CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-902

STATISTICAL REVIEW(S)

US Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

Statistical Review and Evaluation

New Drug Application

Clinical Studies

NDA/Serial Number:	21-902/SN000
Drug Name:	Polyphenon E® ointment([redacted] 15%)
Indication(s):	Genital warts
Applicant:	MediGene
Dates:	Submitted: 09/30/2005 PDUFA: 07/28/2006
Review Priority:	Standard Review
Biometrics Division:	Division of Biometrics III
Statistics Reviewer:	Mat Soukup, Ph.D.
Concurring Reviewer:	Mohamed Alosh, Ph.D.
Medical Division:	Division of Dermatology and Dental Products
Clinical Reviewer:	Elektra Papadopolous, M.D.
Clinical Team Leader:	Jill Lindstrom, M.D.
Project Manager:	Millie Wright
Keywords:	Botanical Product

Contents

1 Executive Summary

1.1 Conclusions and Recommendations

Polyphenon E® is a proprietary extract of green tea leaves containing more than 80% of tea polyphenols. MediGene, Inc. is seeking approval of this new botanical drug product for the treatment of genital warts. To examine the safety and efficacy of this product two international, pivotal trials were conducted using two doses of Polyphenon E®, 10% ointment and 15% ointment applied three times daily. In both trials, both active treatments were superior to vehicle in each study where the primary endpoint was defined as the absence of all warts (both baseline and any new warts) by week 16. Also both studies did not show either of the active treatments to be significantly different from each other in terms of efficacy. The adverse event rate of 15% ointment and 10% ointment were similar. Two subjects treated with 15% ointment had serious adverse events related to application site reactions, and one subject treated with 10% ointment had a serious adverse event for vulvovaginitis.

1.2 Brief Overview of Clinical Studies

Two pivotal Phase 3 safety and efficacy studies, CT1017 and CT1018 were conducted to compare 15% ointment and 10% ointment to vehicle. Treatment duration was up to 16 weeks where subjects could complete the trial prior to week 16 if all warts were resolved. Study CT1017 is a randomized, double-blind, multi-center, three-arm, placebo-controlled, international trial conducted in 46 investigative centers in the Czech Republic, Denmark, Netherlands, Norway, Poland, Romania, Russia, and South Africa. Study CT1018 is a randomized, double-blind, multi-center, three-arm, placebo-controlled, international trial conducted in 50 investigative centers in Argentina, Chile, Columbia, Mexico, Peru, Romania, and the United States. Study CT1017 enrolled 503 subjects which were included in the ITT population while study CT1018 enrolled 502 subjects. The objective of both trials was to show the superiority of each active, 15% ointment and 10% ointment, to vehicle based upon the percent of subjects that had no warts by week 16.

1.3 Statistical Issues and Findings

The initial submission of the efficacy analyzable data set contained only imputed data values for the endpoints of interest. In the 74 day letter (December 8, 2005), the Agency asked the sponsor to also submit the raw data. The sponsor's response was received on January 6, 2006, but the raw data files did not contain data definition files and it was unclear how to use such data. On February 1, 2006 the Agency then sent a request for information for the sponsor to submit the data in specific format which should contain data for each subject at every visit; if

the subject missed the visit the data should be stored as NA. On March 6, 2006 the sponsor resubmitted the data which did not meet the Division's request (i.e. not all subjects contained all visits as was requested).

On May 23, 2006 the Agency had a teleconference with the sponsor again requesting an efficacy analyzable data set which could be used to perform analyses on the protocol defined primary population (ITT-LOCF), and the Agency also requested the sponsor submit data to assess relapse rates. On May 31, 2006 the sponsor submitted SAS code which used the raw data to create modified data sets and the analysis was based on the modified data sets. However, the modified data sets used by the sponsor are the same as those previously submitted to the NDA which the Agency conveyed were deficient for the assessment of the primary analysis on the protocol defined primary population. Consequently the reviewer filled in the missing structures in the sponsor's data set to more accurately capture the data contained in the CRF to allow for the assessment of the protocol defined primary analysis population. However, it should be noted that such an attempt at creating a working data set for efficacy evaluation may not fully address all the missing data issues as the reviewer does not have access or resources to ensure the quality of the data.

Based upon the protocol defined primary analysis of the proportion of subjects that had no warts by week 16, both studies showed 15% ointment and 10% ointment to be statistically significant when comparing to vehicle (refer to Table 1). The results from Study CT1018 were robust, but the results from Study CT1017 were not as robust but did meet protocol defined primary analysis. Note that the sponsor's p-values differ from the reviewer's as the sponsor did not include subjects with baseline data only which were considered to be part of the protocol defined ITT population.

Table 1: Primary Endpoint Efficacy Results (ITT-LOCF)

	Study CT1017			**Study CT1018**		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
Fail	100 (50.3)	99 (49.3)	65 (63.1)	91 (45)	85 (43.4)	69 (66.3)
Success	99 (49.7)	102 (50.7)	38 (36.9)	111 (55)	111 (56.6)	35 (33.7)
p-value[1]	0.0384	0.0284	-	<0.001	<0.001	-
p-value[2]	0.0280	0.0143	-	<0.001	<0.001	-

[1] Source: Reviewer's Analysis using Fisher's exact test.

[2] Source: Table 11.11 in CTD 5.3.5.1.2 (CT1017) and Table 11.9 in CTD 5.3.5.1.3 (CT1018) using Fisher's exact test.

Only 10% of subjects enrolled in Study CT1018 were from the United States (no US sites were used in Study CT1017) and results by country revealed the US to have lower percentages

of subjects with resolution of all warts (refer to Figure 15 on page 34). As an attempt to explain this phenomenon a Classification and Regression Tree (CART) was used to see what factors would predict subject success. This analysis found that subjects who are diagnosed further in the past (CART root node split of 447 days) tend to not respond to treatment as well as subjects who are recently diagnosed. Looking at the subjects enrolled in the US revealed that a large portion of US subjects (62.5%) were diagnosed more than 447 days (CART defined split). Further, the sponsor provided documentation that treatment in circumcised males is not as responsive as in uncircumcised males for which the US has a higher proportion of circumcised males.

According to the system organ classification (MedDRA nomenclature) 85.2% and 86.6% of subjects experienced general disorders and application site reactions for 10% ointment and 15% ointment, respectively, which were larger than the vehicle arm (66.0%). The second most frequently recorded AE's occurred for skin and subcutaneous tissue disorders at rates of 23%, 25%, and 9.2% for 10% ointment, 15% ointment, and vehicle, respectively. One subject receiving 10% ointment and two subjects receiving 15% ointment reported serious AE's that were recorded as having a probable casual relationship to treatment.

2 INTRODUCTION

In the original NDA the data sets referenced by the data definition file contained only derived data (i.e. after imputation of missing data without a flag indicating it was missing) making it very difficult to identify which values were imputed. On December 8, 2005 the Agency requested raw data files (i.e. prior to imputation of missing data). The sponsors response to the Agency's request on January 6, 2006 contains two folders CT1017rn and CT1018rn with raw data files. However, the format and names of these data sets differ from the data sets in the original submission, which are the only data sets with a corresponding data definition file. The sponsor was then asked on February 1, 2006 to provide data sets which contain raw data (prior to imputation) for each of the pivotal studies in which *each* subject had visits 1 through 11 in the data set. If the subject did not attend the visit, all missing data were to be depicted by NA (communication sent to the sponsor depicted an example of how the data should be structured). Note that visit 1 corresponds to baseline, visits 9 is the final treatment visit (if needed) at week 16, and visits 10 and 11 are follow-up visits four and twelve weeks after clearance of all warts.

On March 6, 2006 the sponsor submitted the efficacy data sets (`WART17.XPT` and `WART18.XPT`). However, in close inspection of the efficacy data sets it was noted that several subject profiles did not follow the Division requests (i.e. not all subjects contained a row in the data set which corresponded to each visit: 1 through 11). For each pivotal study and corresponding data set, the discrepancies are described in Section 2.2. Also, for a vast majority of subjects who achieved treatment clearance by week 16, the treatment follow-up visits were recorded as missing data.

Thus, for subjects that resolved all warts by week 16, it is not clear if these subjects actually missed and did not attend any of the follow-up visits or if no warts were seen in the follow-up visits and subsequently recorded as missing rather than as 0 warts seen at the follow-up visit. As a result it is difficult to assess relapse of genital warts.

On May 23, 2006 the Agency had a teleconference with the sponsor again requesting an efficacy analyzable data set which could be used to perform analyses on the protocol defined primary population (ITT-LOCF), and the Agency also requested the sponsor submit data to assess relapse rates. On May 31, 2006 the sponsor submitted SAS code which used the raw data to create modified data sets and the analysis was based on the modified data sets. However, the modified data sets used by the sponsor are the same as those previously submitted to the NDA which the Agency conveyed were deficient for the assessment of the primary analysis on the protocol defined primary population. To address the issue of assessing relapse, the sponsor referred to the `WRTFUALL` data sets (one for each pivotal study). However, it should be noted that these data sets which should include all subjects who cleared during the treatment phase of the trial contained fewer subjects than the number of subjects reported to be clear in the efficacy data sets.

2.1 Overview

Polyphenon E® is a proprietary extract of green tea leaves containing more than 80% of tea polyphenols. The sponsor has conducted a number of clinical studies in the clinical development of Polyphenon E® ointment. However, based upon the submission only 3 clinical studies were submitted with sufficient details to assess safety and efficacy. Study descriptions for the three clinical studies are provided in Table 2. To establish the safety and efficacy of Polyphenon E® ointment the review uses data only from Studies CT1017 and CT1018. The data for Study CT1005 is not included as dosing duration and formulation differ from that used in the two pivotal studies. Throughout the review of Polyphenon E® ointment 15% is abbreviated as 15% ointment within the body of the review and Oint15 in tables and figures (similarly for Polyphenon E® ointment 10%).

Table 2: Efficacy and Safety Studies Overview

Study	Phase Objective	Drug Products	Number Subjects	Treatment Duration	Date[1]
CT1005	Phase 2/3	15% Oint	80	Up to 12 weeks	12/2000 – 06/2001
		Oint Vehicle	40		
		10% Cream	79		
		Cream Vehicle	43		
CT1017	Pivotal Phase 3 Superiority	10% Oint	199	Up to 16 weeks	08/2002 – 08/2003
		15% Oint	201		
		Oint Vehicle	103		
CT1018	Pivotal Phase 3 Superiority	10% Oint	202	Up to 16 weeks	07/2003 – 08/2004
		15% Oint	196		
		Oint Vehicle	104		

[1] Dates correspond to the start and end of the study.

2.2 Data Sources

The review of safety and efficacy uses the data from two studies, CT1017 and CT1018. The narrative below describes the problems with the data submitted to the NDA for statistical review for each study.

- Study CT1017
  - Eleven subjects had only a baseline visit. In the `WART17` data set only one line of data was included rather than 12 with missing data for missed visits. Note that these are considered ITT evaluable per the protocol definition of ITT population.
  - One subject, ID: 02673, is not included in `WART17.XPT`. This subject treated with 15% ointment had a protocol violation (not otherwise specified) and included only baseline data. Baseline data from the CRF was included in the efficacy data set.
- Study CT1018
  - Seven subjects had only a baseline visit. Again, note that these seven subjects are considered ITT evaluable per the protocol definition of ITT population.
  - Three subjects had no visit 9 data. Visit 9 data was recorded as missing.
  - Two subjects had no data for visit 5. Week 5 data for these two subjects was recorded as missing.
  - Many subjects did not have baseline visit data (visit 1). However they did have visit 0 (screening) and visit 2 (week 2) data. Thus baseline data (i.e. visit 1) was imputed

using visit 0 data when no visit 1 data was available. By imputing data in such a manner, it is under the assumption that the screening visit and the baseline visit are the same visit.

As subjects with only baseline data did not contain missing data recordings at visits 2-11 these subjects were not reported in the sponsor's efficacy results despite being recorded as ITT evaluable. New data sets were created making the corrections described above. Specifically, the subjects with only baseline data excluded from the sponsor's ITT-LOCF efficacy results are listed in Tables 25 and 26 in Section A.3 of the Appendix on page 41. Consequently, no specific efficacy data sets in the EDR contain all the efficacy data used in this review.

In the WART17 and WART18 data sets unless a subject *had* warts present at the follow-up visits, the number of warts were recorded as missing. Using such a method of data record-keeping makes it difficult to assess whether a subject had no warts at the visit or if the subject missed the visit. In response to the teleconference held with the Agency on May 23, 2006 the sponsor referred the Agency to the WRTFUALL data sets. The WRTFUALL data sets contained a variable (MISS) to indicate which values were missing. However, for both data sets none of the data were recorded as missing implying all subjects attended both follow-up visits which seems to be unlikely. Also, both data sets failed to include all subjects who were recorded as treatment successes in the efficacy data sets. Specifically, in Study CT1017 239 subjects achieved treatment success, yet 236 subjects were included in the WRTFUALL data set. In Study CT1018 257 subjects achieved treatment success, but 251 subjects were included in the WRTFUALL data set.

It should also be noted that the DISP.XPT data set in the isedern folder does not match the DISP.XPT data set in the sponsor's issdern folder. The review uses the data in the sponsor's isedern folder as this is the data set that matches the study reports.

Reviewer Comment: As the sponsor's submitted data file does not account for all patient visits and some data sets contain unmatched data, the efficacy data sets used in the efficacy analysis are based on the most accurately record of the data provided in the individual CRF's while also recording missing data. However, it should be noted that such an attempt at creating a working data set for efficacy evaluation may not fully address all the missing data issues as the reviewer does not have access or resources to ensure the quality of the data.

3 Statistical Evaluation

3.1 Evaluation of Efficacy

3.1.1 Study Design

The sections below describe protocol descriptions of the study design for each study along with any protocol amendments that may impact the study design. The initial design of the study provided in the protocol was agreed upon with the Division according to the Division's communications with the sponsor.

3.1.1.1 Study CT1017 Study CT1017 is a randomized, double-blind, multi-center, three-arm, placebo-controlled, international trial conducted in 46 investigative centers in the Czech Republic, Denmark, Netherlands, Norway, Poland, Romania, Russia, and South Africa. Treatment duration is a maximum of 16 weeks. Subjects are randomized in a 2:2:1 ratio to 15% ointment, 10% ointment, and vehicle, respectively. The objectives of the trial were to demonstrate the safety and efficacy of either active arm.

At enrollment subjects were to be at least 18 years of age with between 2 and 30 genital warts over a wart area between 2 and 600 mm^2. Subjects who received treatment for genital warts 30 days within study enrollment and those who have previously treated with Polyphenon E® were excluded from the study. Overall, the study enrolled 503 subjects that were included in the ITT population.

Eleven visits are scheduled occurring at baseline, weeks 2, 4, 6, 8, 10, 12, 14, 16, 20, and 28. The last two visits are post-treatment follow-up visits to assess for recurrence of new or baseline warts. However, if a subject cleared of all warts prior to week 16, they were to return 4 weeks and 12 weeks following the visit in which he/she was as cleared.

3.1.1.2 Study CT1018 Study CT1018 is a randomized, double-blind, multi-center, three-arm, placebo-controlled, international trial conducted in 50 investigative centers in Argentina, Chile, Columbia, Mexico, Peru, Romania, and the United States. Overall, the study enrolled 502 subjects that were included in the ITT population. The only difference in the design of Study CT1018 is that a protocol amendment was made on May 22, 2003 to include a screening a visit of up to 14 days. This resulted in 64 subjects that were screened but not randomized. It does not appear that such a change to the study design was communicated to the Agency prior to making the change.

3.1.2 Endpoints

The primary endpoint is the proportion of subjects who have clearance of all warts (baseline and new) by week 16. The secondary endpoints include

- time to complete clearance
- proportion with at least 75% clearance of all warts
- Proportion of subject who clear and then have recurrent genital warts.

3.1.3 Patient Disposition, Demographic, and Baseline Characteristics

3.1.3.1 Patient Disposition Ninety-two subjects out of 503 discontinued Study CT1017. The reason for discontinuation is shown below in Table 3. Results show a higher percentage in the 15% ointment arm discontinued due to an adverse event than in the other treatment arms.

Table 3: Compliance for Study CT1017*

	Oint10 $N = 199$	Oint15 $N = 201$	Vehicle $N = 103$
Total Discontinued	29 (14.6%)	40 (19.9%)	23 (22.3%)
Reason for Discontinuation			
Adverse Event	0% (0)	15% (6)	4% (1)
Death	0% (0)	0% (0)	0% (0)
Warts require treatment @ follow-up	0% (0)	0% (0)	0% (0)
Non-compliance	24% (7)	15% (6)	9% (2)
Lost to follow-up	0% (0)	8% (3)	9% (2)
Withdrew consent	34% (10)	40% (16)	39% (9)
Protocol violation	3% (1)	10% (4)	4% (1)
Lack of efficacy	21% (6)	5% (2)	26% (6)
Inclusion criteria violation	0% (0)	0% (0)	0% (0)
Administrative reasons	0% (0)	0% (0)	0% (0)
Investigator decision	0% (0)	0% (0)	0% (0)
Other	17% (5)	8% (3)	9% (2)
Missing	0% (0)	0% (0)	0% (0)

* Numbers after percent of subjects that discontinued are frequencies.

Source: DISP.XPT data set in the isedern folder.

Ninety-eight subjects withdrew from Study CT1018. In all three arms a large portion of the discontinued subjects were the results of lost to follow-up, withdrawn consent, and lack of efficacy. Results are provided in Table 4.

Table 4: Compliance for Study CT1018*

	Oint10 $N = 202$	Oint15 $N = 196$	Vehicle $N = 104$
Total Discontinued	40 (19.8%)	37 (18.8%)	21 (20.2%)
Reason for Discontinuation			
Adverse Event	0% (0)	3% (1)	0% (0)
Death	0% (0)	0% (0)	0% (0)
Warts require treatment @ follow-up	8% (3)	3% (1)	0% (0)
Non-compliance	10% (4)	11% (4)	5% (1)
Lost to follow-up	15% (6)	11% (4)	19% (4)
Withdrew consent	38% (15)	38% (14)	19% (4)
Protocol violation	5% (2)	8% (3)	0% (0)
Lack of efficacy	10% (4)	19% (7)	29% (6)
Inclusion criteria violation	0% (0)	0% (0)	0% (0)
Administrative reasons	0% (0)	3% (1)	0% (0)
Investigator decision	2% (1)	3% (1)	0% (0)
Other	12% (5)	3% (1)	29% (6)
Missing	0% (0)	0% (0)	0% (0)

* Numbers after percent of subjects that discontinued are frequencies.

Source: DISP.XPT data set in the isedern folder.

The protocol defined ITT population consists of all subjects randomized and dispensed medication. The protocol defined PP population consists of all subjects in the ITT population who complete the study without any major protocol violations. The resulting number of subjects per treatment arm for each population and study is provided in Table 5.

Table 5: Analysis Populations

	Study CT1017			**Study CT1018**		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
ITT pop^n	199	201	103	202	196	104
PP pop^n	174	171	93	171	165	95

Source: Reviewer's analysis.

Reviewer Comment: Note the study reports submitted by the sponsor do not follow the protocol defined ITT population. Rather the sponsor uses two different ITT populations each summarized below.

- *Efficacy analyzable: All subjects who were randomized and had at least one post-randomization observation.*

- *16-week completers: All subjects with available data for the week 16 timepoint.*

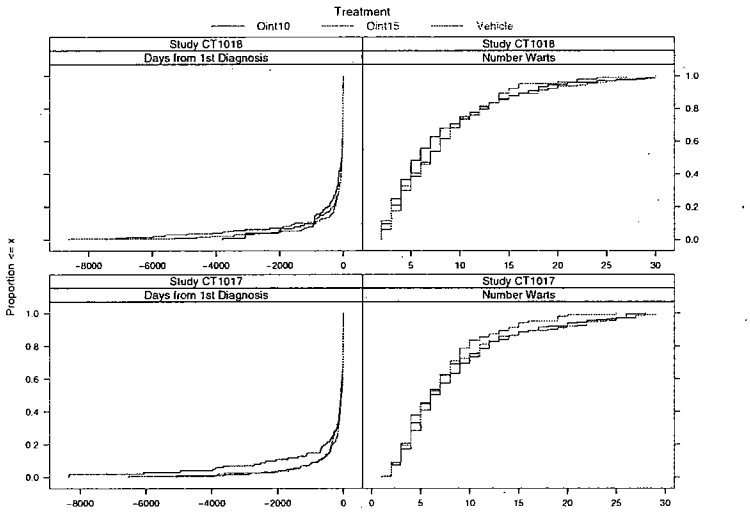
The sponsor cites E9 for the choice of the above ITT populations, but as the populations are defined post hoc and possibly after study unblinding, these populations are not considered to be primary.

3.1.3.2 Baseline Assessment Baseline characteristics for gender, race, circumcision (yes or no), age, and BMI were assessed for each study. Comparisons across treatment groups did not reveal any noticeable differences. Tabled results are provided in the Appendix section A.1 on page 40.

In addition to these baseline factors, two prognostic factors which may have an impact on efficacy were explored. The first factor was the number of days from date of first diagnosis. It is theorized that subjects with longer dates of first diagnosis may be harder to treat. The other factor is the number of warts present at baseline. It is theorized that subjects with fewer warts would be easier to treat than subjects with a large number of warts present.

Figure 1 shows the empirical distribution function for each of these prognostic factors by treatment arm. As there is a high degree of overlap in the three lines this implies the distributions for these two prognostic factors are quite similar for each treatment arm.

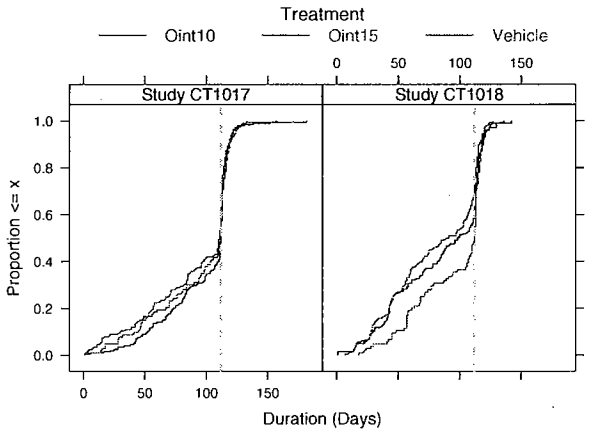
Figure 1: Baseline Factors



3.1.3.3 Treatment Duration The time on treatment is provided in Figure 2. The vertical line in the figure corresponds to Day 112 (end of week 16) which is the last day of treatment. Overall, Study CT1018 tended to have shorter treatment times than Study CT1017. In fact,

more than 50% of subjects in Study CT1018 randomized to 15% ointment and 10% ointment were on treatment less than 112 days. Both studies show that treatment duration tended to be shortest for the 15% ointment treatment. Tabled output of treatment duration is provided in the Appendix section A.2 on page 41.

Figure 2: Treatment Duration



In the two studies 39 subjects had recorded treatment durations of more than 122 days. In fact, one subject on the 15% ointment treatment arm actually had a recorded dosing of 182 days in Study CT1017 (ID: 02132). As the longer treatment duration may impact efficacy claims, the effect of longer treatment durations is explored as a sensitivity analysis in Section 3.1.7.1 on page 17.

3.1.4 Statistical Methodology

The sponsor had several communications with the Agency described below with notes on major statistical issues raised.

- Guidance Meeting on June 11, 2001
  - ITT should be primary analysis population and LOCF is acceptable for imputation.
  - Study should be planned to enroll 10 subjects per treatment arm per center.
- End of Phase 2 Meeting on November 19, 2001
  - Test of significance should combine both genders and then test for significance.

  - Study should be planned to enroll 8 subjects per treatment arm per center.

- Special Protocol Assessment on June 12, 2002

  - ITT should be primary analysis population and LOCF is acceptable for imputation.

- fax sent on September 6, 2002

  - ITT defined as primary analysis population and LOCF is acceptable for imputation.
  - Since two doses, test should be carried out at the two-sided significance level of $\alpha = 0.025$.
  - All other protocol defined analyses are acceptable.

- fax sent on December 24, 2002

  - Reiterated that since two doses, test should be carried out at the two-sided significance level of $\alpha = 0.025$.

- fax sent on Dec 31, 2002

  - Agreed to testing the two doses using Hochberg procedure at the two-sided $\alpha = 0.05$ level.

With each communication the sponsor was provided statistical comments based on the Phase 3 protocol to which the sponsor and Agency were in agreement. The agreements for the statistical methodology which are provided in the protocol follow.

The primary analysis of the proportion of subjects with no warts by the week 16 will be conducted using Fisher's exact test using the ITT population. The ITT population is defined as all subjects randomized and dispensed study medication. Primary method of imputation will use LOCF. Hochberg's procedure will be used to adjust for the multiple comparisons of the two actives versus placebo.

All analysis on the ITT population will follow the protocol defined statistical methods. As mentioned in Section 3.1.3.1, the efficacy results based on the ITT population provided in the study reports uses alternate definitions of the ITT population. As it is not clear if this occurred after study unblinding, this in turn may affect the Type I error. Consequently, to control the Type I error, all analyses are based upon the protocol definition of ITT population.

3.1.5 Primary Efficacy Results (ITT)

The protocol defined the ITT population as all subjects randomized and dispensed treatment. The results provided in the study reports do not completely follow this definition as several subjects with only baseline data were excluded from the results in the study reports. Specifically,

twelve subjects were excluded from the analysis in Study CT1017 (4-10% ointment , 7-15% ointment , 1-vehicle) and seven subjects from Study CT1018 (5-10% ointment , 2-15% ointment , 0-vehicle). Subjects excluded from the analysis are provided in the Appendix section A.3 on page 41.

Efficacy results for the protocol defined ITT population are provided in Table 6. Also provided in the summaries are the results listed in the study reports which exclude the patients with baseline data only. Recall that the protocol defined multiplicity adjustment is based upon Hochberg's procedure. Thus, if both p-values are less than $\alpha = 0.05$ both reach statistical significance. If one of the two p-values is not significant at $\alpha = 0.05$, the other must be less than $\alpha = 0.025$ to reach statistical significance.

As all p-values are below $\alpha = 0.05$ for both studies, the superiority comparisons of 10% ointment and 15% ointment to vehicle reach statistical significance. In study CT1018, response rates of the active arms were higher than in Study CT1017 and vehicle response was lower. Consequently p-values for Study CT1018 are smaller than p-values in Study CT1017.

Table 6: Primary Endpoint Efficacy Results (ITT-LOCF)

	Study CT1017			**Study CT1018**		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
Fail	100 (50.3)	99 (49.3)	65 (63.1)	91 (45)	85 (43.4)	69 (66.3)
Success	99 (49.7)	102 (50.7)	38 (36.9)	111 (55)	111 (56.6)	35 (33.7)
p-value[1]	0.0384	0.0284	-	<0.001	<0.001	-
p-value[2]	0.0280	0.0143	-	<0.001	<0.001	-

[1] Source: Reviewer's Analysis using Fisher's exact test.

[2] Source: Table 11.11 in CTD 5.3.5.1.2 (CT1017) and Table 11.9 in CTD 5.3.5.1.3 (CT1018) using Fisher's exact test.

Based upon Table 6, it can be seen the treatment effects in Study CT1018 are larger than in Study CT1017. The increase for the larger treatment effect in Study CT1018 versus Study CT1017 is caused by an increased response in the actives and a decrease in the response of the vehicle. Overall, this increase in the treatment effect elicits the increased statistical significance in Study CT1018 in comparison to Study CT1017.

3.1.6 Primary Efficacy Results (PP)

As a supportive analysis the primary endpoint was analyzed using the PP population. The PP population consisted of subjects that completed the study without major protocol violations. Results are shown in Table 7. The treatment effects (active-vehicle) observed in the ITT are

similar to those seen in the PP population, but due to the smaller sample size, the comparison of 10% ointment to vehicle does not reach significance at the $\alpha = 0.05$ level for Study CT1017.

Table 7: Primary Endpoint Efficacy Results (PP)

	Study CT1017			**Study CT1018**		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
Fail	86 (49.4)	77 (45)	57 (61.3)	73 (42.7)	67 (40.6)	62 (65.3)
Success	88 (50.6)	94 (55)	36 (38.7)	98 (57.3)	98 (59.4)	33 (34.7)
p-value[1]	0.072	0.0143	-	<0.001	<0.001	-

[1] Source: Table 11.15 (CT1017) and Table 11.13 (CT1018) using Fisher's exact test.

3.1.7 Sensitivity Analysis of the Primary Endpoint

The following sections are sensitivity analyses performed by the reviewer and not defined in the protocol. As the results for Study CT1018 are strong and not likely to be impacted by small changes in the number of successes, this study is not included in the sensitivity analyses that follow.

3.1.7.1 Examination of Subjects Treated More than 122 Days For subjects treated more than 122 days a breakdown of success rate for each study and treatment arms is provided in Table 8. Note that durations were based upon the sponsor's DEMO data set which defined treatment duration as:

- Duration = Last Day Dosed - First Day Dosed + 1

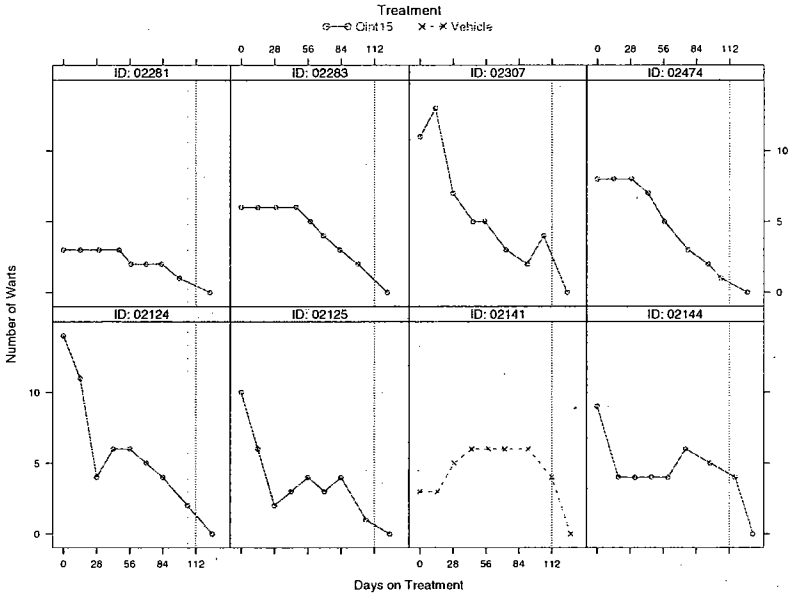
Table 8: Efficacy for Subjects Treated More than 122 days

	Oint10	Oint15	Vehicle
Study CT1017	0/6 (0%)	7/13 (53.8%)	1/4 (25%)
Study CT1018	1/8 (12.5%)	2/4 (50%)	0/4 (0%)
Total	1/14 (7.1%)	9/17 (52.9%)	1/8 (12.5%)

Source: Reviewer's Analysis.

Of the 102 treatment successes on the 15% ointment arm, seven were treated more than 122 days. And of the thirty-eight vehicle successes, one was treated more than 122 days. No subjects

Figure 3: Efficacy Profiles for Eight Subjects Treated More than 122 Days in Study CT1017



treated with 10% ointment more than 122 days resulted in treatment success. The profiles for each of these eight subjects is provided in Figure 3.

A vertical line is placed at day 112 (end of week 16) along with dotted lines at days 105 and 119 which are ± 7 days from end of treatment as defined in the protocol. For six of the eight subjects, the time of their evaluation at visit 8 fell into this window, and at this time each had at least one wart. Thus, these subjects might be considered treatment failures by week 16. Two of the subjects (ID's: 02281 and 02283) had their visit 8 evaluation less than 105 days from first dose. Consequently it is unclear how to treat these subjects. Therefore, two partitions of this analysis will take place. The first partition treats all subjects who were treated more than 122 days as failures and the second considers subjects 02281 and 02283 to be treatment success (results shown in Table 9).

Based upon this analysis, the comparison of 10% ointment to vehicle improves from the original analysis as the response rate in vehicle slightly decreases and the response rate in 10% ointment does not change ($p = 0.02791$). When all subjects are treated as failures, the comparison of 15% ointment to vehicle fails to reach statistical significance ($p = .06644$). This would ultimately result in a failure of the comparison of 10% ointment to vehicle to reach statistical significance at the $\alpha = 0.025$ level (Hochberg's procedure). Even when subjects 02281 and 02283 are not treated as treatment failures, neither comparison reaches statistical significant.

Table 9: Sensitivity Analysis Efficacy Results

	All Failures			022881 and 02283 Success		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
Fail	100 (50.3)	106 (52.7)	66 (64.1)	100 (50.3)	104 (51.7)	66 (64.1)
Success	99 (49.7)	95 (47.3)	37 (35.9)	99 (49.7)	97 (48.3)	37 (35.9)
p-value[1]	0.02791	0.06705	-	0.02791	0.0507	-

Source: Reviewer's analysis using Fisher's exact test.

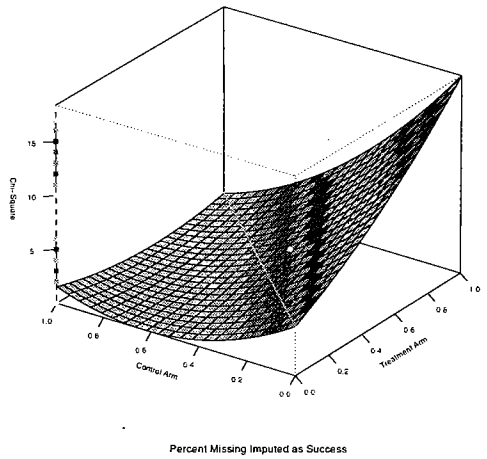
3.1.7.2 Sensitivity Analysis to Method of Data Imputation In the following sensitivity analysis to data imputation, all missing data are imputed using various proportions of successes for the missing data. This can vary from the extremes, all missing data for the control arm are imputed as successes and all missing data from the active arm are imputed as failures to the case where all missing controls are failures and all missing active are success. Everything in between the extremes is covered in this analysis. Once imputed these data are combined with the complete data and a Chi-square test is performed. The Chi-square test is performed for every possible proportion of imputed successes and the response surface of the Chi-Square statistic is plotted in a perspective plot.

First the sensitivity analysis is performed comparing 10% ointment to vehicle. Figure 4 is a perspective plot depicting the response surface of all possible ways to impute success. The gray line dissecting the surface in half corresponds to the cases where the imputation of the active and control arms are equal. To reach statistical significance at the $\alpha = 0.05$ level (i.e. assuming no multiplicity adjustment), the value of the Chi-square statistic should be 3.84 or greater. This value is represented between blue ($\chi^2 = 3$) and cyan or light blue ($\chi^2 = 4$) in the perspective plot. Thus, for points falling in the cyan range, this area would correspond to statistical significance. Any range above this would also correspond to statistical significance.

The rate used by LOCF is shown by a white dot on the figure and is consistent with conclusions based on Fisher's exaxt test. Considering the cases when both arms are imputed with the same ratio, only cases where the proportion missing imputed as successes is near zero does the sensitivity analysis show significance. Even cases when the missing data in the treatment arm is imputed with higher success rates than the control arm fail to reach statistical significance. Also note that if all missing were imputed as failures, statistical significance is achieved. If all missing were imputed as success, the test would fail to reach statistical significance. Thus the conclusion is that while results are not robust to the proportion imputed as success, the primary method of imputation, LOCF, did meet the pre-specified criteria.

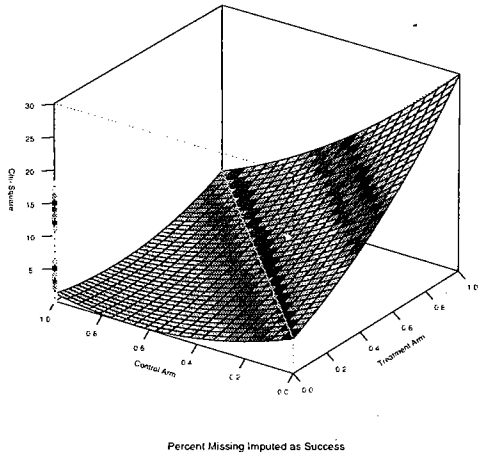
Similar to the sensitivity analysis comparing 10% ointment and vehicle, the same type of analysis is performed comparing 15% ointment and vehicle. Results are displayed in Figure 5.

Figure 4: Sensitivity Analysis CT1017 (10% ointment vs. vehicle)



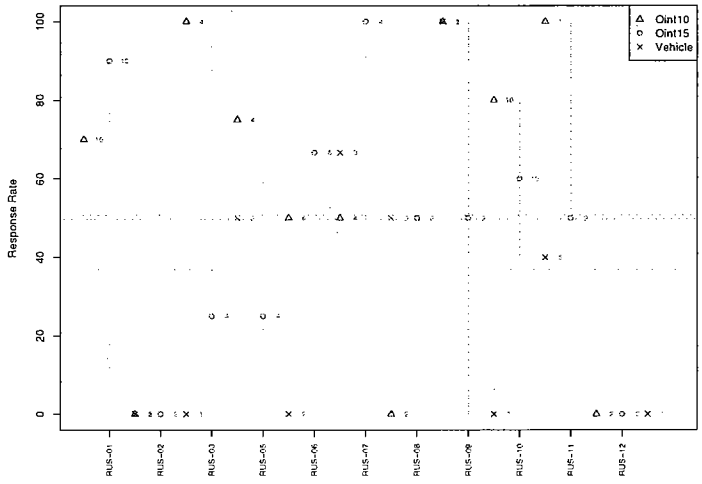
This sensitivity analysis shows that when the same proportion imputed as successes is equal for both the treatment and placebo arm, the test reaches statistical significance. Even in cases when the proportion imputed as successes is slightly lower in the treatment arm than the control arm, the test still reaches statistical significance. Thus, results from this study are quite robust to method of imputation. Note that the above sensitivity analyses were performed treating the subjects discussed in Section 3.1.7.1 above as successes despite them having been treated more than 122 days.

Figure 5: Sensitivity Analysis CT1017 (15% ointment vs. vehicle)



3.1.7.3 Sensitivity Analysis of Influential Center(s) A sensitivity analysis was performed to check for any influential centers that may drive efficacy claims. In Study CT1017 a single Russian site, denoted as RUS-01 in the sponsor's datasets, enrolled a relatively large number of subjects and treatment effects were much larger than the observed treatment effects including all the data. Figure 6 depicts the size of the observed treatment effect for each Russian site, the sample size enrolled for each arm, and the overall study means (horizontal lines).

Figure 6: Efficacy Results within Russia



Based upon Figure 6, Russian site RUS-01 may have an effect on efficacy as the sample size is relatively large and the observed treatment effects are quite large in comparison to the overall study response rate. To further explore this a sensitivity analysis was performed first by removing the data from the analysis, and secondly by imputing the missing as the overall average response. In the latter case, the Russian site data imputed the two active arms as 5 treatment responders and 5 treatment failures, while the 5 RUS-01 vehicle cases were imputed as 2 successes and 3 failures. Results from this analysis are provided in Table 10.

The p-values obtained by Fisher's Exact Test are above $\alpha = 0.05$ for both treatment arms when excluding the RUS-01 data and also when imputing it using overall study means. Based upon the above analysis, the Division of Scientific Investigations (DSI) sent an investigator to inspect site RUS-01. The results of this inspection did not reveal any clear misconduct on the part of the investigators at this site.

Table 10: Sensitivity Analysis of Russian Site 01

	Oint10	Oint15	Vehicle
Overall			
X/n	99/199 (49.7)	102/201 (50.7)	38/103 (36.9)
Treatment Effect[2]	12.8	13.8	-
p-value[1]	0.0384	0.0284	-
Exclude RUS-01			
X/n	92/189 (48.7)	93/191 (48.7)	38/98 (38.8)
Treatment Effect[2]	9.9	9.9	-
p-value[1]	0.1335	0.1341	-
Imputed RUS-01			
X/n	97/199 (48.7)	98/201 (48.8)	40/103 (38.8)
Treatment Effect[2]	9.9	10	-
p-value[1]	0.1136	0.1141	-

[1] Fisher's Exact Test.

[2] Response for active minus response for vehicle.

Source: Reviewer's Analysis.

3.1.8 Analysis of Secondary Endpoints

All analyses of the secondary endpoints below are performed on the ITT population imputing missing data by LOCF.

3.1.8.1 Time to Complete Clearance The time to complete clearance was defined as the number of days from baseline until complete clearance of warts. The protocol did not list a specific analysis method for the time to event analysis other than specify, "The time to complete clearance will be estimated using survival analysis methods..." As no specific analysis method was pre-specified, a log-rank test comparing each of the actives to vehicle were used to test for statistical significance. Subjects that either dropped out of the study without complete clearance of all warts or subjects that still had warts present at the end of the study were right-censored at the day from the last visit. Results from this analysis are provided in Table 11. In this analysis, Study CT1018 shows clear statistical significance, however, Study CT1017 does not establish a statistically significant difference. Further, an examination of the median times to complete clearance *for subjects that cleared of all warts during the trial* shows a shorter time to complete clearance of the two actives versus the vehicle in Study CT1018 whereas the medians of the actives in Study CT1017 are greater than the vehicle.

As the vehicle arm in Study CT1017 has a smaller median time to complete clearance than the active arms, several baseline characteristics were compared with the objective of explaining the larger median time to complete clearance in the active arms compared to the vehicle. Table 12 shows three baseline factors which might have contributed to such differences in the time

Table 11: Time to Complete Clearance Efficacy Results (ITT)

	CT1017			CT1018		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
Quartiles[1]	76 **110** 113	73 **109** 113	84 **104** 114	49 **83** 113	49.5 **72** 112	67.5 **97** 117
p-value[2]	0.108	0.067	-	0.001	< 0.001	-

Source: Reviewer's Analysis

[1] a **b** c correspond to the 25^{th}, 50^{th}, and 75^{th} percentiles (in days) for subjects that cleared during the trial.

[2] Log-rank test for difference in survival.

to complete clearance. It can be seen from the table that there is a larger proportion of males in the vehicle arm compared to the active arms. In addition, subjects on the vehicle arm also tended to have a larger number of baseline warts and larger days from diagnosis.

Table 12: Descriptive Statistics by Treatment

	Vehicle $N = 38$	Oint15 $N = 102$	Oint10 $N = 99$	Test Statistic
Gender : Male	55% $\frac{21}{38}$	47% $\frac{48}{102}$	47% $\frac{47}{99}$	$\chi^2_2 = 0.82$, $P = 0.663^2$
Baseline Warts	3.00 4.50 8.75	4.00 6.00 9.00	4.50 7.00 11.00	$F_{2,236} = 2.11$, $P = 0.124^1$
Days From Diagnosis	-214.75 -62.00 -2.50	-108.25 -21.50 -2.00	-144.50 -47.00 -3.00	$F_{2,236} = 1.75$, $P = 0.176^1$

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. Tests used: [1]Kruskal-Wallis test; [2]Pearson test

3.1.8.2 Proportion with 75% Clearance of all Warts The percent change in the number of warts from baseline was calculated as

$$100 \times \left[\left(\text{warts}_{base} - \text{warts}_{end}\right) / \text{warts}_{base}\right].$$

This percent change was then dichotomized to a success if the change was greater than or equal to 75%. Results are provided in Table 13 on the following page. Results show that both active treatment arms are superior to the vehicle in both studies.

Table 13: Proportion with at Least 75% clearance of all warts

	Study CT1017			Study CT1018		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
Fail	89 (44.7)	89 (44.3)	64 (62.1)	86 (42.6)	74 (37.8)	62 (59.6)
Success	110 (55.3)	112 (55.7)	39 (37.9)	116 (57.4)	122 (62.2)	42 (40.4)
p-value	0.0060	0.00471	-	0.0068	0.0005	-

Source: Reviewer's analysis using a Chi-square test.

3.1.8.3 Proportion that Relapse For each subject that cleared, subjects were instructed to return to the investigator at 4 weeks and 12 weeks after treatment clearance. Table 14 below shows the number and percentage of subjects that have relapsed within 12 weeks of treatment clearance. Note that relapse is defined as either new warts or recurrent appear after treatment success. This table shows that the relapse rate was quite similar across treatments and relatively low. However, recall that subjects who cleared by week 16 either a) did not attend the follow-up visit(s) or b) had zero recorded warts at the follow-up visit(s). Yet, the data sets recorded both situations as missing in the data sets without clarification if the subject attended the follow-up visit or not (refer to Section 2.2 for a discussion about details of the data recorded for follow-up visits). As the results reported in Table 14 assumed all missing follow-up data had no warts, it should be noted that the estimates reported in Table 14 likely underestimate the relapse rate as it is not known which subjects had zero counts and those of which the data is actually missing. All comparisons of active versus vehicle failed to reach statistical significance ($p > 0.75$). Note that this p-value should be interpreted with caution as each study was not powered for assessing relapse rates.

Table 14: Relapse Rates for Those Cleared

	Study CT1017			Study CT1018		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
x/n (%)	8/99 (8.1)	7/102 (6.9)	3/38 (7.9)	15/111 (13.5)	10/111 (9.0)	3/35 (8.6)
x/n (%)*	9/99 (9.1)	7/102 (6.9)	2/38 (5.3)	14/111 (12.6)	9/111 (8.1)	2/35 (5.7)

Source: Reviewer's analysis using the primary efficacy data set and assuming subjects with data recorded as NA's correspond to subjects with no relapse.

* Results reported in the study reports which adds the number of subjects with new warts and the number of subjects with recurrent warts.

Reviewer Comment: It should be noted that among the three protocol listed secondary endpoints.

only the proportion with 75% clearance of all warts achieved statistical significance. Even without a multiplicity adjustment for secondary endpoints to control the Type I error rate, time to complete clearance and proportion that relapse are not statistically different than those of the vehicle for each of the two active concentrations.

3.2 Evaluation of Safety

In the following sections, adverse events are reported only for those subjects that the sponsor lists as treatment emergent. Further, event rates are reported by subject (i.e. if an AE occurs multiple times for subject i, the tabulations only report that subject i experienced this AE and not the number of times it occurred). Safety tabulations combine results from Studies CT1017 and CT1018.

3.2.1 Treatment Emergent Adverse Event Rates

No large differences in the percent of AE's listed by System Organ Classification (SOC–MedDRA nomenclature) were seen between 10% ointment and 15% ointment (results shown in Table 15). However, both the active arms appear to have a higher percentage reporting the AE's listed by SOC than the vehicle arm.

Table 15: Treatment Emergent AEs listed by System Organ Class

System Organ Classification (MedDRA)	10% Oint $N = 392$	15% Oint $N = 388$	Vehicle $N = 206$
General disorders and administration site conditions	334 (85.2)	336 (86.6)	136 (66)
Skin and subcutaneous tissue disorders	90 (23)	97 (25)	19 (9.2)
Infections and infestations	77 (19.6)	69 (17.8)	27 (13.1)
Gastrointestinal disorders	25 (6.4)	20 (5.2)	9 (4.4)
Reproductive system and breast disorders	16 (4.1)	20 (5.2)	7 (3.4)
Blood and lymphatic system disorders	12 (3.1)	15 (3.9)	2 (1)
Nervous system disorders	29 (7.4)	15 (3.9)	13 (6.3)
Injury, poisoning and procedural complications	11 (2.8)	9 (2.3)	4 (1.9)
Musculoskeletal and connective tissue disorders	9 (2.3)	8 (2.1)	2 (1)
Respiratory, thoracic and mediastinal disorders	10 (2.6)	5 (1.3)	2 (1)

Results depict # of subjects and (percentages)

Source: Reviewer's Analysis

Table 16 lists the treatment emergent AE's which occurred in more than 1% of subjects in a given treatment arm. The AE names listed correspond to MedDRA preferred terms. For frequently occurring AE's (i.e. rates $\geq 2.0\%$), the 15% ointment has slightly higher rates than the 10% ointment and both of these have higher rates than the vehicle arm. Overall, the most

frequent AE's according to the MedDRA preferred terms are for local application site conditions as reported in Table 15.

Table 16: Treatment Emergent AEs

Preferred Term	10% Oint $N = 392$	15% Oint $N = 388$	Vehicle $N = 206$
Application Site Reactions			
Erythema	269 (68.6)	273 (70.4)	67 (32.5)
Itching	260 (66.3)	269 (69.3)	94 (45.6)
Burning	253 (64.5)	260 (67)	65 (31.6)
Pain	185 (47.2)	216 (55.7)	30 (14.6)
Erosion/ulceration	183 (46.7)	185 (47.7)	20 (9.7)
Edema	159 (40.6)	173 (44.6)	23 (11.2)
Induration	109 (27.8)	136 (35.1)	23 (11.2)
Vesicles	75 (19.1)	78 (20.1)	13 (6.3)
Desquamation	10 (2.6)	13 (3.4)	0 (0)
Lymphadenitis	8 (2)	10 (2.6)	2 (1)
Bleeding	4 (1)	6 (1.5)	0 (0)
Scaling	1 (0.3)	6 (1.5)	0 (0)
Yellow secretion	5 (1.3)	6 (1.5)	0 (0)
Systemic Reactions			
Headache	23 (5.9)	10 (2.6)	9 (4.4)
Gastritis	2 (0.5)	7 (1.8)	3 (1.5)
Pharyngitis	2 (0.5)	5 (1.3)	1 (0.5)

Results depict number of subjects and (percentages)

Source: Reviewer's Analysis

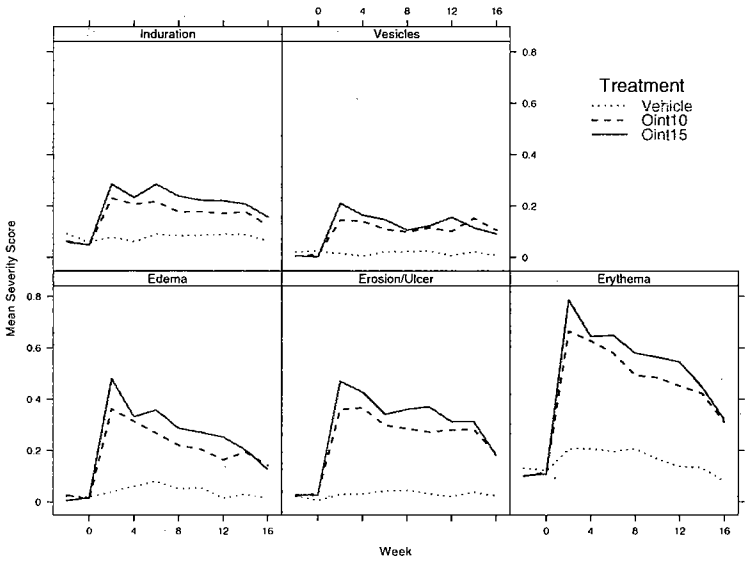
3.2.2 Local Safety Assessment

At the baseline visit and all subsequent treatment visits (weeks 0 through 16), the investigator rated five local skin reactions on a scale of 0 to 3; where 0 = 'none', 1 = 'mild', 2 = 'moderate', 3 = 'severe'. At each of the visits (excluding the follow-up visits for subjects that cleared), the mean score was calculated for each of the investigator rated local skin reactions.

Figure 7 depicts the mean score across time for each of the investigator rated local skin reactions. The figure shows that both the active ointments are more irritating than the vehicle. Specifically, it can be seen that the vehicle causes a slight, if any, increase in irritation while both actives increase the irritation very quickly with the irritation continuing while on treatment tending towards a resolution as time progresses. The graphic also depicts that the mean scores of irritation for 15% ointment tend to dominate the mean scores of the irritation scores for 10% ointment.

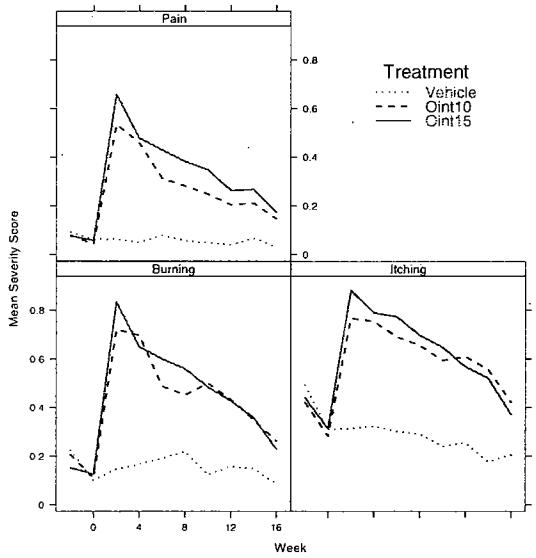
Similar to the investigator rated local skin reactions, subjects rated three skin symptoms based on a scale from 0 (none) to 3 (severe). Figure 8 on page 27 depicts the mean scores across visits for each subject rated skin symptom. Basically, similar to the investigator rated skin

Figure 7: Investigator Rated Local Safety



reactions, subject's reported skin symptoms follow the same trend with sharp increases early in the treatment and then gradually resolving. Again, it is also seen that the 15% ointment tends to dominate 10% ointment and both actives are more irritable than the vehicle.

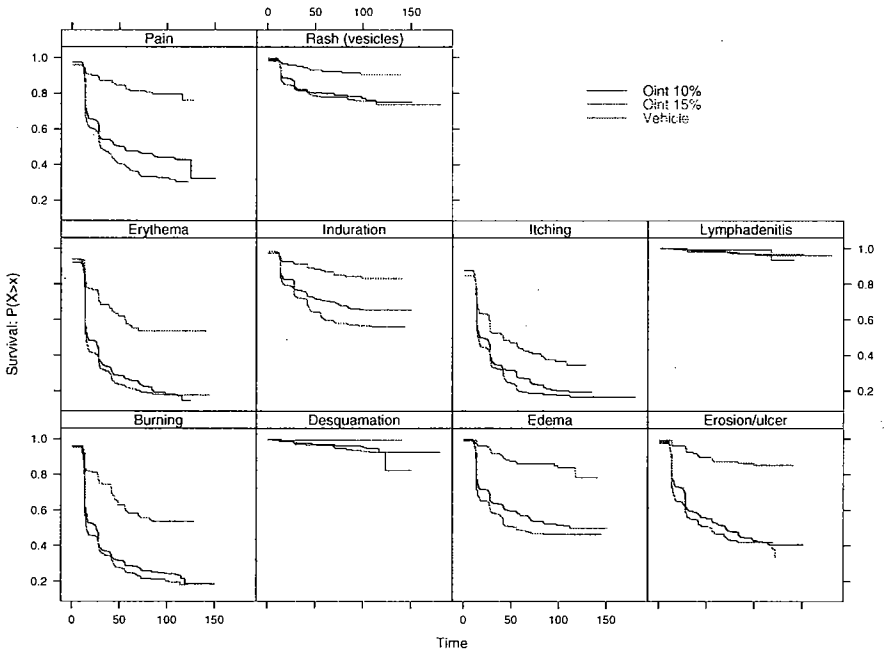
Figure 8: Subject Rated Local Safety



3.2.3 Time to First Event

Figure 9 depicts the time to *first* event for the 10 most frequent application site reactions reported in Table 16. While the AE rates for the 10% ointment and the 15% ointment were quite similar as reported in Table 16, Figure 9 shows that the 15% ointment treatment arm was more likely to experience the AE at an earlier time than the 10% ointment arm. Overall, it can be seen that the time to first event for the two active arms appears to be faster than the vehicle arm. With a small percentage of subjects experiencing desquamation and lymphadenitis, the curves do not appear to have large degrees of separation as the more frequently occurring AE's.

Figure 9: Time to First Event



The overall safety trend shows that both 10% ointment and 15% ointment have higher rates of treatment emergent AE's than the vehicle arm. The most frequent AE's tend to be local and application site orientated. While the AE rates of the two active ointments are similar, the time to first event shows that 15% ointment is more likely to experience the first event sooner than 10% ointment.

3.2.4 Serious Adverse Events

Serious adverse events reported in the trial are shown in Table 17. Both 10% ointment and 15% ointment had an equal number of serious AE's reported, however it appears that the majority of

the serious AE's reported in the 15% ointment arm are treatment related whereas serious AE's reported for the 10% ointment do not appear to be treatment related.

Table 17: Treatment Emergent Serious AEs*

ID	AE Name (MedDRA)	Day Onset	Drug Related	Action Taken	Outcome
Ointment 10%					
CT17 DEU-08 2372	Knee Injury	Day 93	No	Not applicable	Resolved w/o sequelae
CT18 COL-02 0163	Pregnancy	Day -13	No	Med. Discontinued	Resolved w/o sequelae
CT18 COL-02 0166	Pregnancy	Day 68	No	Med. Discontinued	Resolved w/o sequelae
CT18 COL-02 0328	Vulvovaginitis	Day 35	Possible	Med. Interrupted	Resolved w/o sequelae
CT18 COL-06 0320	Lower limb fracture	day 85	No	Med. Interrupted	Resolving
Ointment 15%					
CT17 DEU-01 2311	Application Site Reactions	Day 11	Probable	Med. Discontinued	Resolved
CT18 COL-02 0330	Pregnancy	Day 41	No	Med. Discontinued	Not Resolved
CT18 COL-02 0304	Pain	Day 12	Probable	Med. Interrupted	Resolved w/o sequelae
	Erythema	Day 12	Probable	Med. Interrupted	Resolved w/o sequelae
	Rash vesicular	Day 12	Probable	Med. Interrupted	Resolved w/o sequelae
CT18 USA-06 0508	Diabetic ketoacidosis	Day 70	No	Med. Interrupted	Resolved w/o sequelae

* Source: Study Reports CTD 5.3.5.1.2 page 205 and CTD 5.3.5.1.3 page 204.

4 Findings in Special/Subgroup Populations

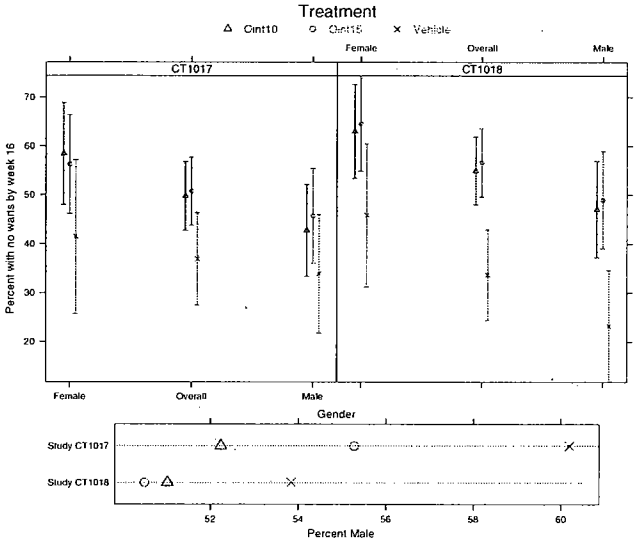
Results for the primary endpoint, percent with no warts by week 16 are provided according to gender, race, and age for Studies CT1017 and CT1018. Results are shown graphically with summary tables provided in the Appendix section A.4 on page 42. All graphical depictions include an unadjusted 95% confidence interval for the point estimates.

4.1 Gender, Race, and Age

Overall efficacy and efficacy by gender is shown in Figure 10. Baseline distributions of gender showed a slightly higher percentage of males for both studies (refer to Tables 22 and 23 in the Appendix). Figure 10 shows that females tended to have slightly increased efficacy rates over males for all treatment arms in each study. However, the observed difference in active and vehicle tended to be larger in male subjects in study CT1018.

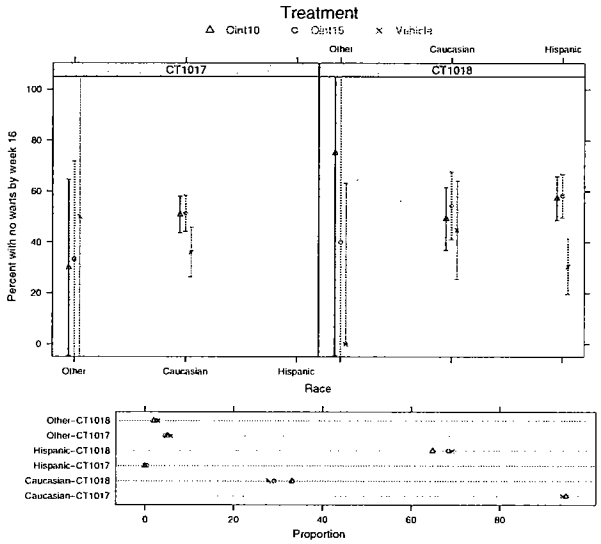
In the examination of efficacy by race all non-Caucasian and non-Hispanic subjects are lumped into the 'Other' category. Also, only one Hispanic subject was enrolled in CT1017 so results for Hispanics is not provided for Study CT1017. For both studies, the number of subjects in the 'Other' subgroup is small and hence the large unadjusted confidence bands. Efficacy in Caucasians is clear in Study CT1017, yet this same trend is not seen in Study CT1018. In Study

Figure 10: Efficacy Results by Gender



CT1018 the vehicle has a fairly high response rate in comparison to the vehicle response rate in Hispanic patients. This trend is further explored in Section 4.2.2.

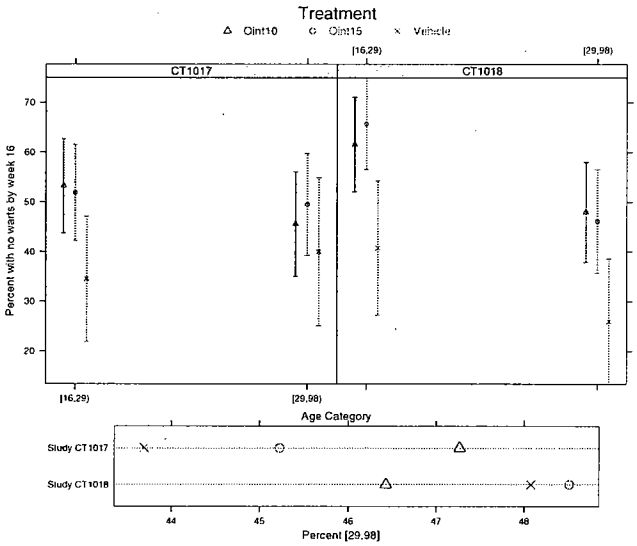
Figure 11: Efficacy Results by Race



Age was dichotomized based on the median age level for both studies. Efficacy results for dichotomized age are depicted in Figure 12. The trends in this study tend to show lower response rates in older patients in addition to slightly smaller deltas for the active and vehicle. Generally

speaking, results are consistent across studies.

Figure 12: Efficacy Results by Age

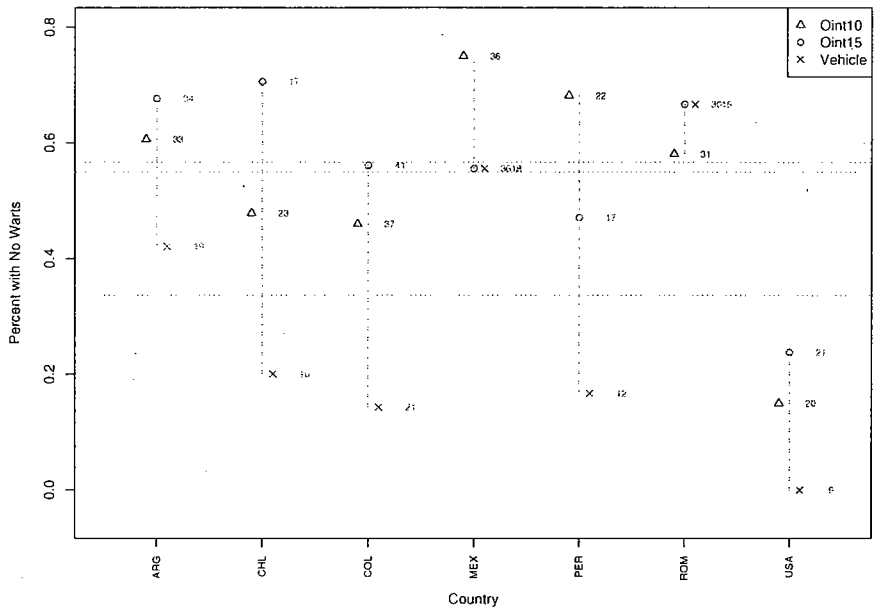


4.2 Other Special/Subgroup Populations

4.2.1 Efficacy by Country

While Study CT1017 was conducted entirely out of the US, roughly 10% of subjects enrolled in Study CT1018 came from the US. Figure 13 below shows the efficacy across countries. Overall study means are represented by horizontal dotted lines in the plot. Sample sizes within country and treatment arm are provided next to the point estimate. From this graphic it clearly shows the lower response rates in the US population versus other countries. In fact, the active arms do not even achieve the level of clearance as the overall estimate of clearance for the vehicle. Despite the smaller response rates, the same dose trend does appear in the US as other countries. The next section is an analysis to try and explain the reduced response in the US.

Figure 13: Efficacy by Country in Study CT1018



4.2.2 Examination of Lower Response in US

For study CT1018 a classification and regression tree (CART) was used to predict which subjects would achieve 100% clearance of all warts at the end of treatment. Note that the following sensitivity analysis excludes subjects with only a baseline visit. The goal was to identify what predictors are chosen based upon the CART algorithm. Specifically it is of interest to see if the US or non-US variable would be selected to predict 100% clearance.

The possible predictors (along with `variable name` used in the analysis) were:

- country (US or non-US) (`us`)
- treatment arm (`trt.x`);
- race (`race`);
- gender (`sex`);
- age (`age`);
- BMI (`BMI`);

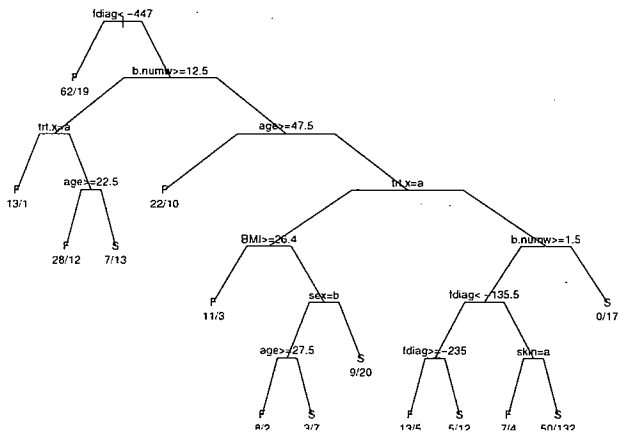
- wart located in moist area (yes for females and uncircumcised males, no for circumcised males) (`circum`);
- time from first diagnosis (in days) (`fdiag`);
- number of previous episodes (`nepi`);
- the number of warts present at the first treatment visit (`b.numw`); and
- area of baseline warts (`barea`).

To construct the tree, the `rpart` package in R (also available in Splus) was used. The default settings of the `rpart` function were used which makes the following assumptions:

- node splitting is based on the Gini rule;
- loss is 0/1; and
- prior probabilities are proportional to the observed data.

The resulting tree is shown in Figure 14. Based on the above model assumptions, the root node was if the patient had been first diagnosed more than 447 days ago. For subjects diagnosed more than 447 days ago, 62 out of 81 subjects (75.5%) did not achieve 100% clearance. Note this is for all treatment arms. A more detailed examine of these subjects will follow.

Figure 14: Tree using `rpart` in Study CT1018.

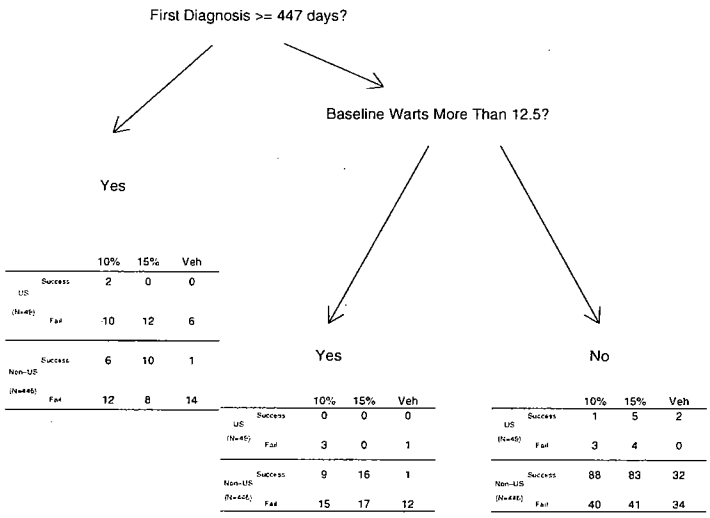


For subjects with a diagnosis less than 447 days, the next node is for baseline warts greater than 12.5. When baseline warts are greater than 12.5 and subjects receive placebo (`trt.x:a`),

only 1 out of 14 subjects (6.7%) achieved 100% clearance of all warts. Subjects can continually be broken down as shown on the tree. However, as the tree becomes more complex the robustness of the tree decreases. Based on cross-validation, the tree selected contains splits at `fdiag < -447`, `b.numw >=12.5`, and age `>= 47.5`, and this tree is a subset of the larger tree shown in Figure 14. Note that the tree selected does not include treatment arm as one of the predictors of 100% clearance.

As it appears that 100% clearance is dependent upon disease severity, defined in terms of days since first diagnosis and number of baseline warts, efficacy results for treatment arm and country (US versus non-US) are provided in Figure 15. Looking at the table located on the far left of the figure, it can be seen that 30/49 (61.2%) of US subjects were first diagnosed more than 447 days ago. And of the 30 enrolled with this condition only 2 (all receiving 10% ointment) achieved 100% clearance. As discussed previously, the overall success rate for subject diagnosed more than 447 days ago is 24.5% in Study CT1018. Similarly in Study CT1017, 21 out of 79 subjects (26.6%) who were diagnosed more than 447 days ago achieved 100% clearance regardless of treatment received. A breakdown in success by treatments is provided in Table 18 for the two pivotal studies on the following page.

Figure 15: Success Rates by Disease Severity (CT1018)



For subjects diagnosed less than 447 days ago with more than 12.5 warts at baseline, 35.0% achieved 100% clearance (refer to the middle table in Figure 15). In this category there appears to be a dose response with the 15% ointment achieving the highest efficacy. Note that with such a limited number of US subjects enrolled in this subgroup, it is hard to reach any conclusions on efficacy in US alone.

Table 18: Efficacy results for subjects diagnosed more than 447 days from first treatment with Polyphenon E® .

		10% ointment	15% ointment	Vehicle
CT1017	N	28	29	22
	Success	32.1%	17.2%	31.8%
CT1018	N	22	20	21
	Success	36.4%	50.0%	5.0%

For subjects diagnosed less than 447 days ago and who had less than 12.5 warts at baseline the efficacy is higher than in the previous subgroups. Table 19 below shows efficacy results from both studies CT1017 and CT1018 when subjects met the above criteria. Study CT1018 failed to show a clear dose response in the 10% ointment and 15% ointment treatment groups, but Study CT1017 did show a dose response. Again, few subjects in the US were enrolled that fall into this subgroup to reach any conclusions about efficacy in the US alone.

Table 19: Efficacy results for subjects diagnosed less than 447 days with fewer than 12.5 warts at baseline.

		10% ointment	15% ointment	Vehicle
CT1017	N	133	133	66
	Success	55.6%	61.7%	42.4%
CT1018	N	132	132	68
	Success	67.4%	65.9%	50.0%

Based upon the model selected from `rpart` and using supporting evidence from Study CT1017, the reason for the decreased efficacy in US sites appears to be influenced by the fact that US sites enrolled subjects with more severe disease at baseline. More than 60% of subjects enrolled in the US had been first diagnosed more than 447 days prior to treatment which was shown to be the primary predictor using CART. Constructing a tree using Study CT1017 also resulted in a root node based on the time from first diagnosis. In this tree the split of the root node occurred at `fdiag < - 175.5`. While the location of the split differs in the two trees, it shows the importance of how the time from first diagnosis is influential in determining treatment success. Specifically it emphasizes that those who receive treatment many days after first diagnosis have a lower chance of reaching treatment success. While success is much lower for these subjects, the active treatment arms do seem to show some effect over placebo.

In the Agency's 74 day letter to the sponsor, the Agency requested the sponsor to provide supportive information regarding the generalizability of the clinical data to the United States population. The sponsor's response included two written statements from dermatologists that stated that efficacy was not likely due to race or ethnic factor. Rather, the two dermatologists stated that females and uncircumcised males tend to respond better to treatment than circumcised males. The variable `circum` was used in the analysis above and neither in Study CT1017 nor CT1018 was this variable selected as a predictor of treatment success. However, of the 36 males recruited in the US, 33 were circumcised.

In conclusion, it appears that efficacy in the US was impacted by possibly two factors: the number of days from diagnosis and whether males were circumcised or not. The first factor, the number of days from first diagnosis, was found using CART in two separate studies and the later factor was based upon prior clinical information. Based upon the collective evidence it appears to be more difficult to achieve 100% clearance of all warts in circumcised males and in subjects whose first diagnosis occurred many days from first treatment with Polyphenon E® (possibly more than 6 months ago). Of the 49 US subjects enrolled with at least one post-baseline visit, Table 20 below depicts the number of US subjects enrolled for each of the two factors: days from first diagnosis (more or less than 180 days) and if the subject was circumcised. This table depicts that US sites tended to enroll subjects who were less likely to achieve success based upon the two factors. Consequently, if the small sample of subjects enrolled in Study CT1018 is representative of the US population as a whole, it is unclear if the efficacy results from CT1017 and CT1018 would generalize to the US as the US population may consist of subjects that are less likely to achieve 100% clearance of all warts.

Table 20: Baseline Enrollment for US subjects

Circumcised	First Diag. > 180 days	First Diag. ≤ 180 days
Yes	22	11
No	8	8

Source: Reviewer's analysis

5 Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

The initial submission of the efficacy analyzable data set contained only imputed data values for the endpoints of interest. In the 74 day letter (December 8, 2005), the Agency asked the

sponsor to also submit the raw data. The sponsor's response was received on January 6, 2006, but the raw data files did not contain data definition files and it was unclear how to use such data. On February 1, 2006 the Agency then sent a request for information for the sponsor to submit the data in specific format which should contain data for each subject at every visit; if the subject missed the visit the data should be stored as NA. On March 6, 2006 the sponsor resubmitted the data which did not meet the Division's request (i.e. not all subjects contained all visits as was requested).

On May 23, 2006 the Agency had a teleconference with the sponsor again requesting an efficacy analyzable data set which could be used to perform analyses on the protocol defined primary population (ITT-LOCF), and the Agency also requested the sponsor submit data to assess relapse rates. On May 31, 2006 the sponsor submitted SAS code which used the raw data to create modified data sets and the analysis was based on the modified data sets. However, the modified data sets used by the sponsor are the same as those previously submitted to the NDA which the Agency conveyed were deficient for the assessment of the primary analysis on the protocol defined primary population. Consequently the reviewer filled in the missing structures in the sponsor's data set to more accurately capture the data contained in the CRF to allow for the assessment of the protocol defined primary analysis population. However, it should be noted that such an attempt at creating a working data set for efficacy evaluation may not fully address all the missing data issues as the reviewer does not have access or resources to ensure the quality of the data.

In each of the sponsor's pivotal Phase 3 trials, both 10% ointment and 15% ointment were superior to the vehicle based upon protocol defined analysis of the primary endpoint. The results for Study CT1018 were strong and robust, whereas some sensitivity analyses (not protocol defined) of Study CT1017 data failed to demonstrate that 10% ointment and 15% ointment were superior to the vehicle. In both studies 15% ointment had a slightly larger percentage of subjects with complete clearance of all their warts than 10% ointment, but this difference was not statistically significant. Results shown in Table 21.

Only 10% of subjects were enrolled in Study CT1018 (no US sites were used in Study CT1017) and results by country revealed the US to have lower percentages of subjects with resolution of all warts (refer to Figure 15 on page 34). As an attempt to explain this phenomenon a Classification and Regression Tree (CART) was used to see what factors would predict subject success. This analysis found that subjects who are diagnosed further in the past (CART root node split of 447 days) tend to not respond to treatment as well as subjects who are recently diagnosed. Looking at the subjects enrolled in the US revealed that a large portion of US subjects (62.5%) were diagnosed more than 447 days (CART defined split). Further, the sponsor provided documentation that treatment in circumcised males is not as responsive as in uncircumcised males for which the US has a higher proportion of circumcised males.

According to the system organ classification (MedDRA nomenclature) 85.2% and 86.6% of

Table 21: Primary Endpoint Efficacy Results (ITT-LOCF)

	Study CT1017			Study CT1018		
	Oint10	Oint15	Vehicle	Oint10	Oint15	Vehicle
Fail	100 (50.3)	99 (49.3)	65 (63.1)	91 (45)	85 (43.4)	69 (66.3)
Success	99 (49.7)	102 (50.7)	38 (36.9)	111 (55)	111 (56.6)	35 (33.7)
p-value[1]	0.0384	0.0284	-	<0.001	<0.001	-
p-value[2]	0.0280	0.0143	-	<0.001	<0.001	-

[1] Source: Reviewer's Analysis using Fisher's exact test.

[2] Source: Table 11.11 in CTD 5.3.5.1.2 (CT1017) and Table 11.9 in CTD 5.3.5.1.3 (CT1018) using Fisher's exact test.

subjects experienced general disorders and application site reactions for 10% ointment and 15% ointment, respectively, which was less than the vehicle arm (66.0%). The second most frequently recorded AE's occurred for skin and subcutaneous tissue disorders at rates of 23%, 25%, and 9.2% for 10% ointment, 15% ointment, and vehicle, respectively. One subject receiving 10% ointment and two subjects receiving 15% ointment reported serious AE's that were recorded as having a probable casual relationship to treatment.

5.2 Conclusions and Recommendations

The efficacy analysis based upon protocol defined methods showed both 10% ointment and 15% ointment to be superior to the vehicle in the treatment of genital warts. However, the difference in the percentage of subjects with no warts between the two active treatment arms was small in both studies. Further, examinations of the safety data reveal that the higher dose has the potential for additional safety concerns than the lower dose such as faster onset of application site reactions and more treatment related serious adverse events.

The following comments correspond to the sponsor's proposed label.

- The sponsor's proposed labeling includes results based on study CT1005 where subjects were treated for 12 weeks and different formulations. At the June 11, 2001 meeting the Agency advised the sponsor that Study CT1005 will not be considered a pivotal trial. Consequently, in terms of efficacy claims data from CT1005 should not be included in the label.

- The sponsor includes a table with point estimates only that pools results for those with complete clearance and $\geq$ 75% clearance of all warts from all three studies using the population of observed cases. The data from CT1005 should not be used in this summary

per the previous comment. As both studies established the significance of active versus vehicle for both endpoints (complete clearance and $\geq$ 75% clearance) and no p-values are reported from a statistical perspective inclusion of summaries of the percent who achieved either criteria from Studies CT1017 and CT1018 is acceptable.

- The sponsor's label also provides the following statement, "Examination of age, race, gender subgroups suggested a larger absolute treatment effect in women and a more pronounced treatment effect for patients below the median age of 28 years." As no formal testing of the subgroups were performed, the following is suggested for statements about efficacy in subgroups, "Examination of age and gender subgroups showed a trend with higher efficacy in females subjects and subjects below 28 years of age,"

- The sponsor also includes a section about the recurrence of genital warts. In the sponsor's proposed label, they estimate the relapse rate to be 6.7% for 15% ointment. The Division has the following comments pertaining to labeling the relapse rate.

  - The sponsor's data does not appear to permit estimation of relapse rate due to the difficult recording of 'missing data' as discussed in Sections 2.2 and 3.1.8.3.
  - The comparisons of each active concentration to vehicle did not reach statistical significance even without a multiplicity adjustment.
  - The sponsor's reported relapse rate of 6.7% for 15% ointment appears to assume all 'missing data' do not relapse and this likely underestimates the relapse rate.

With a number of inconsistencies in the data submitted to the Agency, this calls into question the quality of the data. While the Agency was not able to find any definitive misrepresentations in the data, this does not preclude the chance for such occurrences. Based upon the data used in the above review; from a statistical perspective, the collective evidence establishes that both 10% ointment and 15% ointment are superior to the vehicle for the primary endpoint, complete clearance of all warts.

APPENDIX

A.1 Baseline Descriptive Statistics

The following tables contain comparisons across treatment arms for each of the baseline characteristics. Overall, both studies appear to have balance in the baseline characteristics across treatment arms.

Table 22: Baseline Descriptive Statistics (CT1017)

	N	Oint10 $N = 199$	Oint15 $N = 201$	Vehicle $N = 103$	Test Statistic
Sex : male	503	55% $\frac{110}{199}$	52% $\frac{105}{201}$	60% $\frac{62}{103}$	$\chi^2_2 = 1.8$, $P = 0.417$[1]
Race : African	503	3% $\frac{6}{199}$	3% $\frac{6}{201}$	4% $\frac{4}{103}$	$\chi^2_8 = 2.3$, $P = 0.97$[1]
Asian		1% $\frac{1}{199}$	0% $\frac{1}{201}$	1% $\frac{1}{103}$	
Caucasian		95% $\frac{189}{199}$	95% $\frac{191}{201}$	94% $\frac{97}{103}$	
Hispanic		0% $\frac{0}{199}$	0% $\frac{1}{201}$	0% $\frac{0}{103}$	
Other		2% $\frac{3}{199}$	1% $\frac{2}{201}$	1% $\frac{1}{103}$	
Circumcised : No	277	88% $\frac{97}{110}$	93% $\frac{98}{105}$	90% $\frac{56}{62}$	$\chi^2_2 = 1.7$, $P = 0.431$[1]
Age in Years	503	23 27 36	22 28 35	22 26 36	$F_{2,500} = 0.24$, $P = 0.789$[2]
Body Mass Index	502	21 23 25	21 23 25	21 23 25	$F_{2,499} = 0.32$, $P = 0.727$[2]

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. N is the number of non–missing values. Tests used: [1]Pearson test; [2]Kruskal-Wallis test

Table 23: Baseline Descriptive Statistics (CT1018)

	N	Oint10 $N = 202$	Oint15 $N = 196$	Vehicle $N = 104$	Test Statistic
Sex : male	502	50% $\frac{102}{202}$	51% $\frac{100}{196}$	54% $\frac{56}{104}$	$\chi^2_2 = 0.33$, $P = 0.85$[1]
Race : African	502	1% $\frac{3}{202}$	3% $\frac{5}{196}$	2% $\frac{2}{104}$	$\chi^2_8 = 7$, $P = 0.538$[1]
Asian		0% $\frac{1}{202}$	0% $\frac{0}{196}$	0% $\frac{0}{104}$	
Caucasian		33% $\frac{67}{202}$	29% $\frac{57}{196}$	28% $\frac{29}{104}$	
Hispanic		65% $\frac{131}{202}$	68% $\frac{134}{196}$	69% $\frac{72}{104}$	
Other		0% $\frac{0}{202}$	0% $\frac{0}{196}$	1% $\frac{1}{104}$	
Circumcised : No	258	83% $\frac{85}{102}$	76% $\frac{76}{100}$	77% $\frac{43}{56}$	$\chi^2_2 = 1.9$, $P = 0.393$[1]
Age in Years	502	23 28 36	23 27 35	23 28 38	$F_{2,499} = 0.36$, $P = 0.7$[2]
Body Mass Index	502	21 24 26	21 24 27	22 24 27	$F_{2,499} = 0.69$, $P = 0.504$[2]

a b c represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. N is the number of non–missing values. Tests used: [1]Pearson test; [2]Kruskal-Wallis test

A.2 Treatment Duration

The following provides summary statistics of length of treatment for each study. This data is meant to augment distribution depictions provided in Figure 2 on page 14.

Table 24: Percentiles (5%, 50%, 95%) of Treatment Duration by Treatment Arm

	Oint10				Oint15				Vehicle			
	N	5%	50%	95%	N	5%	50%	95%	N	5%	50%	95%
Protocol												
CT1017	199	40.7	112.0	121.1	201	13.0	111	124.0	103	28.0	112	121.9
CT1018	202	22.1	98.5	120.0	196	16.8	90	119.2	104	41.2	112	121.0
Overall												
	401	27.0	111.0	121.0	397	15.0	105	120.0	207	28.3	112	121.7

Source: Reviewer's Analysis.

A.3 ITT Eligible Subjects Excluded from Sponsor's Report

The following tables provide ID numbers for subjects which were not included in the sponsor's primary efficacy analysis for the ITT-LOCF population. These subjects were included in the reviewer's primary efficacy analysis and based on LOCF each was considered a treatment failure.

Table 25: Subjects Excluded from CT1017 Study Reports

ID	Site	Treatment	ID	Site	Treatment
2311	DEU-01	15% Oint	2361	ROM-01	10% Oint
2192	DEU-08	15% Oint	2245	RUS-02	10% Oint
2193	DEU-08	10% Oint	2260	RUS-08	15% Oint
2373	DEU-08	15% Oint	2518	ZAF-06	15% Oint
2006	NOR-02	10% Oint	2672	ZAF-06	Vehicle
2009	NOR-02	15% Oint	2673	ZAF-06	15% Oint

Table 26: Subjects Excluded from CT1018 Study Reports

ID	Site	Treatment	ID	Site	Treatment
0253	CHL-03	10% Oint	0229	ARG-05	15% Oint
0272	CHL-03	10% Oint	0775	ROM-01	10% Oint
1023	MEX-10	15% Oint	0987	USA-10	10% Oint
1046	MEX-13	10% Oint	-	-	-

A.4 Efficacy Results by Subgroup Tables

The following tables provide further description to the figures displayed in Section 4.1 on page 29.

Table 27: Efficacy Results by Gender

	Oint10	Oint15	Vehicle
CT1017			
Female	58.4% $\frac{52}{89}$	56.2% $\frac{54}{96}$	41.5% $\frac{17}{41}$
Male	42.7% $\frac{47}{110}$	45.7% $\frac{48}{105}$	33.9% $\frac{21}{62}$
CT1018			
Female	63% $\frac{63}{100}$	64.6% $\frac{62}{96}$	45.8% $\frac{22}{48}$
Male	47.1% $\frac{48}{102}$	49% $\frac{49}{100}$	23.2% $\frac{13}{56}$

Table 28: Efficacy Results by Race

	Oint10	Oint15	Vehicle
CT1017			
Other	30% $\frac{3}{10}$	33.3% $\frac{3}{9}$	50% $\frac{3}{6}$
Caucasian	50.8% $\frac{96}{189}$	51.3% $\frac{98}{191}$	36.1% $\frac{35}{97}$
Hispanic	0% $\frac{0}{0}$	100% $\frac{1}{1}$	0% $\frac{0}{0}$
CT1018			
Other	75% $\frac{3}{4}$	40% $\frac{2}{5}$	0% $\frac{0}{3}$
Caucasian	49.3% $\frac{33}{67}$	54.4% $\frac{31}{57}$	44.8% $\frac{13}{29}$
Hispanic	57.3% $\frac{75}{131}$	58.2% $\frac{78}{134}$	30.6% $\frac{22}{72}$

Table 29: Efficacy Results by Age

	Oint10	Oint15	Vehicle
CT1017			
[16,29)	53.2% $\frac{58}{109}$	51.9% $\frac{55}{106}$	34.5% $\frac{20}{58}$
[29,98]	45.6% $\frac{41}{90}$	49.5% $\frac{47}{95}$	40% $\frac{18}{45}$
CT1018			
[16,29)	63% $\frac{63}{100}$	64.6% $\frac{62}{96}$	45.8% $\frac{22}{48}$
[29,98]	47.1% $\frac{48}{102}$	49% $\frac{49}{100}$	23.2% $\frac{13}{56}$

Signatures/Distribution List

Primary Statistical Reviewer: Mat Soukup, Ph.D.
Date: July 25, 2006

Statistical Team Leader: Mohamed Alosh, Ph.D.

cc:
Archival NDA
DDDP/Walker
DDDP/Lindstrom
DDDP/Papadapolous
DDDP/Wright
OBIO/O'Neill
OBIO/Patrician
DBIII/Wilson
DBIII/Alosh
DBIII/Soukup

July 25, 2006

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

/s/

Matt Soukup
8/15/2006 01:59:04 PM
BIOMETRICS

Mohamed Alosh
8/16/2006 10:08:58 AM
BIOMETRICS
Concur with review