbial counts will be used for primary analysis.
- Secondary Endpoints: The secondary endpoints are safety data including evaluations for skin reactions using the modified Berger Bowman irritation assessment scale and the incidence of adverse events reported during the study for all randomized subjects (the full Intent-to-Treat data set). Post treatment skin irritation scores of 3 on those scales are also considered adverse events.
- Exploratory Endpoints: The exploratory endpoints are $\log_{10}$ CFU/cm^2 reductions from baseline for each treatment per body site per post-application time point.

- Informational Endpoints: The informational endpoints are 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.

3.5 Acceptance Criteria

The following table summarizes the minimum baseline criteria for each of the test sites and the expected minimum efficacy standards (superiority criteria).

Table 2: Minimum Treatment Baseline and Expected Mean Log_{10} Reduction per Anatomical Site

Anatomical Site	Treatmen Day Baseline Criteria	Expected Efficacy Standards
Abdomen	3.00 to 5.50 $\log_{10}$ CFU/cm^2	At 10 minutes, the lower two-sided 95% confidence bound of the average treatment effect of IP vs. VC and IP vs. SC should be greater than 1.2 $\log_{10}$
Groin	5.00 to 7.50 $\log_{10}$ CFU/cm^2	At 10 minutes, the lower two-sided 95% confidence bound of the average treatment effect of IP vs. VC and IP vs. SC should be greater than 1.2 $\log_{10}$

4 Sample Size

This is a phase 2 study to assess the immediate antimicrobial effect of the Investigational Products relative to the two negative controls using a log reduction study 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.

- 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.

- 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 secondary analysis.

For $\log_{10}$ CFU/cm^2 determinations, missing data will not be imputed for either mITT data or Per Protocol data and will be excluded from analysis. For data points excluded from per-protocol analysis due to protocol deviations, those data points may be included in mITT analysis.

For primary analysis (average treatment effect correcting for pre-treatment $\log_{10}$ bacterial counts) and exploratory analysis (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 informational analysis (product expression volumes), inclusion of data points will be based on whether the matching data point is included for efficacy for primary analysis. 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.

6.2 Analysis Population Set(s)

For each treatment per body site, the number of subjects randomized with treatments received, the number of subjects available for mITT analysis and / or for per-protocol analysis at 10 minutes post-treatment time point will be provided in tables below (cf. Table 3, Table 4 and Table 5).

Table 3: Number of Subjects Randomized with Treatment Received

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

Table 4: Number of subjects available for mITT analysis

Treatment	Body Site	N Subjects for mITT Treatment Effect	N Subjects for mITT $\log_{10}$ CFU Reduction	N Subjects for mITT Product Expression Volumes
A	Abdomen			
B	Abdomen			
C	Abdomen			
A	Groin			
B	Groin			
C	Groin			

Table 5: Number of subjects available for Per-protocol analysis

Treatment	Body Site	N Subjects for Per-protocol Treatment Effect	N Subjects for Per-protocol $\log_{10}$ CFU Reduction	N Subjects for Per-protocol Product Expression Volumes
A	Abdomen			
B	Abdomen			
C	Abdomen			
A	Groin			
B	Groin			
C	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:

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

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 (SS) added to the sampling cylinder (in this study, F = 6 mL for all samples)
$\frac{\sum_{i=1}^{2} C_i}{n}$ = 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 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 $\log_{10}$ CFU/cm^2 determinations, missing data will not be imputed for either mITT data or Per Protocol data and will be excluded from analysis.

7.3 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 6 and 7 below for $\log_{10}$ CFU/cm^2 and $\log_{10}$ CFU/cm^2 reduction.

Table 6: 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	24	5.92	5.92	5.82	0.53	5.01 - 7.23
$\log_{10}$ CFU/cm^2	A	10 Minutes	24	3.29	3.29	3.44	1.10	0.89 - 5.89
$\log_{10}$ CFU/cm^2	B	Baseline	24	5.78	5.78	5.66	0.54	5.05 - 7.04
$\log_{10}$ CFU/cm^2	B	10 Minutes	24	3.27	3.27	3.43	1.09	0.00 - 4.95
$\log_{10}$ CFU/cm^2	C	Baseline	24	5.92	5.92	5.82	0.53	5.01 - 7.23
$\log_{10}$ CFU/cm^2	C	10 Minutes	24	3.05	3.05	3.09	0.72	1.22 - 4.09

Table 7: 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 - 10 Minutes	24	2.63	2.63	2.57	1.18	0.20 - 4.80
$\log_{10}$ CFU/cm^2 reduction	B Baseline - 10 Minutes	24	2.83	2.83	2.83	0.89	1.24 - 4.86
$\log_{10}$ CFU/cm^2 reduction	C Baseline - 10 Minutes	24	2.52	2.52	2.33	1.09	0.83 - 5.50

7.4 Analysis Methods

For each body site, when per-protocol data are different from mITT data for $\log_{10}$ CFU/cm^2, 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.

7.4.1 Primary Analysis

An analysis of variance (ANOVA) of the baseline $\log_{10}$ CFU/cm^2 values will be performed separately for abdomen and groin. In each model, there will be a random subject effect, a fixed treatment effect, a fixed body side effect and the interaction between treatment and body side. Pairwise comparisons with Tukey's method will be made for $\log_{10}$ CFU/cm^2 to determine if there is any significant difference between treatments at baseline after adjusting body side (e.g., randomization has produced treatment arms with similar baseline CFU values) for each body site.

The primary purpose of the study is to compare the immediate antimicrobial activity of the Investigational Products to the Negative Control according to the methods described in the FDA TFM with the statistical analyses outlined in a deferral letter published January 19, 2017 for superiority.

Evaluate the treatment effect at 10 minutes for Investigational Product (IP) vs. Vehicle Control (VC) / Saline Control (SC) for abdomen and groin by comparing to a superiority margin of 1.2 $\log_{10}$.

- Ho: treatment effect of IP vs. VC/SC correcting for pre-treatment log $_{10}$ bacterial counts $\leq$ 1.2 $\log_{10}$
- Ha: treatment effect of IP vs. VC/SC correcting for pre-treatment log $_{10}$ bacterial counts $>$ 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 analyses. 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 two-way interaction between the treatment and pre-treatment $\log_{10}$ bacterial counts will also be explored. If there is no significant interaction detected, the interaction term will be removed from the model and the average difference between treatments correcting for pre-treatment bacterial loads will be estimated from the linear regression model. 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.

The average treatment effect along with 95% confidence interval between IP and VC/SC will be estimated from the model above for each body site and compared to a superiority margin of 1.2. If the two-sided 95% lower confidence bound $>$ 1.2 $\log_{10}$, the null hypothesis for superiority is rejected and it is concluded that IP is superior to VC/SC by passing the superiority margin of 1.2 $\log_{10}$. Otherwise the superiority criteria cannot be passed.

The primary analysis will be based on the modified Intent to Treat (mITT) data set and the supportive analysis may be conducted using the per-protocol data set, as described in Section 6.

Since the study will declare success when and only when superiority is demonstrated for both IP vs. VC and IP vs. SC for both abdomen and groin, no multiplicity adjustment for multiple primary objectives is required[1].

An example for results of comparisons above can be found in Table 8. All results that do not meet the targets are highlighted.

Table 8: Example: results for for IP vs. VC and SC for Groin

Comparison	N Pairs	Average Treatment Effect	Conf Int	Meet Superiority Criteria
IP vs. VC	24	1.23	0.90, 1.50	No
IP vs. SC	24	1.65	1.31, 1.99	Yes

7.4.2 Secondary 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 treatment day or the 10 minute post-application time point) in any category for any site.

Based on the protocol, subjects who receive an irritation score of 1 or greater (any redness, swelling, rash, or dryness present at any treatment area) for any individual skin condition prior to the treatment Day baseline sample collection will be excluded from the study and will not be randomized.

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 Product, Vehicle Control and Saline 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 a two-proportion test may be calculated along with the confidence interval for each category at the certain time interval.

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, 22, 23, 21),
+        nrow = 3,
+        dimnames =
+        list(c("A", "B", "C"),
+             c("Positive (Score > 0)", "Non-positive (Score = 0)")))
> irr_tab
```

```
  Positive (Score > 0) Non-positive (Score = 0)
A                    2                       22
B                    1                       23
C                    3                       21
```

```
> 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 = 0.8654
alternative hypothesis: two.sided
```

```
> # subgroup comparisons
> library(PropCIs)
> print(diffscoreci(2, 22, 1, 23, conf.level = 0.984))
```

```
data:

98.4 percent confidence interval:
 -0.1934747  0.3014313
```

7.4.3 Exploratory Analyses

7.4.3.1 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 predictor is the treatment as a fixed effect. The average difference in log reduction between treatments at 10 minutes post-treatment time point will be estimated. However, all conclusions for the average treatment effect will be based on the primary analysis only.

An example for results of comparisons above can be found in Tables 9.

Table 9: Example: Log reduction comparisons for Groin

Comparison	N Pairs	Difference	Conf Int
IP Baseline-IP 10 min vs. VC Baseline-VC 10 min	24	0.13	-0.44, 0.70
IP Baseline-IP 10 min vs. SC Baseline-SC 10 min	24	0.11	-0.65, 0.87

7.4.3.2 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.

7.5 Graphs

7.5.1 Average Treatment Effect Per Primary Analysis

Graphs for least squares means with 95% intervals estimated from the regression model for each treatment will be provided (see an examples in Figure 1).

Graphs for the average treatment effect between IP vs. VC/SC with 95% intervals (as in Table 8) will be provided (see an example in Figure 2).

Groin

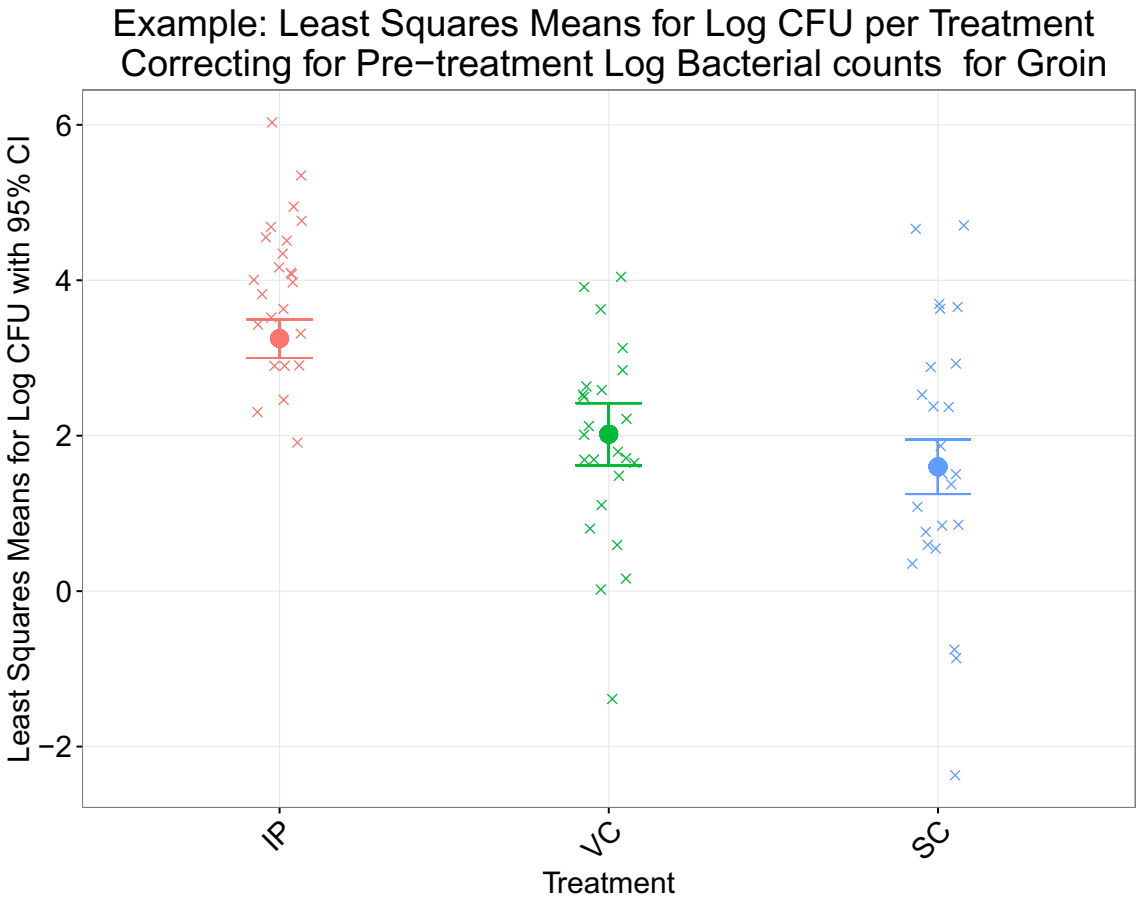

Figure 1: Mean with confidence interval plots for log reduction for Groin

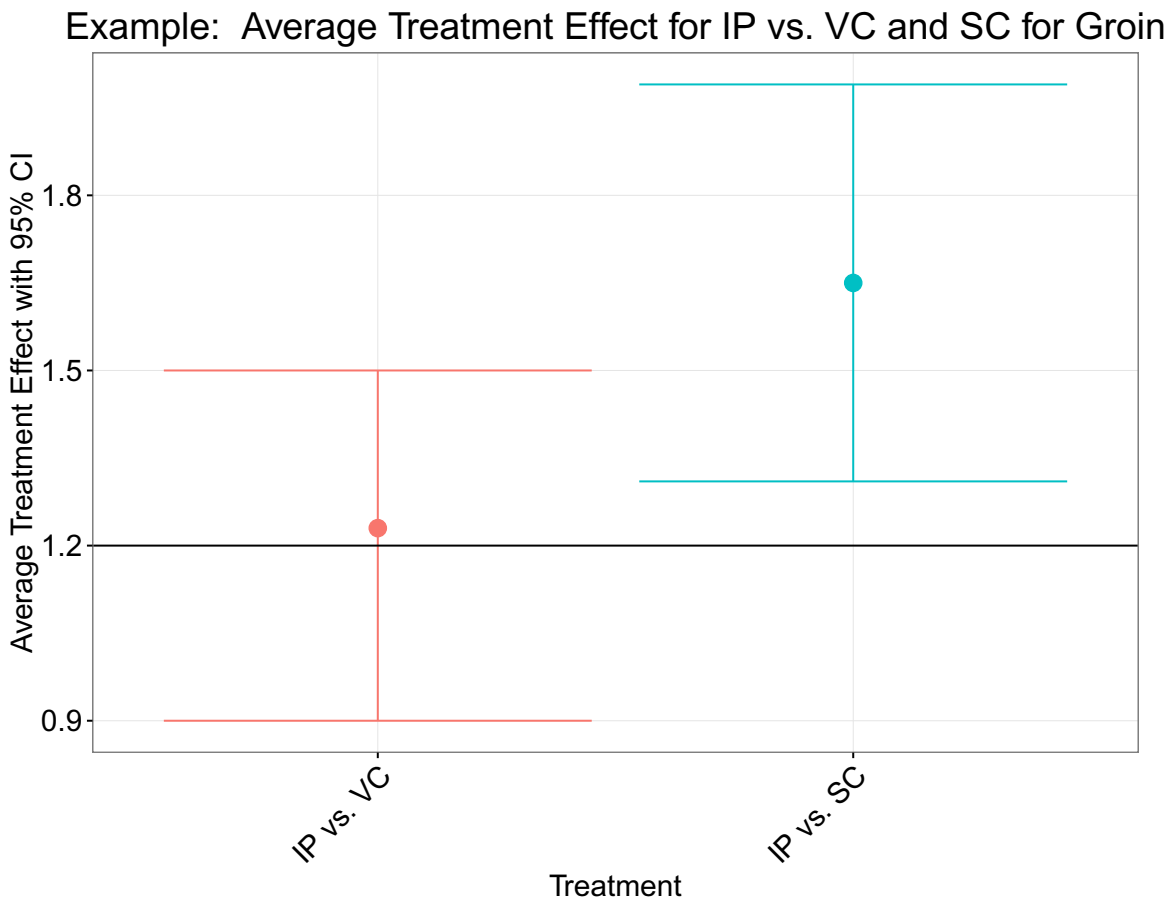

Figure 2: Average difference between products in log reduction from baseline for Groin

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.

9 References

References

[1] Alex Dmitrienko, Ajit C.Tamhane, Frank Bretz Multiple Testing Problems in Pharmaceutical Statistics

Signature Page for MPS-17IPVAW10_Statistical_Analysis_Plan

Approved	Lingzhi Li Statistician, Corporate/Shared Services Statistician 23-Jun-2017 13:40:59 GMT+0000

Approved	Julie Berube Manager Statistics, Corporate/Shared Services Functional Management 26-Jun-2017 12:56:35 GMT+0000

Approved	Carole Schatz Study Manager 29-Jun-2017 22:17:31 GMT+0000

Signature Page for MPS-17IPVAW10_Statistical_Analysis_Plan - System Version No.