

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

21-794

STATISTICAL REVIEW(S)



U.S. 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: 21-794 / N-000
Drug Name: Aczone (dapsons) topical gel, 5%
Indication(s): Acne
Applicant: Atrix Laboratories/QLT
Dates: Submitted: 9/7/04
PDUFA: 7/7/05

Review Priority: Standard review

Biometrics Division: Division of Biometrics III (HFD-725)
Statistics Reviewer: Kathleen Fritsch, Ph.D.
Concurring Reviewer: Mohamed Alosch, Ph.D.

Medical Division: Division of Dermatologic and Dental Drug Products
(HFD-540)
Clinical Team: Brenda Vaughan, M.D.
Project Manager: Frank Cross

Keywords: effect size

Table of Contents

1	EXECUTIVE SUMMARY	4
1.1	Conclusions and Recommendations.....	4
1.2	Brief Overview of Clinical Studies	4
1.3	Statistical Issues and Findings	5
2	INTRODUCTION	6
2.1	Overview	6
2.2	Data Sources	6
3	STATISTICAL EVALUATION	7
3.1	Evaluation of Efficacy.....	7
3.1.1	Study Design – Studies 203 and 204	7
3.1.2	Subject Disposition	8
3.1.3	Baseline Data	9
3.1.4	Primary Efficacy Endpoints.....	10
3.1.4.1	Analysis Populations.....	10
3.1.4.2	ITT Analyses.....	11
3.1.4.3	MITT Analyses	12
3.1.4.4	Per Protocol Analyses.....	13
3.1.5	Secondary Efficacy Endpoints.....	13
3.1.6	Statistical Issues	14
3.1.6.1	Clear/Minimal on the GAAS with at least Two Grades Reduction	14
3.1.6.2	Efficacy Results by Center.....	15
3.1.6.3	Pooling of Centers	16
3.1.6.4	Lesion Counts and the GAAS.....	17
3.1.6.5	Nodules.....	18
3.1.7	Study 0004	18
3.2	Evaluation of Safety.....	19
3.2.1	Studies 0203 and 0204.....	19
3.2.1.1	Extent of Exposure.....	19
3.2.1.2	Adverse Events	20
3.2.1.3	G-6-PD Deficiency	20
3.2.2	Study 0114	20
3.2.2.1	Extent of Exposure.....	21
3.2.2.2	Adverse Events	22
3.2.2.3	Dapsone Concentrations	22

4	FINDINGS IN SPECIAL/SUBGROUP POPULATIONS.....	24
4.1	Gender, Race, and Age.....	24
4.2	Other Special/Subgroup Populations.....	27
5	SUMMARY AND CONCLUSIONS.....	27
5.1	Statistical Issues and Collective Evidence.....	27
5.2	Conclusions and Recommendations.....	29
APPENDIX	30
	Additional Figures.....	30
SIGNATURES/DISTRIBUTION LIST	31

1 Executive Summary

1.1 Conclusions and Recommendations

Dapsone is statistically superior to its vehicle in two studies for all primary endpoints (percent change in inflammatory, non-inflammatory, and total lesions, and success on the Global Acne Assessment Scale or GAAS). The studies were adequately powered to detect relatively small treatment effects. In these studies the efficacy benefit of dapsone over vehicle is relatively modest, with dapsone efficacy results about 4 to 9% better than vehicle results. The two Phase 3 studies had overall results that were very similar to each other, but within studies there was a lot of center to center variation. Although it was not a pre-specified hypothesis, there is some evidence from the subgroup analyses that female subjects had better overall results than males, and that adult subjects had better overall results than adolescent subjects. The treatment differences between dapsone and vehicle were roughly the same, however, in the different subgroups. One design flaw with the studies was there was no baseline requirement for the GAAS and a number of subjects were enrolled with a GAAS score of 1 (minimal) at baseline. Consequently, those subjects would not have to improve any from baseline to be considered a success on the GAAS at the end of the study. This necessitated a post hoc MITT population to exclude those subjects. However, the ITT, MITT, and an additional post hoc analysis that required at least two grades reduction in addition to achieving none or minimal were all statistically significant.

Adverse event rates were similar between the dapsone and vehicle arms. Subjects were screened for G-6-PD deficiency and a total of 24 dapsone and 25 vehicle subjects in Studies 0203, 0204, and 0114 were G-6-PD deficient. In the 12-month open-label safety study (Study 0114), subjects' dapsone blood levels were assessed at each visit. Of the 5 G-6-PD deficient subjects in Study 0114, two of the subjects had higher dapsone levels relative to the other subjects (85th to 99th percentile of measurements), but these two subjects only had dapsone concentrations through month 3 rather than through month 12.

1.2 Brief Overview of Clinical Studies

The sponsor conducted two pivotal Phase 3 studies (Studies 0203 and 0204), a supportive Phase 3 study (Study 0004), and a long-term, open-label safety study (Study 0114). The pivotal Phase 3 studies are the primary focus of this review. The sponsor is seeking an indication for the topical treatment of acne vulgaris. Studies 0203, 0204, and 0004 were 12-week, vehicle-controlled studies that enrolled 1485, 1525, and 496 subjects with acne, respectively. In each study subjects were evaluated for inflammatory, non-inflammatory, and total lesions and on a global acne assessment scale. Study 0004 used slightly different endpoints than are usually recommended by the Division (final acne counts instead of change or percent change from baseline, and a global evaluation scale that included half-steps). Study 0004 is not independent of Studies 0203 and 0204 as many investigators from Study 0004 were also used in Study 0204. Study 0114 was an open-label single-arm 12-month study of the safety of dapsone in 506 subjects with acne. Adverse events, lab results, and dapsone blood concentrations were monitored throughout the study.

1.3 Statistical Issues and Findings

In both of the Phase 3 studies, dapson topical gel was statistically superior to its vehicle in terms of the percent reduction in inflammatory, non-inflammatory, and total lesions and in terms of success on the GAAS, as specified in the protocol. The Phase 3 studies were very large (1485 and 1525 subjects) and were adequately powered to detect relatively small differences between treatments. In Study 0203, the average net benefit of dapson over vehicle in terms of reductions in lesions was of 4.2% or 1.4 more inflammatory lesions and 7.2% or 4.3 more non-inflammatory lesions. In Study 0204, the average net benefit of dapson over vehicle was 7.3% or 2.3 more inflammatory lesions and 8.5% or 4.2 more non-inflammatory lesions. These results are summarized in Table 1.

Table 1 – Percent Reduction and Absolute Reduction in Lesions from Baseline (ITT)

	Study 0203			Study 0204		
	Dapson N=745	Vehicle N=740	p-value	Dapson N=761	Vehicle N=764	p-value
Infl. (% Red)	45.9%	41.7%	0.0302	47.6%	40.3%	<0.0001
(Abs Red)	13.7	12.3	0.0265	14.3	12.0	0.0001
Non-Inf. (% Red)	31.1%	23.9%	0.0022	29.6%	21.1%	<0.0001
(Abs Red)	16.4	12.1	0.0001	13.9	9.7	0.0001
Total (% Red)	38.3%	32.0%	0.0004	37.4%	29.3%	<0.0001
(Abs Red)	30.4	24.6	0.0001	28.4	21.7	<0.0001

Table presents least squares means adjusted for baseline lesion count and center. P-values are based on an ANCOVA model with terms for treatment, center, and baseline lesion count.

Interpretation of the Global Acne Assessment Score (GAAS) is complicated by the fact that 6.7% of subjects in Study 0203 and 3.8% of subjects in Study 0204 were enrolled with baseline GAAS scores of 1 (minimal). Success at Week 12 was defined as a GAAS score of 0 (none) or 1 (minimal), thus these subjects were enrolled at a severity considered a success post-treatment. Study entry criteria were only defined in terms of the number of lesions, and a minimum GAAS score was not required at baseline. The sponsor defined an ad hoc MITT population that excluded subjects with baseline GAAS of 1.

The interpretation of a “few” or “no” lesions also seemed to vary from investigator to investigator. The success categories were defined as 0 - None: no evidence of facial acne vulgaris, and 1 - Minimal: a few non-inflammatory lesions (comedones) are present; a few inflammatory lesions (papules/pustules) may be present. However, subjects scored as 0 (none) had as many as 9 inflammatory lesions or 51 non-inflammatory lesions and subjects scored as 1 (minimal) had as many as 49 inflammatory lesions or 115 non-inflammatory lesions.

All analyses based on the GAAS, both those specified in the protocol (success = 0,1 for ITT, PP) and post hoc (success=0,1 for MITT, success= 0,1 + 2 grades reduction for MITT), were all statistically significant. The success rates for the MITT analysis

(success=0,1) are: 41.6% (dapson) versus 32.5% (vehicle) for Study 0203, and 34.7% (dapson) versus 27.9% (vehicle) for Study 0204.

Efficacy rates varied greatly among centers. A number of centers had higher efficacy with the vehicle than with dapson. Because each study had more than 50 centers and the fact that the treatment difference between dapson and vehicle was relatively modest—in the range of 4-9%—it does not seem too surprising that a number of centers favored vehicle over dapson. In addition, the presence of one or two extreme outliers within a center could greatly influence the center mean. Examination of subgroups indicates that gender and age have an impact on efficacy results with females and adults generally having better outcomes than males and adolescents, though treatment differences across subgroups did not vary greatly (i.e. females and adults had better results on both the dapson and vehicle arms). Some of this variability may be due to the fact that males and adolescents generally had higher baseline lesion counts than females and adults, though the effect persists even after adjusting for baseline. Some of the center variability may be confounded with gender and age variability as different centers enrolled different demographics (i.e. some centers enrolled more adult females while other centers enrolled more adolescent males).

2 Introduction

2.1 Overview

Dapson topical gel 5% has been studied in the treatment of acne vulgaris. The sponsor has conducted three Phase 3 studies (two pivotal and one supportive) and a 12-month open label safety study. The two pivotal Phase 3 studies were the subject of a Special Protocol Assessment. All studies were conducted in the United States and Canada. Since Study 0004 used endpoints which were slightly different than those generally used by the Division, and most of the investigators in Study 0004 were also used in Study 0204, Study 0004 is reviewed in less detail than Studies 0203 and 0204. The relevant studies in the clinical program are presented in Table 2.

Table 2 – Dapson Clinical Program

Study	Description	Duration	# Subjects
0004	Safety and Efficacy (Supportive)	12 weeks	330 dapson 166 vehicle
0203	Safety and Efficacy (Pivotal)	12 weeks	745 dapson 740 vehicle
0204	Safety and Efficacy (Pivotal)	12 weeks	761 dapson 764 vehicle
0114	Long term Safety	12 months	506 dapson

2.2 Data Sources

The materials reviewed include the study reports and clinical summaries. The final study reports for Studies 0004, 0203, and 0204 were submitted with the initial NDA submission on September 7, 2004. The amended study report for Study 0114, which included the

dapsone concentrations from the blood samples that were not analyzed in the original final report, was submitted on February 22, 2005. The datasets used in this review are archived at '\\cdesub1\n21794\N_000\2004-08-31\SAS Transport Datasets'. The datasets with the updated dapsone concentrations for Study 0114 are archived at '\\cdesub1\n21794\2005-02-22A\DAP0114 datasets and variables table'.

3 Statistical Evaluation

3.1 Evaluation of Efficacy

The sponsor conducted three Phase 3 studies evaluating dapson topical gel in the treatment of acne. The two larger studies, DAP0203 and DAP0204, follow identical protocols and were the subject of a Special Protocol Assessment. The third study (DAP0004) was conducted before the two pivotal trials and has slightly different endpoints than those usually considered by the Division (total lesions rather than change or percent change, and an investigator's global that used intermediate half scores). In addition, many of the investigators from Study 0004 were used again in Study 0204. Because of the lack of independence of the investigators and the choice of endpoints, Study 0004 will be considered supportive and reviewed in less detail than Studies 0203 and 0204.

3.1.1 Study Design – Studies 203 and 204

Studies 0203 and 0204 were conducted under identical protocols. Protocol 0204 was submitted as a Special Protocol Assessment in June 2002. The studies are randomized, double-blind, vehicle-controlled studies evaluating the safety and efficacy of dapson topical gel 5% in the treatment of acne vulgaris. Subjects were 12 years old or older and had 20 to 50 inflammatory and 20 to 100 non-inflammatory lesions at baseline. Subjects applied test medication twice daily for 12 weeks. Subjects were evaluated at baseline, Weeks 2, 4, 6, 8, and 12. The primary efficacy timepoint was Week 12.

The primary efficacy endpoints were the percent reduction in inflammatory, non-inflammatory, and total lesions (2 out of 3 must be significant) and success on the Global Acne Assessment Scale (GAAS), defined as a score of 0 or 1 (clear or minimal). The following is the full GAAS:

- 0 - None: no evidence of facial acne vulgaris
- 1 - Minimal: a few non-inflammatory lesions (comedones) are present; a few inflammatory lesions (papules/pustules) may be present
- 2 - Mild: several to many non-inflammatory lesions (comedones) are present; a few inflammatory lesions (papules/pustules) are present
- 3 - Moderate: many non-inflammatory (comedones) and inflammatory lesions (papules/pustules) are present; no nodulo-cystic lesions are allowed
- 4 - Severe: significant degree of inflammatory disease; papules/pustules are a predominant feature; a few nodulo-cystic lesions may be present; comedones may be present

The secondary efficacy endpoints were the mean lesion counts at Week 12, and the mean reductions from baseline. The GAAS was analyzed with a Cochran-Mantel-Haenszel test stratified on center. Percent reduction, reduction, and total lesion counts were analyzed with an ANCOVA using the corresponding baseline lesion count as a covariate along with factors for treatment and center, and treatment by center interaction.

The ITT population was defined as all subjects dispensed test article. The inclusion criteria only required that subjects have lesion counts within a certain range and did not require a minimum value for the GAAS at baseline. Consequently, some subjects were enrolled with a GAAS of 1 putting them into the success criteria for the GAAS at the start of the study. After discovering this problem upon review of the data, the sponsor defined an ad hoc MITT population defined as subjects with a minimum score of 2 on the GAAS at baseline.

3.1.2 Subject Disposition

Study 0203 enrolled 1485 subjects, 745 to dapsone and 740 to vehicle, at 51 centers. Study 0204 enrolled 1525 subjects, 761 to dapsone and 764 to vehicle, at 53 centers. Approximately 15% of subjects in Study 0203 and 18% of subjects in Study 0204 discontinued the study early. The reasons for study discontinuation are listed in Table 3. The most common reason for study discontinuation was lost to follow-up. About 8% of subjects in Study 0203 and 10% of subjects in Study 0204 were lost to follow-up. The second most common reason was 'patient voluntarily withdrew' at around 5% in both studies. All other reasons, including adverse events and lack of efficacy, were cited by less than 1% of subjects each. The discontinuation rates are similar in the dapsone and vehicle arms.

Table 3 – Reason for Study Discontinuation (All Randomized)

	Study 0203		Study 0204	
	Dapsone	Vehicle	Dapsone	Vehicle
Number of Subjects	745	740	761	764
Subjects who Discontinued	107 (14.4%)	123 (16.6%)	133 (17.5%)	139 (18.2%)
Reasons for Discontinuation				
Adverse Experience	3 (0.4%)	5 (0.7%)	3 (0.4%)	4 (0.5%)
Application Site AE	1 (0.1%)	4 (0.7%)	1 (0.1%)	4 (0.5%)
Non-Application Site AE	2 (0.3%)	1 (0.1%)	2 (0.3%)	0 (0.0%)
Lack of Efficacy	3 (0.4%)	7 (0.9%)	6 (0.8%)	8 (1.0%)
Administrative	101 (13.6%)	111 (15.0%)	124 (16.3%)	127 (16.6%)
Patient Noncompliance	3 (0.4%)	2 (0.3%)	7 (0.9%)	4 (0.5%)
Protocol Violation	1 (0.1%)	4 (0.5%)	2 (0.3%)	3 (0.4%)
Lost to Follow-up	58 (7.8%)	59 (8.0%)	74 (9.7%)	74 (9.7%)
Patient Voluntarily Withdrew	32 (4.3%)	41 (5.5%)	39 (5.1%)	43 (5.6%)
Other	7 (0.9%)	5 (0.7%)	2 (0.3%)	3 (0.4%)

Source: Mod 5, Vol 18, pg 81 and Mod 5, Vol 42, pg 84.

3.1.3 Baseline Data

The baseline demographic variables were generally balanced across treatment arms. Studies 0203 and 0204 enrolled similar numbers of males and females, with slightly more females than males enrolled. About 73% of subjects were white. The average subject age was about 19 years and the range was 11 to 81 years. The baseline demographic data is presented in Table 4.

Table 4 – Baseline Demographic Data

		Study 0203		Study 0204	
		Dapsone N=745	Vehicle N=740	Dapsone N=761	Vehicle N=764
Gender	Male	358 (48.1%)	339 (45.8%)	367 (48.2%)	359 (47.0%)
	Female	387 (51.9%)	401 (54.2%)	394 (51.8%)	405 (53.0%)
Race	White	548 (73.6%)	542 (73.2%)	559 (73.5%)	546 (71.5%)
	Black	94 (12.6%)	83 (11.2%)	115 (15.1%)	128 (16.8%)
	Hispanic	73 (9.8%)	81 (10.9%)	65 (8.5%)	64 (8.4%)
	Asian	19 (2.6%)	19 (2.6%)	12 (1.6%)	16 (2.1%)
	Other	11 (1.5%)	15 (2.0%)	10 (1.3%)	10 (1.3%)
Age	Mean	19.0	19.5	19.5	19.6
	Range	12-53	11-59	12-81	12-57

Source: Mod 5, Vol 17, pg 22 and Mod 5, Vol 41, pg 22.

Subjects could be enrolled in the study if they had a clear diagnosis of acne vulgaris, as defined by having 20–50 inflammatory acne lesions and 20–100 non-inflammatory acne lesions. A few subjects had baseline lesion counts outside the specified ranges. There was no specific requirement in the protocol for a minimum score on the GAAS. Consequently, 99 subjects in Study 0203 and 58 subjects in Study 0204 were enrolled with a baseline GAAS equal to 1. A GAAS of 1 is defined as “Minimal: a few non-inflammatory lesions (comedones) are present; a few inflammatory lesions (papules/pustules) may be present” and is one of the success categories at the end of the study. The sponsor defined an ad hoc MITT population that excluded subjects who had a GAAS of 1 at baseline. Most subjects were classified as having mild (2) or moderate (3) acne at baseline. A small number of investigators enrolled most of the subjects who had a baseline GAAS of 1. In Study 0203, Centers 4, 39, and 44 enrolled 56 of the 99 subjects excluded from the MITT population with 34, 12, and 10 subjects respectively enrolled with a baseline GAAS of 1. In Study 0204, Centers 1, 29, and 54 enrolled 40 of the 58 subjects excluded from the MITT population with 13, 16, and 11 subjects respectively enrolled with a baseline GAAS of 1. The distributions of the GAAS and the lesion counts at baseline are presented in Table 5.

Table 5 – Baseline Acne Endpoints

	Study 0203		Study 0204	
	Dapsone N=745	Vehicle N=740	Dapsone N=761	Vehicle N=764
<i>GAAS</i>				
0=None	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
1=Minimal	46 (6.2%)	53 (7.2%)	32 (4.2%)	26 (3.4%)
2=Mild	236 (31.7%)	243 (32.8%)	264 (34.7%)	273 (35.7%)
3=Moderate	447 (60.0%)	425 (57.4%)	447 (58.7%)	440 (57.6%)
4=Severe	16 (2.1%)	19 (2.6%)	18 (2.4%)	25 (3.3%)
<i>Inflammatory</i>				
<20	7 (0.9%)	1 (0.1%)	2 (0.3%)	3 (0.4%)
20-50	716 (96.1%)	722 (97.6%)	745 (97.9%)	745 (97.5%)
>50	22 (3.0%)	17 (2.3%)	14 (1.8%)	16 (2.1%)
Mean	30.8	30.2	30.8	30.4
Range	14-84	18-114	11-114	11-88
<i>Non-Inflam</i>				
<20	3 (0.4%)	3 (0.4%)	0 (0.0%)	1 (0.1%)
20-100	731 (98.1%)	726 (98.1%)	757 (99.5%)	763 (99.9%)
>100	11 (1.5%)	11 (1.5%)	4 (0.5%)	0 (0.0%)
Mean	48.9	49.8	47.5	45.9
Range	13-240	8-172	20-190	9-100

Source: Mod 5, Vol 17, pg 22 and Mod 5, Vol 41, pg 22 and reviewer analysis.

3.1.4 Primary Efficacy Endpoints

3.1.4.1 Analysis Populations

The sponsor analyzed three analysis populations: the ITT, MITT, and per protocol populations. The ITT population was defined as all subjects randomized and dispensed test article. The MITT population was defined after the data were collected and the discovery that a number of subjects were enrolled at baseline with a GAAS equal to 1 (minimal) which was a success category. The MITT population was defined as all ITT subjects with a baseline GAAS ≥ 2 . The per protocol population was defined in the protocol as those subjects without major protocol violations such as violating inclusion criteria, inadequate compliance, use of prohibited medications, or starting birth control pills during the study. The actual violations used to exclude subjects from the per protocol population was determined by the study medical officer and biostatistician prior to unblinding. The final list of protocol violations used to exclude subjects from the per protocol population was: baseline GAAS < 2, patient discontinued, pregnancy, violated inclusion or exclusion criteria (changed contraceptive, baseline lesion counts outside protocol-specified range), use of prohibited medication, and adverse event (sunburn). The numbers of subjects in each population are presented in Table 6.

Table 6 – Analysis Populations

	Study 0203		Study 0204	
	Dapsone	Vehicle	Dapsone	Vehicle
ITT	745 (100.0%)	740 (100.0%)	761 (100.0%)	764 (100.0%)
MITT	699 (93.8%)	687 (92.8%)	729 (95.8%)	738 (96.6%)
Per Protocol	573 (76.9%)	555 (75.0%)	586 (77.0%)	592 (77.5%)

Source: Mod 5, Vol. 17, Page 19 and Mod 5, Vol. 41, Page 19.

3.1.4.2 ITT Analyses

The protocol specified the ITT population as the primary analysis population. The ITT population was defined as all subjects dispensed test article. Missing data was handled via last observation carried forward (LOCF). The protocol specified four primary endpoints: the percent reduction in inflammatory, non-inflammatory, and total lesions and success (clear or minimal) on the GAAS. The results for the percent reduction in lesion count endpoints for the ITT population are presented in Table 7. All three lesion count endpoints are statistically significant in both studies and the results are consistent across the two studies.

Table 7 – Mean Percent Reductions in Lesions at Week 12 and Standard Errors (ITT)

	Study 0203			Study 0204		
	Dapsone N=745	Vehicle N=740	p-value	Dapsone N=761	Vehicle N=764	p-value
Inflammatory	45.9% (1.8)	41.7% (1.8)	0.0302	47.6% (1.7)	40.3% (1.7)	<0.0001
Non-Inflam.	31.1% (2.1)	23.9% (2.2)	0.0022	29.6% (1.9)	21.1% (1.9)	<0.0001
Total	38.3% (1.6)	32.0% (1.7)	0.0004	37.4% (1.5)	29.3% (1.5)	<0.0001

Least squares means and p-values for the treatment effect based on an ANCOVA model with factors treatment and center, and baseline lesion count as covariate.

Source: Mod. 5, Vol. 17, pg. 27, and Mod. 5, Vol. 41, pg. 27.

The analyses for the GAAS endpoint for the ITT population are also statistically significant in both studies. The results for the success on the GAAS for the ITT population are presented in Table 8. Because some subjects were enrolled with a GAAS of 1 at baseline, some subjects are counted as successes in this analysis even though their GAAS score did not improve over-baseline. The sponsor defined an ad hoc MITT population to address the problem of subjects starting the study in the success category. The need for the MITT population was not anticipated by the sponsor in the protocol.

Table 8 – Success (0 or 1) on GAAS at Week 12 (ITT)

	Dapsone	Vehicle	p-value
Study 0203	329/745 (44.2%)	266/740 (36.0%)	0.0003
Study 0204	281/761 (36.9%)	228/764 (29.8%)	0.0017

P-value based on the CMH test stratified on center.

Source: Mod. 5, Vol. 17, pg 26, and Mod. 5, Vol. 41, pg. 26.

3.1.4.3 MITT Analyses

After collecting the data, the sponsor realized that approximately 5% of subjects in Studies 0203 and 0204 were enrolled with a GAAS score equal to 1 at baseline. Since this put these subjects into a success category for the GAAS already at baseline, the sponsor defined an ad hoc MITT population defined as all subjects with GAAS ≥ 2 at baseline. The results for the percent reduction in lesion counts for the MITT population are presented in Table 9. All three lesion count endpoints are statistically still significant in both studies. The MITT results are similar to the ITT results, with the MITT results slightly more significant. The percent reductions on the dapstone arm were comparable in the MITT and ITT populations, but the percent reductions on the vehicle arm were slightly higher in the ITT population than the MITT population.

Table 9 – Mean Percent Reductions in Lesions and Standard Errors (MITT)

	Study 203			Study 204		
	Dapstone N=699	Vehicle N=687	p-value	Dapstone N=729	Vehicle N=738	p-value
Inflammatory	45.8% (1.8)	40.8% (1.9)	0.0158	47.2% (1.7)	39.8% (1.7)	<0.0001
Non-Inflam.	30.6% (2.2)	22.5% (2.2)	0.0009	29.4% (1.9)	20.6 (1.9)	<0.0001
Total	37.9% (1.7)	30.8% (1.7)	0.0001	37.1% (1.5)	28.8% (1.5)	<0.0001

Least squares means and p-values for the treatment effect based on an ANCOVA model with factors treatment and center, and baseline lesion count as covariate.

Source: Mod. 5, Vol. 17, pg. 32, and Mod. 5, Vol. 41, pg.32.

The MITT GAAS analysis is significant in both studies. For this analysis, all subjects must have shown at least one grade improvement over their baseline GAAS classification. The success rates on both the dapstone and vehicle arms are slightly lower after excluding the subjects with baseline GAAS equal to 1, but the treatment effect difference is similar in both the MITT and ITT populations. The MITT results for the GAAS are presented in Table 10.

Table 10 – Success (0 or 1) on GAAS (MITT)

	Dapstone	Vehicle	p-value
Study 203	291/699 (41.6%)	223/687 (32.5%)	<0.0001
Study 204	253/729 (34.7%)	206/738 (27.9%)	0.0032

P-value based on the CMH test stratified on center.

Source: Mod. 5, Vol. 17, pg 31, and Mod. 5, Vol. 41, pg. 31.

3.1.4.4 Per Protocol Analyses

The results of the per protocol analysis are similar to those of the ITT and MITT analyses. The percent change in inflammatory, non-inflammatory, and total lesions and success on the GAAS are all significant in both Studies 0203 and 0204. Thus the conclusions based on the ITT, MITT and per protocol populations are all consistent. Table 11 presents the results of the percent reduction in lesion counts analyses and Table 12 presents the results of the GAAS analyses for the per protocol population.

Table 11 – Mean Percent Reductions in Lesions and Standard Errors (PP)

	Study 203			Study 204		
	Dapsone N=573	Vehicle N=555	p-value	Dapsone N=586	Vehicle N=591	p-value
Inflammatory	50.6% (1.6)	46.1% (1.6)	0.0291	52.4% (1.7)	45.0% (1.8)	0.0002
Non-Inflam.	34.7% (1.9)	27.9% (1.9)	0.0064	35.1% (2.1)	23.9% (2.1)	<0.0001
Total	42.2% (1.5)	35.7% (1.5)	0.0007	42.9% (1.6)	33.1% (1.6)	<0.0001

Least squares means and p-values for the treatment effect based on an ANCOVA model with factors treatment and center, and baseline lesion count as covariate.

Source: Mod. 5, Vol. 17, pg. 33, and Mod. 5, Vol. 41, pg. 33.

Table 12 – Success (0 or 1) on GAAS (PP)

	Dapsone	Vehicle	p-value
Study 203	261/573 (45.6%)	209/555 (37.7%)	0.0022
Study 204	229/586 (39.1%)	178/591 (30.1%)	0.0010

P-value based on the CMH test stratified on center.

Source: Mod. 5, Vol. 17, pg 32, and Mod. 5, Vol. 41, pg. 33.

3.1.5 Secondary Efficacy Endpoints

The mean reduction in inflammatory, non-inflammatory, and total lesions were the secondary endpoints in Studies 0203 and 0204. The reduction in lesion endpoints were analyzed similarly to the percent reduction endpoints. The reductions were analyzed with an ANCOVA model with factors for treatment, center, and baseline lesion count. All reductions in lesion counts were significant in both studies. Dapsone reduced about 1.4 to 2.3 more inflammatory lesions and about 4.2 to 4.3 more non-inflammatory lesions than vehicle in the two studies. The results are presented in Table 13.

Table 13 – Mean Reductions in Lesions and Standard Errors (ITT)

Study 0203					
	Dapsone N=745		Vehicle N=740		p-value
	Baseline	Reduction	Baseline	Reduction	
Inflammatory	30.8	13.7 (0.6)	30.2	12.3 (0.6)	0.0265
Non-Inflam.	48.9	16.4 (1.0)	49.8	12.1 (1.1)	0.0001
Total	79.7	30.4 (1.4)	80.0	24.6 (1.4)	0.0001
Study 0204					
	Dapsone N=761		Vehicle N=764		p-value
	Baseline	Reduction	Baseline	Reduction	
Inflammatory	30.8	14.3 (0.5)	30.4	12.0 (0.5)	0.0001
Non-Inflam.	47.5	13.9 (0.9)	45.8	9.7 (1.0)	0.0001
Total	78.3	28.4 (1.3)	76.2	21.7 (1.3)	<0.0001

Least squares means and p-values for the treatment effect based on an ANCOVA model with factors treatment and center, and baseline lesion count as covariate.

Source: Mod. 5, Vol. 17, pg. 30, and Mod. 5, Vol. 41, pg. 30.

3.1.6 Statistical Issues

3.1.6.1 Clear/Minimal on the GAAS with at least Two Grades Reduction

Recently the Division has been recommending that sponsors define success on the global evaluation as achieving clear or almost clear with at least a two grade reduction from baseline. Especially in light of the fact that a number of subjects in Studies 0203 and 0204 were enrolled with GAAS scores equal to 1, this definition of success may be a useful supportive analysis. Table 14 presents the results of the analysis where success on the GAAS is defined as achieving a 0 or 1 with at least 2 grades reduction from baseline. This analysis is conducted on the MITT population, as subjects who enroll with a score of 1 cannot achieve a reduction of at least two grades. The success rates are reduced from those seen in the protocol analysis of success being defined as achieving a 0 or 1, since those subjects moving from 2 to 1 are no longer counted as successes. This endpoint is, however, statistically significant in both studies.

Table 14 – Success (0 or 1 and at least 2 grades decrease) on GAAS (MITT)

	Dapsone	Vehicle	p-value
Study 203	161/699 (23.0%)	113/687 (16.5%)	0.0009
Study 204	115/729 (15.8%)	81/738 (11.0%)	0.0077

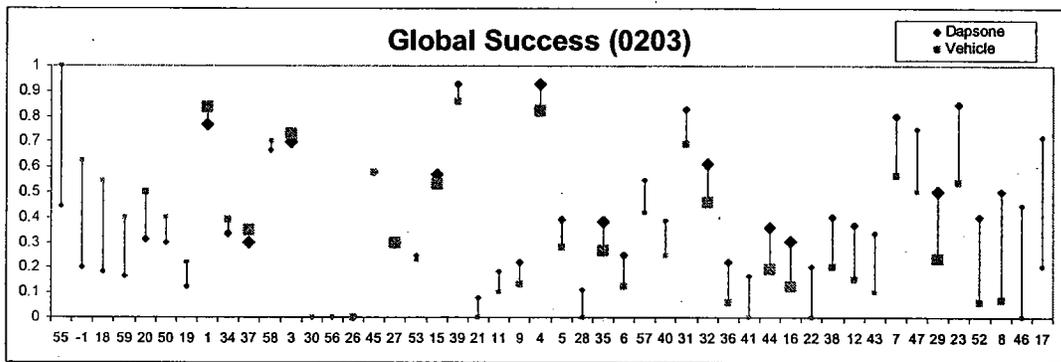
P-value based on the CMH test stratified on center.

Source: Reviewer analysis

3.1.6.2 Efficacy Results by Center

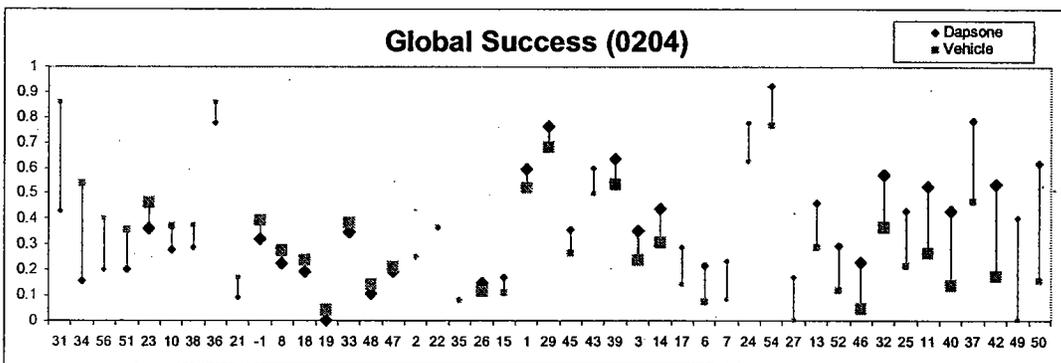
Studies 0203 and 0204 demonstrated a fair amount of center to center variability. On the GAAS, some centers had close to a 90% success rate on both arms while other centers had no successes on either arm. Because of the relatively modest treatment effect and the large variability, about one third of the centers had higher success rates on vehicle than on dapson. The centers with the largest point estimates for the treatment differences (either favoring dapson or vehicle) were generally the smaller centers with 5-12 subjects per arm, though there are exceptions. Figure 1 and Figure 2 display the GAAS success rates by center for Studies 0203 and 0204, respectively. Results for the percent change in inflammatory and non-inflammatory lesions are similar and the results are displayed in the Appendix in Figure 7 and Figure 8.

Figure 1 – Success (0,1) on the GAAS by Center (Study 0203, ITT)



Small symbols = 10-25 subjects per center, Medium symbols = 26-39 subjects per center, Large symbols = 40-60 subjects per center
 Center -1 represents the pooled results from Centers 10, 24, 33, and 60 (the centers enrolling fewer than 5 subjects per treatment arm).

Figure 2 – Success (0,1) on the GAAS by Center (Study 0204, ITT)



Small symbols = 10-25 subjects per center, Medium symbols = 26-39 subjects per center, Large symbols = 40-60 subjects per center
 Center -1 represents the pooled results from Centers 4, 9, 12, 16, 28, 44, 53, 55, and 57 (the centers enrolling fewer than 5 subjects per treatment arm).

Studies 0203 and 0204 were run concurrently with Study 0114, the open-label 12-month safety study. Thirteen of the investigators from Studies 0203 and 0204 also participated in Study 0114. There may be some question as to whether investigators concurrently participating in an open-label study might be more easily unblinded in the pivotal studies due to potential knowledge of characteristics of the test medication from the open-label study. Investigators 3, 20, 26, 28, 41, and 46 from Study 0203 and Investigators 11, 12, 18, 26, 39, and 40 from Study 0204 also participated in Study 0114. These investigators appear to range from those having higher success on vehicle to those having higher success on dapstone, with no clear trends present or evidence that possible unblinding impacted the results.

3.1.6.3 Pooling of Centers

The protocol and the statistical analysis plan stated that centers enrolling fewer than 5 subjects per treatment arm would be pooled for analyses involving center. However, in the study reports the sponsor did not pool any centers together, even though Study 0203 had 4 centers and Study 0204 had 9 centers enrolling fewer than 5 subjects per treatment arm. The study reports do not explain why pooling was not performed per the protocol. The analyses presented above use the individual centers without pooling. This reviewer compared the analyses using pooled centers as defined in the protocol and those presented by the sponsor in the study report that do not use pooling. The results are similar and all conclusions are the same. P-values from each approach for the ITT population are presented in Table 15. The only test for interaction that was significant was the Breslow-Day test for success on the GAAS in Study 0203. The success rates by pooled center are presented in Figure 1 above.

Table 15 – P-value Comparison using Unpooled and Pooled Centers (ITT)

Endpoint	Unpooled		Pooled ¹	
	Treatment Effect	Interaction ²	Treatment Effect	Interaction ²
Study 0203				
GAAS	0.0003	0.0477	0.0003	0.0491
% Ch. Infl	0.0302	0.6847	0.0295	0.7796
% Ch. Non	0.0022	0.3919	0.0025	0.3410
% Ch. Total	0.0004	0.2336	0.0004	0.1965
Study 0204				
GAAS	0.0017	0.3402	0.0024	0.3102
% Ch. Infl	<0.0001	0.1996	0.0001	0.2700
% Ch. Non	<0.0001	0.5610	<0.0001	0.6781
% Ch. Total	<0.0001	0.3954	<0.0001	0.5094

¹ For Study 0203, pooling combines Centers 10, 24, 33, and 60 into one analysis center. For Study 0204, pooling combines Centers 4, 9, 12, 16, 28, 44, 53, 55, and 57 into one analysis center.

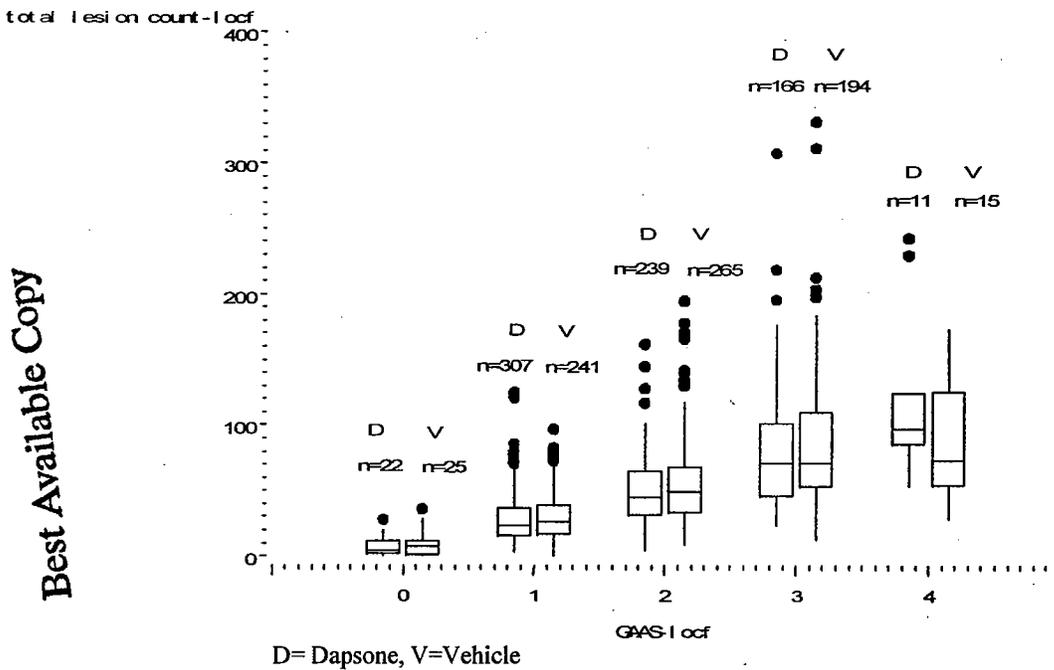
² For the GAAS, the p-value is from the Breslow-Day test. For percent change, the p-value is for the treatment by center interaction in an ANCOVA model with terms for baseline lesion count, treatment, center, and treatment*center. Treatment effect p-values for the percent change endpoints are from the ANCOVA model with terms for baseline lesion count, treatment, and center.

Source: Reviewer analysis.

3.1.6.4 Lesion Counts and the GAAS

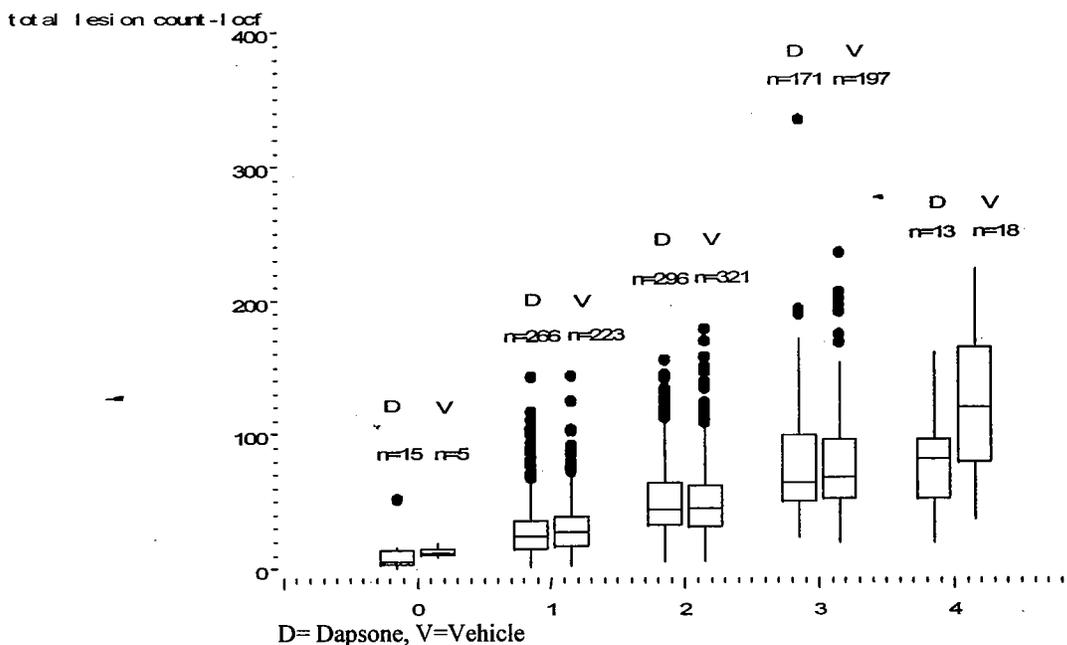
Some subjects counted as successes under the GAAS seemed to have relatively high lesion counts for the definition of 'none' (no evidence of facial acne vulgaris) or 'minimal' (a few non-inflammatory lesions (comedones) are present; a few inflammatory lesions (papules/pustules) may be present). Subjects scored as 0 (none) had as many as 9 inflammatory lesions or 51 non-inflammatory lesions. Subjects scored as 1 (minimal) had as many as 49 inflammatory lesions or 115 non-inflammatory lesions. As can be seen in Figure 3 and Figure 4 below, lesion counts do generally increase with increasing GAAS, but there is considerable overlap between the categories and the distributions are skewed. The success categories of 'none' and 'minimal' appear to contain many subjects with more than a 'few' inflammatory and non-inflammatory lesions.

Figure 3- Total Lesion Count by GAAS at Week 12, ITT (Study 0203)



APPEARS THIS WAY
ON ORIGINAL

Figure 4 - Total Lesion Count by GAAS at Week 12, ITT (Study 0204)



Best Available Copy

3.1.6.5 Nodules

The sponsor did not include nodules as part of either the inflammatory or total lesion counts, though the number of nodules was recorded separately. Subjects were not to have active or developing nodules at baseline, but nodules are not otherwise defined (i.e. size or color) in the protocol. However, 8 subjects in Study 0203 and 3 subjects in Study 0204 were classified as having 1 to 4 nodules at baseline. At week 12, 11 subjects in Study 0203 and 6 subjects in Study 0204 had at least one nodule and were also classified as successes on the GAAS. Two of the 11 subjects in Study 0203 had GAAS scores of 0. (Subject 0122 (vehicle) had 6 inflammatory lesions, 22 non-inflammatory lesions, and 1 nodule with a GAAS of 0, and Subject 1719 (vehicle) had 5 inflammatory lesions, 2 non-inflammatory lesions, and 1 nodule with a GAAS of 0.)

3.1.7 Study 0004

In addition to the two pivotal Phase 3 studies, the sponsor conducted an additional Phase 3 study of dapson for the treatment of acne. The study enrolled subjects 12 years old or older with 20 to 50 inflammatory and ≥ 20 non-inflammatory lesions at baseline. Subjects applied test medication twice daily for 12 weeks. Subjects were evaluated at baseline, Weeks 2, 4, 6, 8, and 12. The primary efficacy timepoint was Week 12. The study enrolled 496 subjects, 330 on dapson and 166 on vehicle. Many of the investigators who conducted Study 0004 were also investigators in Study 0204.

This study uses slightly different endpoints than those recommended by the Division for acne trials. The primary efficacy endpoints were the final lesion counts for inflammatory, non-inflammatory, and total (two out of three must be significant) and

success on the GAAS. The GAAS scale was defined the same way as in Studies 0203 and 0204 except that investigators were permitted to assign half-step scores in-between the defined integer scores. In the study report, success was defined as a score ≤ 1.5 . (However, in the protocol success was defined as improvement by at least 1 grade, but this analysis was not included in the study report.) Secondary endpoints were the reduction and percent reduction in lesion counts. Lesion count endpoints were analyzed with ANCOVA with terms baseline lesion count, center, and treatment by center interaction. Treatment by center interaction was tested at $\alpha=0.10$. Success on the GAAS was analyzed with a Cochran-Mantel-Haenszel test stratified on center. All primary efficacy endpoints were significant in Study 0004. The efficacy results for the Week 12 lesion counts (primary endpoints), success on the GAAS (primary endpoint) and percent reduction in lesion counts (secondary endpoints) are presented in Table 16. The efficacy results from Study 0004 are consistent with and supportive of the results from the pivotal studies 0203 and 0204.

Table 16 – Efficacy Results, ITT (Study 0004)

	Week 12 Mean ^a			Percent Reduction ^b		
	Dapsone N=330	Vehicle N=166	p-value	Dapsone N=330	Vehicle N=166	p-value
Inflammatory	20.7 (0.7)	24.1 (1.0)	0.003	37.2% (2.1)	26.6% (2.8)	0.001
Non-Inflam.	39.5 (1.5)	46.0 (2.0)	0.004	27.5% (2.5)	16.8% (3.3)	0.005
Total	60.2 (1.9)	70.4 (2.5)	<0.001	32.0% (2.0)	21.9% (2.6)	0.001
GAAS Success	88/330 (26.7%)	31/166 (18.7%)	0.042			

Standard errors are presented in parentheses.

^a Primary Endpoints

^b Secondary Endpoints

Source: Mod 5, Vol 66, pg 28 and 30.

3.2 Evaluation of Safety

3.2.1 Studies 0203 and 0204

3.2.1.1 Extent of Exposure

The extent of exposure to dapson and vehicle was comparable in Studies 0203 and 0204. The median number of days on treatment was 84 or 85 days for both treatment arms in both studies, with the mean ranging from 75.5 days to 77.9 days. The total grams of study drug used were also similar among treatment arms. Dapsone subjects used an average of 105.4 to 106.2 g in the two studies (median = 88) and vehicle subjects used an average of 101.8 to 103.4 g in the two studies (medians 88 and 91). For dapson subjects, the maximum total grams exposure was 447.8 g. The average daily drug use was 1.4 g/day (median= 0.6 g/day) for both arms in the two studies. The maximum average daily use for dapson was 8.5 g/day.

3.2.1.2 Adverse Events

Adverse event rates in Studies 0203 and 0204 were similar for the dapson and vehicle arms. Both application site and non-application site events occurred at about equal rates in the dapson and vehicle arms. The most common adverse events were application site reaction (not otherwise specified), application site dryness, and application site erythema. The most common adverse events are listed in Table 17.

Table 17 – Number of Subjects with Adverse Events

	Study 0203		Study 0204	
	Dapson n=730	Vehicle n=726	Dapson n=736	Vehicle n=741
All Adverse Events	458 (63%)	466 (64%)	409 (56%)	407 (55%)
Application Site	335 (46%)	328 (45%)	294 (40%)	304 (41%)
Non-Application Site	266 (36%)	267 (37%)	200 (27%)	206 (28%)
Most Common AEs ^a				
Application site reaction NOS	187 (26%)	192 (26%)	139 (19%)	152 (21%)
Application site dryness	159 (22%)	145 (20%)	142 (19%)	134 (18%)
Application site erythema	139 (19%)	124 (17%)	103 (14%)	111 (15%)
Nasopharyngitis	37 (5%)	44 (6%)	35 (5%)	49 (7%)

^a Those AEs occurring in at least 5% of subjects in either arm

3.2.1.3 G-6-PD Deficiency

Subjects in Studies 0203 and 0204 were screened for G-6-PD deficiency at Week 12. In Study 0203, 5 dapson and 7 vehicle subjects were G-6-PD deficient. In Study 0204, 14 dapson and 18 vehicle subjects were G-6-PD deficient. Subjects from Study 0114 (the 12-month long-term safety study) and Study 0110 (a study comparing blood levels from oral and topical dapson) were also screened for G-6-PD deficiency. Five subjects in Study 0114 and one subject in Study 0110 were G-6-PD deficient. Thus out of the 2030 dapson-treated subjects in studies where G-6-PD deficiency was assessed, 25 dapson-treated subjects were found to be G-6-PD deficient.

3.2.2 Study 0114

Study 0114 is a 12-month open-label long term safety study. Subjects were evaluated for adverse events and laboratory assessments were conducted. Plasma dapson levels were also collected. Subjects were screened for G-6-PD deficiency. Study 0114 enrolled 506 subjects.

The study report for Study 0114 was signed off on June 7, 2004 and was submitted as a final report in the initial NDA submission. However, under the guise of the 120-day safety update, the sponsor submitted a revised dataset and study report containing additional plasma dapson concentration data and corrections to the original study report. The sponsor's explanation for the revisions is as follows (page 1 of Amendment 01 of 2.7.2, of the January 19, 2005 submission)

The study report DAP0114 dated June 2, 2004 included available plasma dapson and n-acetyl dapson information. After reviewing the original data, QLT USA, Inc. (formerly Atrix Laboratories, Inc.) discovered that some laboratory data were missing from the dataset. Dapson and n-acetyl dapson plasma levels were missing for two patients at Week 2, 172 patients at Month 1, and ten patients at Month 3, 27 patients at Month 6, 17 patients at Month 9, at two patients at Month 12.

When contacted about this error, _____ located the remaining samples, which were stored frozen, and sent them to the laboratory _____ for analysis. These results were pending at the time of the 5% Dapson Topical Gel submission. The original study report (dated June 2, 2004) also indicated that a total of four patients in the study were found to be G-6-PD deficient. One G-6-PD deficient patient (Patient #1314), who was tested at the end of the study (Month 12) has been added to the G-6-PD deficient population.

The Agency was *not* advised at the time of the original submission (September 7, 2004) or at any point before the January 19 amendment arrived that the sponsor had discovered missing data and was having samples analyzed and that additional data would be forthcoming. The sponsor notes that the missing samples were shipped for analysis on April 2004—two months before the study report was finalized and five months before the NDA was submitted to the Agency as a complete submission. The sponsor claims that dapson concentrations in plasma are not affected by long term freezer storage based on a validation of assays study they conducted.

3.2.2.1 Extent of Exposure

Study 0114 enrolled 506 subjects. According to the sponsor's study report, the mean number of days on study drug was 253 days with a median of 326 days and a maximum of 380 days. This reviewer could not confirm the sponsor's calculations based on the datasets provided. For the (possibly different) endpoint of *days in the study*, the mean number of days was 266, the median was 336, and the maximum was 393. The number of subjects who discontinued by three-month time intervals is presented in Table 18. Thirty-two percent of the subjects discontinued before 9 months. Based on the 349 subjects with reconcilable drug weights, the mean daily drug use was 1.35 g/day (median=1.07) and the maximum was 11.02 g/day.

Table 18 – Length of Time Enrolled in the Study (Study 0114)

Time in Study	N=506
0 to 3 months (1 to 92 days)	78 (15%)
3 to 6 months (93 to 184 days)	62 (12%)
6 to 9 months (185 to 275 days)	21 (4%)
9 to 12+ months (276 to 393 days)	345 (68%)

Source: Reviewer analysis.

3.2.2.2 Adverse Events

The overall percentage of subjects reporting adverse events in Study 0114 was similar to the percentages in the Phase 3 trials, however, the percentage of subjects reporting application site events was much lower in the long-term study (14%) than in the pivotal trials (40-45%). The most common adverse events are presented in Table 19. All individual application site reactions occurred in 3% or fewer of subjects. The discrepancies between the rates of application site reaction in the pivotal studies and the long-term study may be due in part to the fact that in the pivotal studies, investigators directly queried subjects about erythema, dryness, oiliness, and peeling and rated them on a scale from absent to severe, whereas in Study 0114, all local skin reaction reports were spontaneous reports and were not specifically queried.

Table 19 – Common Adverse Events (Study 0114)

	n=486
All Adverse Events	330 (68%)
Application Site	67 (14%)
Non-Application Site	312 (64%)
Most Common AEs ^a	
Headache NOS	98 (20%)
Nasopharyngitis	75 (15%)
Pharyngitis	43 (9%)
Dysmenorrhea	31 (6%)
Sinusitis NOS	28 (6%)
Upper Resp Tract Inf NOS	24 (5%)
Most Common Appl. Site AEs ^b	
Application Site Dryness	14 (3%)
Application Site Rash	12 (3%)
Sunburn	11 (2%)
Application Site Burning	8 (2%)
Application Site Erythema	8 (2%)
Application Site Pruritus	7 (1%)
Acne Aggravated	6 (1%)
Application Site Reaction NOS	5 (1%)

^a Occurring in more than 5% of subjects

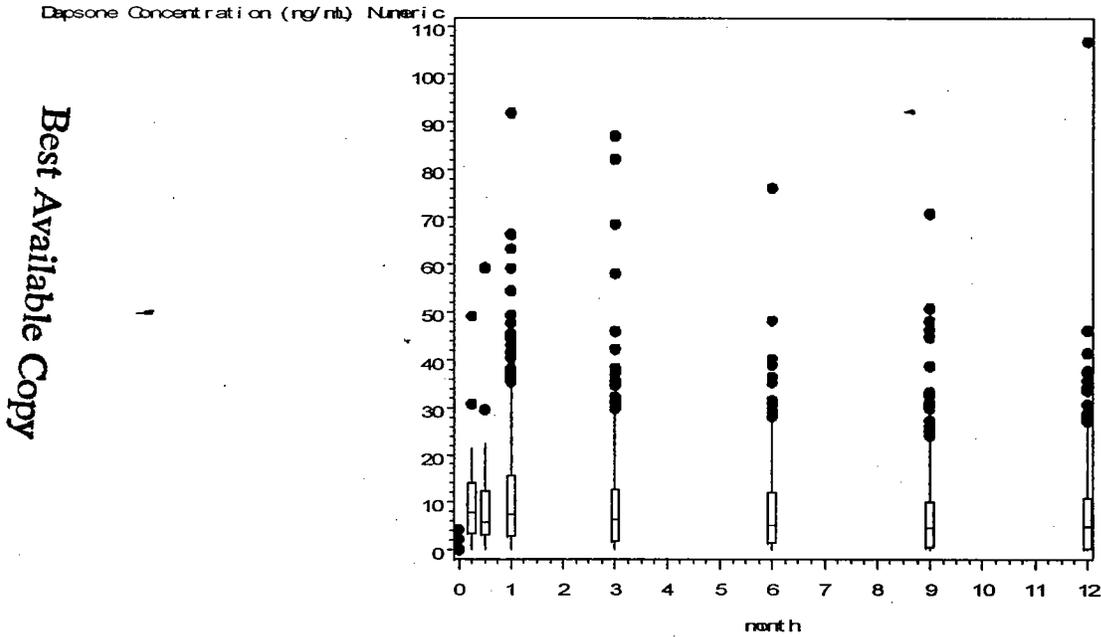
^b Occurring in more than 1% of subjects

3.2.2.3 Dapsone Concentrations

The dapsons concentrations were relatively constant of the course of the 12-month study. The post-baseline mean concentrations ranged from 7.5 to 11.0 ng/mL during the study with the highest mean occurring at Month 1 and the lowest mean occurring at Month 12. The 90th percentiles for the dapsons concentrations ranged from 18.5 ng/mL at Month 9 to 25.6 ng/mL at Month 1. The maximum dapsons concentration for any subject was 107 ng/mL (Subject 1809 at Month 12). Subject 1809 was the subject who used the largest amount of study medication (approximately 10 times the mean usage). Subject 1809 used 3701.1 g of medication during the study, or an average of 11.0 g per day, compared

to the mean usage for the study of a total of 360.2 g or 1.3 g per day. The distribution of the dapson concentrations by visit is presented in Figure 5.

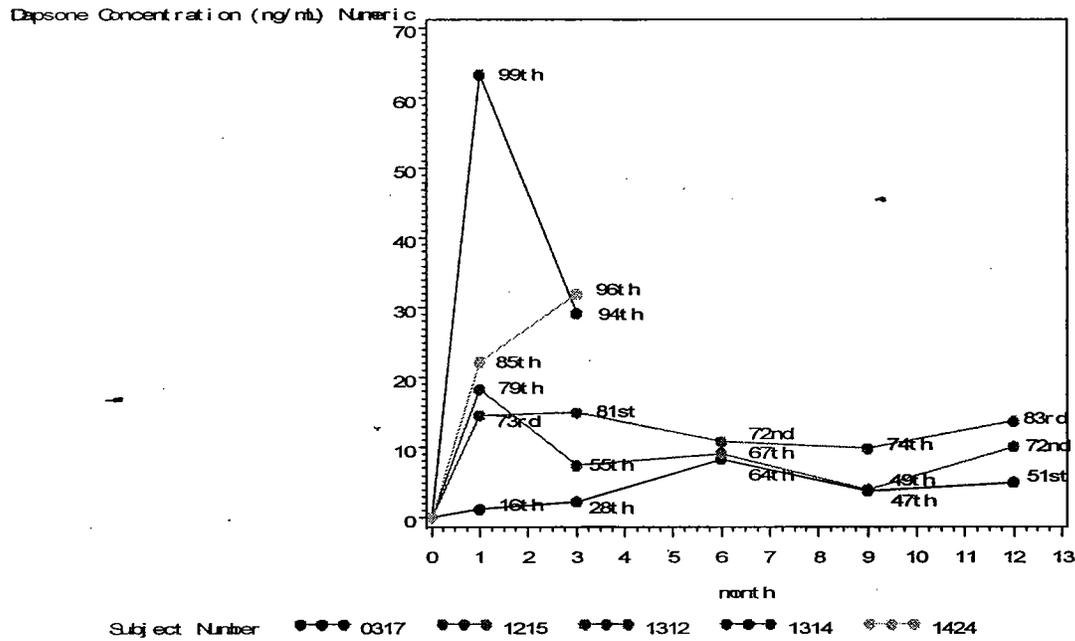
Figure 5 – Dapson Concentrations by Visit (Study 0114)



The dapson concentrations for the 5 G-6-PD deficient subjects are displayed in Figure 6. Only the dapson concentration values through Month 3 are available for Subjects 1314 and 1424, although Subject 1424 reportedly completed the study and attended all visits. Subject 1314 ‘voluntarily withdrew’ from the Study at Month 6. The dapson concentrations for the two subjects with partial data are on the higher end of the observed data for these timepoints (85th to 99th percentile). The remaining three subjects had observations between the 16th and 83rd percentiles.

**APPEARS THIS WAY
ON ORIGINAL**

Figure 6 – Dapsons Concentrations for G-6-PD Deficient Subjects (Study 0114)



Note: The labels represent the Subject's percentile relative to the other subjects with dapsons measurements at each visit.

4 Findings in Special/Subgroup Populations

4.1 Gender, Race, and Age

Females had greater percent reductions in lesions and global success rates than males, though in most cases the treatment differences between dapsons and vehicle were roughly the same. In general, the percent reductions and success on the GAAS were about 10% higher for females than males on both the dapsons and vehicle arms. The efficacy results by gender are presented in Table 20. While there were differences between males and females regarding the number of baseline lesions, with males averaging 10 more lesions than females at baseline (see Table 21), this baseline discrepancy does not appear to fully account for the higher success rates for females. In an exploratory linear model with effects for baseline lesion count, treatment, gender, and center, gender was a highly significant covariate ($p < 0.0001$) for all percent reduction in lesion counts in both studies even after adjusting for baseline count. Within the gender subgroups, the percent reduction in lesions and global success were higher on dapsons than vehicle.

APPEARS THIS WAY
ON ORIGINAL

Table 20 – Percent Reduction in Lesions and Success on the GAAS by Gender (ITT)

	Study 0203			
	Male		Female	
	Dapsone N=358	Vehicle N=339	Dapsone N=387	Vehicle N=401
Inflammatory	39.9%	34.8%	56.6%	52.2%
Non-Inflammatory	29.0%	17.8%	38.9%	34.8%
Total	34.8%	25.6%	46.4%	41.8%
GAAS ^a	35.1%	26.2%	47.7%	38.0%
	Study 0204			
	Male		Female	
	Dapsone N=367	Vehicle N=359	Dapsone N=394	Vehicle N=405
Inflammatory	41.9%	33.3%	54.1%	47.8%
Non-Inflammatory	26.0%	16.6%	38.0%	29.9%
Total	33.1%	24.1%	44.4%	37.4%
GAAS ^b	27.5%	24.5%	41.6%	31.0%

Table displays unadjusted means.

^a Based on the MITT population with 336 male/dapsone subjects, 324 male/vehicle subjects, 363 female/dapsone subjects, and 363 female/vehicle subjects.

^b Based on the MITT population with 356 male/dapsone subjects, 351 male/vehicle subjects, 373 female/dapsone subjects, and 387 female/vehicle subjects.

Table 21 – Baseline and Week 12 Mean Lesion Counts by Gender (ITT)

	Study 0203							
	Male				Female			
	Dapsone N=358		Vehicle N=339		Dapsone N=387		Vehicle N=401	
	BL	W12	BL	W12	BL	W12	BL	W12
Inflam.	33.5	20.0	32.7	21.2	28.4	12.4	28.1	13.6
Non-Inflam.	51.7	35.4	52.3	42.4	46.2	27.6	47.7	31.1
Total	85.2	55.4	85.1	63.6	74.6	39.9	75.8	44.7
	Study 0204							
	Male				Female			
	Dapsone N=367		Vehicle N=359		Dapsone N=394		Vehicle N=405	
	BL	W12	BL	W12	BL	W12	BL	W12
Inflam.	33.9	19.8	32.8	22.6	28.0	13.1	28.2	14.7
Non-Inflam.	49.8	37.3	48.0	40.1	45.3	28.0	44.0	31.6
Total	83.7	57.1	80.8	62.6	73.3	41.0	72.2	46.3

Table displays unadjusted means.

Results in the race subgroups are mixed in Studies 0203 and 0204. In Study 0204, the results on the dapson arm are higher than on the vehicle arm for each endpoint within each racial subgroup. However, in Study 0203 the results are more mixed. In Study 0203 black and Hispanic subjects had more favorable results on vehicle than dapson. It

is not clear why the black and Hispanic dapson subjects did worse than vehicle subjects in Study 0203, especially as these results are not replicated in Study 0204. Race subgroup results are presented in Table 22 for Study 0203 and Table 23 for Study 0204.

Table 22 – Percent Reduction in Lesions and Success on the GAAS by Race (Study 0203, ITT)

	Study 0203			
	Caucasian		Black	
	Dapson N=548	Vehicle N=542	Dapson N=94	Vehicle N=83
Inflammatory	47.2%	40.3%	47.5%	56.6%
Non-Inflammatory	34.9%	24.2%	29.2%	34.1%
Total	41.0%	31.5%	37.6%	42.0%
GAAS ^a	43.7%	31.1%	38.9%	38.0%
	Hispanic		Asian/Nat. Am./Other	
	Dapson N=73	Vehicle N=81	Dapson N=30	Vehicle N=34
Inflammatory	55.3%	56.5%	60.8%	46.9%
Non-Inflammatory	31.8%	34.8%	41.0%	37.2%
Total	40.5%	43.4%	48.2%	40.4%
GAAS ^a	29.6%	32.5%	43.3%	39.4%

Table displays unadjusted means.

^a Based on the MITT population with 508 Caucasian/dapson, 498 Caucasian/vehicle, 90 Black/dapson, 79 Black/vehicle, 71 Hispanic/dapson, 77 Hispanic/vehicle, 30 Other/dapson, and 33 Other/vehicle subjects.

Table 23 – Percent Reduction in Lesions and Success on the GAAS by Race (Study 0204, ITT)

	Study 0204			
	Caucasian		Black	
	Dapson N=559	Vehicle N=546	Dapson N=115	Vehicle N=128
Inflammatory	47.1%	38.7%	56.8%	54.6%
Non-Inflammatory	31.6%	24.7%	32.8%	22.1%
Total	38.2%	31.1%	42.1%	34.1%
GAAS ^a	32.6%	28.2%	50.5%	31.9%
	Hispanic		Asian/Nat. Am./Other	
	Dapson N=65	Vehicle N=64	Dapson N=22	Vehicle N=26
Inflammatory	44.5%	39.6%	44.5%	26.6%
Non-Inflammatory	35.4%	22.8%	34.8%	11.9%
Total	39.8%	30.7%	39.1%	18.7%
GAAS ^a	24.6%	23.4%	40.9%	15.4%

Table displays unadjusted means.

^a Based on the MITT population with 537 Caucasian/dapson, 529 Caucasian/vehicle, 105 Black/dapson, 119 Black/vehicle, 65 Hispanic/dapson, 64 Hispanic/vehicle, 22 Other/dapson, and 26 Other/vehicle subjects.

Similarly to the way that females had better results for both dapson and vehicle than males did, adult subjects had better results on both arms than adolescent subjects did. However, again the treatment differences in both age groups appear to be roughly the same magnitude. Efficacy results by age group are presented in Table 24.

Table 24 – Percent Reduction in Lesions and Success on the GAAS by Age (ITT)

	Study 0203			
	<18 years		≥18 years	
	Dapsone N=456	Vehicle N=424	Dapsone N=289	Vehicle N=316
Inflammatory	44.6%	39.1%	54.8%	51.1%
Non-Inflammatory	27.5%	18.1%	44.6%	39.1%
Total	35.6%	27.1%	49.1%	44.2%
GAAS ^a	38.1%	27.5%	47.1%	39.1%
	Study 0204			
	<18 years		≥18 years	
	Dapsone N=433	Vehicle N=426	Dapsone N=328	Vehicle N=338
Inflammatory	46.6%	36.6%	51.2%	46.6%
Non-Inflammatory	28.5%	18.0%	37.1%	30.8%
Total	35.8%	26.3%	43.2%	37.2%
GAAS ^b	32.5%	26.3%	37.8%	30.0%

Table displays unadjusted means.

^a Based on the MITT population with 425 pediatric/dapsone subjects, 393 pediatric/vehicle subjects, 274 adult/dapsone subjects, and 294 adult/vehicle subjects.

^b Based on the MITT population with 422 pediatric/dapsone subjects, 415 pediatric/vehicle subjects, 307 adult/dapsone subjects, and 323 adult/vehicle subjects.

In summary, the subgroup analyses by gender, race, and age indicate that gender and age may have some impact on the efficacy results with female and adult subjects generally having higher success rates and percent reductions than male and adolescent subjects. These higher success rates pertain to both dapson and vehicle and do not appear to substantially impact the magnitude of the treatment difference between dapson and vehicle, which remains reasonably constant across subgroups.

4.2 Other Special/Subgroup Populations

Not applicable.

5 Summary and Conclusions

5.1 Statistical Issues and Collective Evidence

In both of the Phase 3 studies, dapson topical gel was statistically superior to its vehicle in terms of the percent reduction in inflammatory, non-inflammatory, and total lesions and in terms of success on the GAAS, as specified in the protocol. The Phase 3 studies were very large (1485 and 1525 subjects) and were adequately powered to detect relatively small differences between treatments. In Study 0203, the average net benefit of

dapsone over vehicle in terms of reductions in lesions was of 4.2% or 1.4 more inflammatory lesions and 7.2% or 4.3 more non-inflammatory lesions. In Study 0204, the average net benefit of dapsone over vehicle was 7.3% of 2.3 more inflammatory lesions and 8.5% or 4.2 more non-inflammatory lesions. These results are summarized in Table 25.

Table 25 – Percent Reduction and Absolute Reduction in Lesions from Baseline (ITT)

	Study 0203			Study 0204		
	Dapsone N=745	Vehicle N=740	p-value	Dapsone N=761	Vehicle N=764	p-value
Infl. (% Red)	45.9%	41.7%	0.0302	47.6%	40.3%	<0.0001
(Abs Red)	13.7	12.3	0.0265	14.3	12.0	0.0001
Non-Inf. (% Red)	31.1%	23.9%	0.0022	29.6%	21.1%	<0.0001
(Abs Red)	16.4	12.1	0.0001	13.9	9.7	0.0001
Total (% Red)	38.3%	32.0%	0.0004	37.4%	29.3%	<0.0001
(Abs Red)	30.4	24.6	0.0001	28.4	21.7	<0.0001

Table presents least squares means adjusted for baseline lesion count and center. P-values are based on an ANCOVA model with terms for treatment, center, and baseline lesion count.

Interpretation of the Global Acne Assessment score (GAAS) is complicated by the fact that 6.7% of subjects in Study 0203 and 3.8% of subjects in Study 0204 were enrolled with baseline GAAS scores of 1 (minimal). Success at Week 12 was defined as a GAAS score of 0 (none) or 1 (minimal), thus these subjects were enrolled at a severity considered a success post-treatment. Study entry criteria were only defined in terms of the number of lesions, and a minimum GAAS score was not required at baseline. The sponsor defined an ad hoc MITT population that excluded subjects with baseline GAAS of 1.

The interpretation of a “few” or “no” lesions also seemed to vary from investigator to investigator. The success categories were defined as 0 - None: no evidence of facial acne vulgaris, and 1 - Minimal: a few non-inflammatory lesions (comedones) are present; a few inflammatory lesions (papules/pustules) may be present. However, subjects scored as 0 (none) had as many as 9 inflammatory lesions or 51 non-inflammatory lesions and subjects scored as 1 (minimal) had as many as 49 inflammatory lesions or 115 non-inflammatory lesions.

All analyses based on the GAAS, both those specified in the protocol (success = 0,1 for ITT, PP) and post hoc (success=0,1 for MITT, success= 0,1 + 2 grades reduction for MITT), were all statistically significant. The success rates for the MITT analysis (success=0,1) are: 41.6% (dapson) versus 32.5% (vehicle) for Study 0203, and 34.7% (dapson) versus 27.9% (vehicle) for Study 0204.

Efficacy rates varied greatly among centers. A number of centers had higher efficacy with the vehicle than with dapson. Because each study had more than 50 centers and the

fact that the treatment difference between dapson and vehicle was relatively modest—in the range of 4-9%—it does not seem too surprising that a number of centers favored vehicle over dapson. In addition, the presence of one or two extreme outliers within a center could greatly influence the center mean. Examination of subgroups indicates that gender and age have an impact on efficacy results with females and adults generally having better outcomes than males and adolescents, though treatment differences across subgroups did not vary greatly (i.e. females and adults had better results on both the dapson and vehicle arms). Some of this variability may be due to the fact that males and adolescents generally had higher baseline lesion counts than females and adults. Some of the center variability may be confounded with gender and age variability as different centers enrolled different demographics (i.e. some centers enrolled more adult females while other centers enrolled more adolescent males).

5.2 Conclusions and Recommendations

Dapson is statistically superior to its vehicle in two studies for all primary endpoints (percent change in inflammatory, non-inflammatory, and total lesions, and success on the GAAS). The efficacy benefit of dapson over vehicle is relatively modest, with dapson efficacy results about 4 to 9% better than vehicle results. The two Phase 3 studies had overall results that were very similar to each other, but within studies there was a lot of center to center variation in response. Although it was not a pre-specified hypothesis, there is some evidence from the subgroup analyses that female subjects had better overall results than males, and that adult subjects had better overall results than adolescent subjects. The treatment differences between dapson and vehicle were roughly the same, however, in the different subgroups. One design flaw with the studies was that investigators were not required to enroll subjects with a minimum global evaluation of mild (2) at baseline and consequently a number of subjects were enrolled with a GAAS score of 1 (minimal) at baseline, which meant that those subjects would not have to improve any from baseline to be considered a success on the GAAS at the end of the study. This necessitated a post hoc MITT population to exclude those subjects.

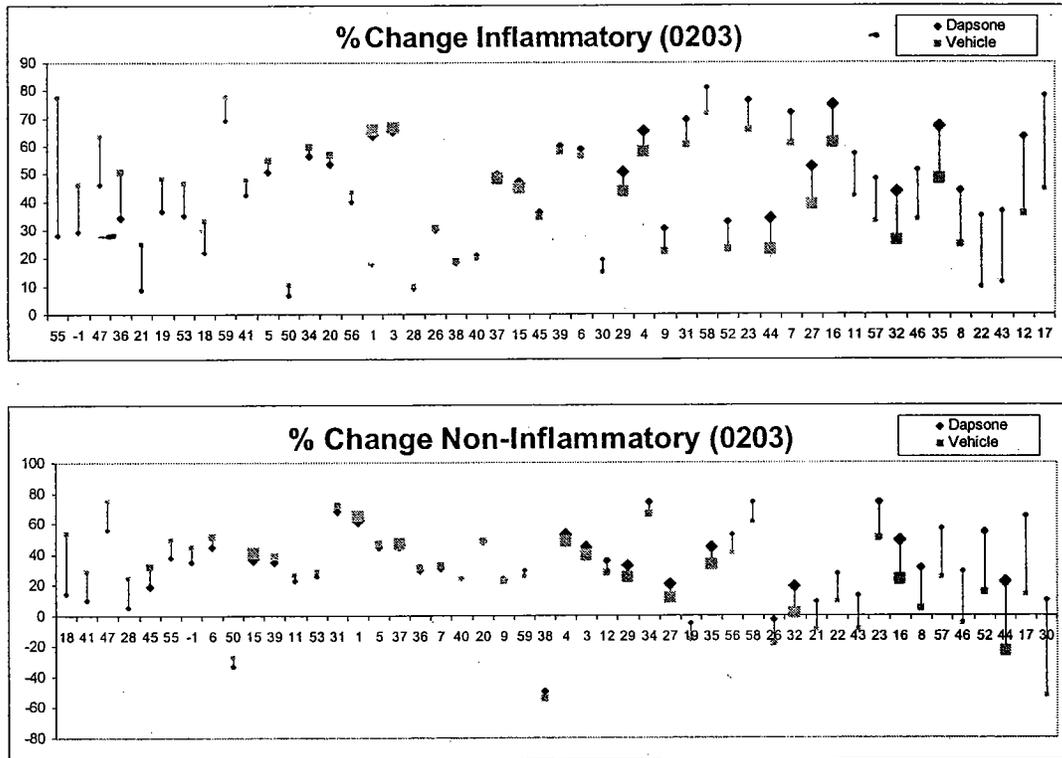
Adverse event rates were similar between the dapson and vehicle arms. Subjects were screened for G-6-PD deficiency. In the 12-month open-label safety study (Study 0114), subjects dapson blood levels were assessed at each visit. Of the 5 G-6-PD deficient subjects in Study 0114, two of the subjects had higher dapson levels (85th to 99th percentile of measurements), but these two subjects only had dapson concentrations through month 3 rather than through month 12.

**APPEARS THIS WAY
ON ORIGINAL**

Appendix

Additional Figures

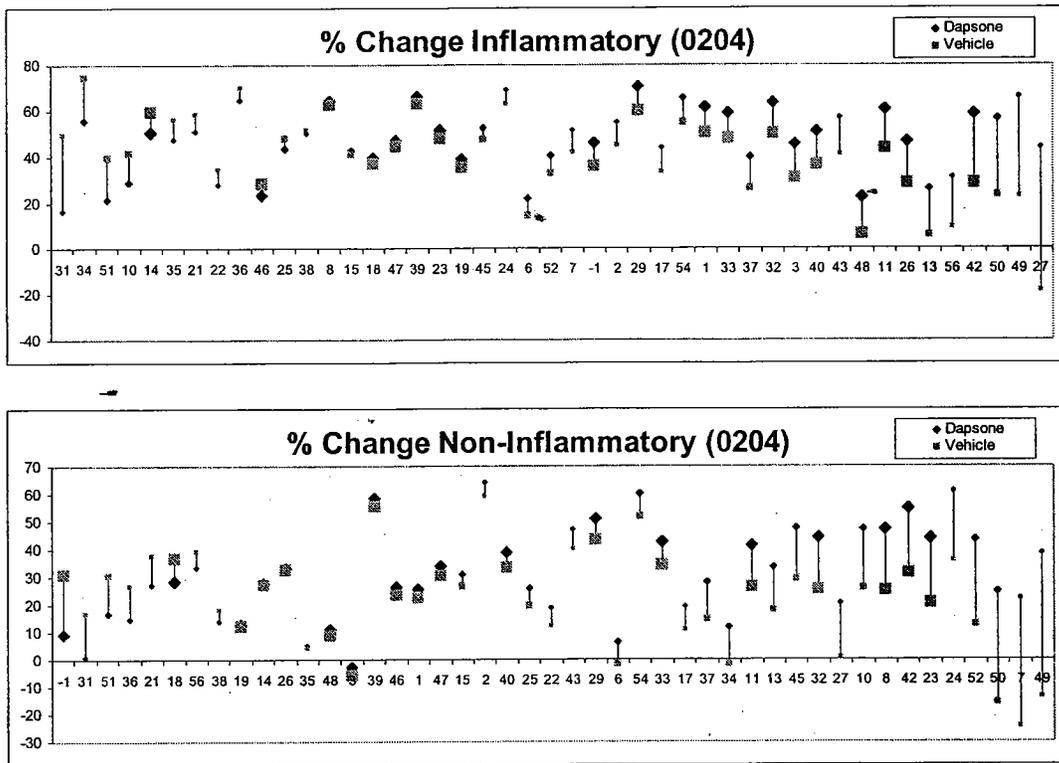
Figure 7 - Percent Change in Inflammatory and Non-Inflammatory Lesions (0203)



Small symbols = 10-25 subjects per center, Medium symbols = 26-39 subjects per center, Large symbols = 40-60 subjects per center
Center -1 represents the pooled results from Centers 10, 24, 33, and 60 (the centers enrolling fewer than 5 subjects per treatment arm).

APPEARS THIS WAY
ON ORIGINAL

Figure 8 - Percent Change in Inflammatory and Non-Inflammatory Lesions (0204)



Small symbols = 10-25 subjects per center, Medium symbols = 26-39 subjects per center, Large symbols = 40-60 subjects per center
 Center -1 represents the pooled results from Centers 4, 9, 12, 16, 28, 44, 53, 55, and 57 (the centers enrolling fewer than 5 subjects per treatment arm).

Signatures/Distribution List

Primary Statistical Reviewer: Kathleen Fritsch, Ph.D.
 Date: May 19, 2005

Statistical Team Leader: Mohamed Alesh, Ph.D.

- cc:
 HFD-540/Wilkin
 HFD-540/Luke
 HFD-540/Vaughan
 HFD-540/Cross
 HFD-700/Anello
 HFD-725/Huque
 FHD-725/Lin
 HFD-725/Alesh
 HFD-725/Fritsch

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

/s/

Kathleen Fritsch
5/19/05 09:57:09 AM
BIOMETRICS

Mohamed Alesh
5/19/05 11:27:10 AM
BIOMETRICS
Concur with review

U.S. 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

CARCINOGENICITY STUDY

NDA Number: 21,794
Drug Name: 5% Dapsone Topical Gel
Indication(s): Topical Treatment of Acne Vulgaris
Applicant: Atrix Laboratories, Inc.
Date(s): Received 9/7/04, user fee (10 months) 7/7/05
Review Priority: Standard

Biometrics Division: Division 3, HFD-725
Statistical Reviewer: Steve Thomson, HFD-725
Concurring Reviewers: Team Leader: Mohamed Alosh, Ph. D., HFD-725
Medical Division: Dermatological and Dental Drug Products, HFD-540
Toxicologist: Reviewer: Norman See, Ph.D., HFD-540
Team Leader: Paul Brown, Ph.D., HFD-540
Project Manager: Frank Cross, Jr., HFD-540

Keywords: Bayesian analysis, Carcinogenicity, Cox regression, Kaplan-Meier product limit, Survival analysis, Trend test

Table of Contents

1. EXECUTIVE SUMMARY	3
1.1. CONCLUSIONS AND RECOMMENDATIONS	3
1.2. BRIEF OVERVIEW OF THE STUDIES	4
1.3. STATISTICAL ISSUES AND FINDINGS	5
1.3.1. <i>Statistical Issues</i>	5
1.3.2. <i>Statistical Findings</i>	7
2. INTRODUCTION	8
2.1. OVERVIEW	8
2.2. DATA SOURCES.....	9
3. STATISTICAL EVALUATION.....	10
3.1. EVALUATION OF EFFICACY.....	10
3.2. EVALUATION OF SAFETY.....	10
3.2.1 <i>Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats</i>	10
3.2.2 <i>Study ATLS-163: 26[33]-Week Dermal Carcinogenicity Study in Tg.AC Mice</i>	13
4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	18
5 SUMMARY AND CONCLUSIONS	18
5.1. STATISTICAL ISSUES AND COLLECTIVE EVIDENCE.....	18
5.2. CONCLUSIONS AND RECOMMENDATIONS	20
APPENDICES:	22
APPENDIX 1. ATLS-123 TWO YEAR RAT STUDY: SURVIVAL ANALYSIS	22
APPENDIX 2. ATLS-123 TWO YEAR RAT STUDY: SURVIVAL TO SPECIFIED TIME.....	24
APPENDIX 3. ATLS-123 TWO YEAR RAT STUDY: BAYESIAN ANALYSIS OF SURVIVAL.....	27
APPENDIX 4. ATLS-123 TWO YEAR RAT STUDY: TUMORIGENICITY ANALYSIS	29
APPENDIX 5. ATLS-163 Tg.AC STUDY: OBSERVED INDIVIDUAL SKIN LESIONS AT END OF STUDY.....	35
APPENDIX 6. ATLS-163 Tg.AC STUDY: SURVIVAL ANALYSIS	37
APPENDIX 7. ATLS-163 Tg.AC STUDY: WEEKLY SOA TUMOR INCIDENCE	41
APPENDIX 8. REFERENCES.....	48
SIGNATURES/DISTRIBUTION LIST	49

1. EXECUTIVE SUMMARY

This submission was intended to assess the carcinogenic potential of daily administration of diethylene glycol monoethyl ether (DGME) and Dapsone administered via oral gavage in a two year study in Sprague-Dawley rats and administered on the skin in a 33-week study in hemizygous Tg.AC transgenic mice.

1.1. Conclusions and Recommendations

The submission involved two studies entitled, respectively:

Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats

Study ATLS-163: 26[33]-Week Dermal Carcinogenicity Study in Tg.AC Mice (Note study actually lasted for 33 weeks.)

Study ATLS-123 involved three Dapsone Gel treatment groups (1, 5, and 15 mg/kg/day nominal dose), a Dapsone vehicle group, and a Diethylene Glycol Monoethyl Ether (DGME) treatment group (540 mg/kg/day). Mortality was high in all treatment groups. Interestingly, for both genders, among the five treatment groups, survival was generally lowest in the vehicle group and highest in the DGME group, though most of the contributions to any apparent differences in survival were due to the high survival in the DGME group (see Appendix 1).

Due to the low survival rates in the vehicle treatment groups, all study treatment groups were terminated early. Females were sacrificed in Week 93, males in Week 100. Histopathology was performed on tissues from animals in the Dapsone high dose (15 mg/kg/day) group, the Dapsone vehicle group, and the DGME treatment group, plus all animals dead or moribund sacrifice and all gross lesions in the low (1 mg/kg/day) and medium (5 mg/kg/day) dose groups. That is, the neoplasms for the low and medium dose groups were not exhaustively analyzed. Thus, for tumorigenicity, the tests of trend over all Dapsone doses were somewhat questionable. Strictly, only the assumptions of the pairwise tests between the high dose and vehicle groups, and the vehicle groups and DGME groups are completely satisfied. With that caveat, systemic hemangiosarcomas and pooled hemangiomas and hemangiosarcomas showed highly statistically significant dose related trends (Hemangiosarcomas Peto test: $p = 0.0037$, Cochran-Armitage test: $p = 0.0036$ and Pooled hemangiomas and hemangiosarcomas Peto test: $p = 0.0164$, Cochran-Armitage test: $p = 0.0786$) and an almost statistically significant difference between the high dose group and the Dapsone control ($p = 0.0750$). Note the Sponsor's submission received 22 April 2005 described these as rare tumors, so following the Haseman-Lin-Rahman rules, discussed below, these tests were statistically significant or close to it. The only other statistically significant dose related trend was in skin papillomas ($p = 0.0018$).

Study ATLS-163 involved three Dapsone Gel treatment groups (3%, 5%, and 10% in 25% DGME), vehicle control #1 (25% DGME), vehicle control #2 (Acetone), 5% Dapsone in Acetone, and two dose groups using tetradecanoyl phorbol acetate (TPA). These latter groups were positive control #1, 20 µg TPA in 25% DGME 3 times/week, and positive control #2, 1.25µg TPA in 0.1mL Acetone 3 times/week. At the recommendation of the Carcinogenicity Assessment Committee, dosing was increased and the study extended to allow for at least 26 weeks of dosing at the higher dose. For both genders, there was an increasing trend in mortality (i.e., decreasing survivability) among the Dapsone treatment groups. As a proof of the model papilloma incidence was high in the two TPA treatment groups. However, among the surviving animals, papilloma incidence was quite low for all non-TPA treatment groups, with no evidence of a dose related trend.

1.2. Brief Overview of the Studies

Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats

Animals were randomized by body weight into five groups of 50 animals per gender. The Sponsor states that the oral (gavage) route of administration was taken "to increase the systemic availability of Dapsone that could not be achieved by topical application of Dapsone topical gel" (page 15). One group was treated with DGME (540 mg/kg), the other four groups were treated with increasing doses of Dapsone Gel, 0 mg/kg - vehicle control, 1 mg/kg, 5 mg/kg, and 10 mg/kg. Due to the low survival rates in the vehicle treatment groups, all study treatment groups were terminated early. Females were sacrificed in Weeks 92 and 93, males in Week 100. Histopathology was performed on tissues from animals in the Dapsone high dose, the Dapsone vehicle group, and the DGME treatment group, plus all animals dead or moribund sacrifice and all gross lesions in the low (1 mg/kg/day) and medium (5 mg/kg/day) dose groups. That is, the neoplasms for the low and medium dose groups were not exhaustively analyzed. The animal phase was initiated on 15 February 2001 and completed 15 January 2003.

Study ATLS-163: 26[33]-Week Dermal Carcinogenicity Study in Tg.AC Mice

Animals were randomized by body weight into seven groups of 25 animals per gender and one group of 10 animals per gender. Four of these groups were treated daily at the site of application (SOA), with increasing doses of Dapsone Gel (0%, i.e., vehicle control #1, 3%, 5%, and 10% Dapsone), each in 25% DGME gel. These defined the four primary Dapsone treatment groups. Animals in the vehicle control group #2 were treated with acetone only, once daily. Animals in another treatment group, consisting of 5% Dapsone in acetone vehicle, were also treated daily. To verify the sensitivity of the animal model, two positive control groups were treated with a known carcinogen, 50 µg tetradecanoyl phorbol acetate (TPA), three times a week. Positive control group #1 was treated with 20 µg TPA in 25% DGME gel, while positive control group #2 (the 10 animal group), was treated with 1.25µg in 0.1mL acetone. The Sponsor indicated that the FDA Carcinogenicity Assessment Committee requested that the dose volume of

the test article and the vehicle control groups be increased from 2 mL/kg to 5 mL/kg. This change was implemented at Week 7 for the 25% DGME groups and at Week 4 for the acetone groups. Further, with the exception of the positive control treatment groups, the study was extended so that animals could be treated continuously at the new dose for at least 26 weeks from that point on, giving 33 weeks in total. Dosing was initiated on May 19 and 21, 2003, for males and females respectively.

1.3. Statistical Issues and Findings

1.3.1. Statistical Issues

Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats

Several issues, typical of such analyses, are considered in the following discussion:

1. Survival Analysis:

Logrank tests were used to test homogeneity of survival among the five treatment groups, including the DGME group. Further, the test of homogeneity among the four Dapsone treatment groups also was conducted, followed by a somewhat more powerful test of trend among these four groups. This involved testing multiple hypotheses, but from the point of view of finding differences among treatment groups, i.e., minimizing Type II error, would have been conservative. Appendices 1-3 review the animal survival.

2. Tests in Neoplasms:

Only the organs of animals in the vehicle group, the DGME group, and the high dose group were exhaustively checked by the pathologist for tumors. That is, animal organs in the low and medium dose groups were not exhaustively inspected. Thus, following the Sponsor, one might argue that tests for trend in tumorigenicity are not generally interpretable, and only the pairwise tests between the high dose and vehicle groups, and the vehicle groups and DGME groups were appropriate. However, generally about 60% or more of the low and medium dose group animals were checked by the pathologist, and this may be sufficient to suggest trends. A problem with the FDA Peto tumorigenicity analysis was that the software used assumes that all animals (50/group) were checked for tumors. A problem with the Sponsor's original statistical analyses of tumors was that neoplasms that are usually analyzed systemically were assigned to specific organs, thus reducing incidence. Further, several neoplasms were assigned to suborgans, potentially masking trends and differences in neoplasms. This was particularly true for hemangiosarcomas.

Finally, due to programming difficulties, the FDA analyzed some of the pooled results using Cochran-Armitage tests of trend and Fisher exact tests instead of Peto tests. These tests were not adjusted for differences in mortality. However, this should have caused no problem, as the only statistically significant evidence of tumorigenicity comes from male rats, and, for male

rats, there was no evidence of differences in mortality. Note that unlike the FDA Peto type analysis, these tests used the actual number of animals analyzed in the low and medium dose groups. In those groups where the organs were not exhaustively analyzed for neoplasms, it is not clear if one should base analysis on all animals at risk (i.e., 50 animals), or just those analyzed. Basing the analysis on all 50 animals would tend to underestimate the true number of neoplasms, since some animals with neoplasms would be missed. However, basing the analysis on animals chosen since they were dead or moribund, might tend to inflate the proportion of animals with apparent neoplasms. In addition, at the request of the FDA, on 22 April 2005, the Sponsor provided Peto mortality adjusted analyses for certain requested combinations of organs and neoplasms. These results are reviewed in Appendix 4.

3. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involved a large number of statistical tests, one for each organ tumor combination. This implies that an adjustment for experiment-wise Type I error is needed to correct for the multiple comparisons involved in such tests. Based on his experience with such analyses, Haseman (1983), proposed a p-value adjustment rule that is applicable to such comparisons. That is, for a roughly 0.10 (10%) overall false positive error rate in tests of differences, rare tumors (with a historical control incidence 1% or below) should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Note that the corresponding rules for trend developed by Lin and Rahman (1996) are that rare tumors should be tested at a 0.025 (2.5%) level and common tumors at a 0.005 (0.5%) level.

Study ATLS-163: 26[33]-Week Dermal Carcinogenicity Study in Tg.AC Mice

1. Survival Analysis:

Logrank tests were used to separately test homogeneity of survival among i) the four Dapsone/DGME treatment groups, and ii) the two acetone groups. Further, there was a test of trend among the four Dapsone dose groups. Note that a test of trend in dose for the two acetone groups would have been superfluous. These results are presented in Appendix 6.

2. Tests in Neoplasms:

According to Dunson, et al. (2000), the current standard for statistical analysis of skin papillomas separately tests for differences between the experimental group and the control group with respect to the following endpoints (in a slightly different order):

- (1) percentage (or count) of animals with tumors, presumably at the end of the study,
- (2) average number of tumors per animal at risk,
- (3) average number of tumors per tumor-bearing animal,
- (4) average latency time to appearance of the first skin tumor,
- (5) average latency time to the appearance of the maximal number of tumors.

In addition, Dunson (2000) proposed a mixed effects Poisson model for the increase in the maximum number of papillomas.

However, as can be seen in Appendices 5 and 7, except for the positive controls, very few animals displayed site of application (SOA) tumors. Since there were so few tumors, it was apparent that latency times would show no statistically significant differences, and the detailed analyses described by Dunson was not attempted. Furthermore, any analysis based only on the number of tumor-bearing animals would utilize too few animals for any conclusions to be drawn. Similarly, there were insufficient tumors to justify any analysis using the Dunson mixed effects Poisson model.

1.3.2. Statistical Findings

Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats

Study ATLS-123 involved three Dapsone Gel treatment groups (1, 5, and 15 mg/kg/day nominal dose), a Dapsone vehicle group, and a DGME treatment group (540 mg/kg/day). Mortality was high in all treatment groups. However, for both genders, among the five treatment groups, survival was lowest in the vehicle group, although any apparent differences were largely due to high survival in the DGME group. There was no consistent pattern of differences in either food consumption or animal weight.

Due to low survival rates in the vehicle treatment groups, all study treatment groups were terminated early. Females were sacrificed in Week 93, males in Week 100. Histopathology was performed on tissues from animals in the Dapsone high dose, the Dapsone vehicle group, and the DGME treatment group, plus all animals dead or moribund sacrifice and all gross lesions in the low (1 mg/kg/day) and medium (5 mg/kg/day) dose groups. That is, the neoplasms for the low and medium dose groups were not exhaustively analyzed by the histopathologist. In male rats systemic hemangiosarcomas and pooled hemangiomas and hemangiosarcomas showed statistically significant dose related trends ($p = 0.0037$ and $p = 0.0164$, respectively) and almost statistically significant differences between the high dose group and the Dapsone control ($p = 0.0750$ for both endpoints). The only other statistically significant dose related trend was in skin papillomas in males ($p = 0.0018$).

Study ATLS-163: 26 [33]-Week Dermal Carcinogenicity Study in Tg.AC Mice

Study ATLS-163 involved three Dapsone Gel treatment groups (3%, 5%, and 10% in 25% DGME), vehicle control #1 (25% DGME), vehicle control #2 (Acetone), 5% Dapsone in Acetone, positive control #1 (20 μ g TPA in 25% DGME 3 times/week), positive control #2 (1.25 μ g TPA in 0.1mL Acetone 3 times/week). For both genders, there was an increasing trend in mortality among the Dapsone treatment groups. Papilloma incidence was high in the two TPA treatment groups.

However, among the surviving animals, papilloma incidence was quite low for all non-TPA treatment groups, with no evidence of a Dapsone Gel dose related trend or differences.

2. INTRODUCTION

2.1. Overview

Results from two studies (Study ATLS-123 and Study ATLS-163), were submitted to assess the carcinogenic potential of Dapsone in two rodent species.

Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats

For each gender, animals were randomized by body weight into five groups of 50 animals. The Sponsor states that the oral (gavage) route of administration was taken "to increase the systemic availability of Dapsone that could not be achieved by topical application of Dapsone topical gel" (study report page 15). One group was treated with Diethylene Glycol Monoethyl Ether (DGME) 540 mg/kg, the other four groups were treated with increasing doses of Dapsone Gel, 0 mg/kg, i.e., vehicle control; 1 mg/kg; 5 mg/kg; and, finally, 15 mg/kg per day. For this study the vehicle used for the Dapsone groups was 0.5% carboxymethyl-cellulose (medium viscosity) in sterile water. Dosing was initiated on May 19 and 21, 2003, for males and females respectively.

Table 1. ATLS-123 Dosage in 104-week Gavage Carcinogenicity Study in Sprague-Dawley Rats

Group No.	Treatment	Dose (mg/kg)	# Males	# Females
0	Vehicle Control	0	50	50
1	DGME	540	50	50
2	Low Dose Dapsone	1	50	50
3	Medium Dose Dapsone	5	50	50
4	High Dose Dapsone	15	50	50
Total			250	250

Histopathology was performed on tissues from animals in the Dapsone high dose, the Dapsone vehicle group, and the DGME group, plus all animals found dead or moribund (and sacrificed), plus all gross lesions in the low and medium dose groups. Since the low dose and the medium dose groups were not exhaustively analyzed, tests of trend over these doses are problematical.

Study ATLS-163: 26[33]-Week Dermal Carcinogenicity Study in Tg.AC Mice

Animals were randomized by body weight into seven groups of 25 animals per gender and one group of ten animals per gender. Table 2 below summarizes the eight treatment groups for this study. Four of these groups were treated daily at the site of application (SOA), with

increasing doses of Dapsone Gel (0%, i.e., Vehicle Control #1; 3%; 5%; and 10% Dapsone) each in 25% DGME gel. These defined the four primary Dapsone treatment groups. Animals in the Vehicle Control Group #2 were treated with acetone only, once daily. Another treatment group (Group #8), consisting of 5% Dapsone in acetone vehicle, were also treated daily. To verify the sensitivity of the animal model, two positive control groups were treated with a known carcinogen, 50 µg tetradecanoyl phorbol acetate (TPA), three times a week. Positive Control Group #1 was treated with 20 µg TPA in 25% DGME gel, while Positive Control Group #2 (the 10 animal group), was treated with 1.25µg in 0.1mL acetone.

The Sponsor indicated that the FDA Carcinogenicity Assessment Committee requested that dose volume of the test article and the vehicle control groups be increased from 2 mL/kg to 5 mL/kg. This recommendation was not received until after initiation of the study and was implemented at Week 7 for the 25% DGME groups and at Week 4 for the acetone groups. Further, with the exception of the positive control treatment groups, the study was extended so that animals could be treated continuously at the new dose for at least 26 weeks, giving 33 weeks in total. Dosing was initiated on May 19 and 21, 2003, for males and females respectively.

Table 2. ATLS-163 Dosage in 26[33]-week Dermal Carcinogenicity Study in Tg.AC Mice`

Group No.	Treatment	# Males	# Females
1	Vehicle Control, 25% DGME	25	25
2	Positive Control #1, 20µg TPA in 25% DGME	25	25
3	3% Dapsone in 25% DGME gel	25	25
4	5% Dapsone in 25% DGME gel	25	25
5	10% Dapsone in 25% DGME gel	25	25
6	Acetone, Vehicle Control	25	25
7	Positive Control #2, 1.25µg TPA in Acetone	10	10
8	5% Dapsone in Acetone	25	25
Total		185	185

The Sponsor reported that animals were observed twice daily for moribundity and mortality, and once a week for clinical signs of toxicity. A detailed examination of the site of application (SOA) was performed weekly. A skin tumor was designated as "latent" after attaining a size of 2 mm in diameter and protruded from the surface of the skin. A skin tumor was designated as "actual" if it was observed for three consecutive weeks. Animals in the two positive control groups above sacrificed when 20 or more SOA tumors were observed.

2.2. Data Sources

Tumorigenicity data for studies ATLS-123 and ATLS-163 were in SAS transport data sets, in the FDA electronic data room directory:

\\Cdsesub1\n50803\N_000\2004-08-23\pharmtox\d datasets\aa81ew.7d8t.btl

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

NA

3.2. Evaluation of Safety

3.2.1 Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats

Study ATLS-123 involved three Dapsone Gel treatment groups (1, 5, and 15 mg/kg/day nominal dose), a Dapsone vehicle group, and a DGME treatment group (540 mg/kg/day). Mortality was high in all treatment groups. However, for both genders, among the five treatment groups, through most of the study survival was lowest in the vehicle group, and highest in the DGME group, particularly for males (see Appendix 1). Data sets were not provided for a detailed analysis of food consumption or animal weights, but it was clear from the Sponsor's figures and tables that for both genders, over the course of the study, there were no statistically significant, consistent results between treatment groups for these endpoints. The following table is a summary description of these treatment groups:

Table 3. ATLS-123: Dose Groups for 104-week Carcinogenicity Study in Sprague-Dawley Rats

Group No.	Treatment	Dose (mg/kg)	# animals /gender
0	Vehicle Control	0	50
1	DGME	540	50
2	Low Dose Dapsone	1	50
3	Medium Dose Dapsone	5	50
4	High Dose Dapsone	15	50

Because of the relatively low survival rates in the vehicle treatment groups, all study treatment groups were terminated early. Females were sacrificed by Week 93, males in Week 100. Appendix 2 has several analyses comparing the percentage of animals that survived to these endpoints in the current study to survival percentages of control groups from a similar strain of rats. All these analyses indicated statistically significant difference in survival between the control group in the current study and the nominally similar control groups in other recent studies.

Histopathology was performed on tissues from animals in the Dapsone high dose, the Dapsone vehicle group, and the DGME treatment group, plus all animals dead or moribund sacrifice and all gross lesions in the low (1 mg/kg/day) and medium (5 mg/kg/day) dose groups.

3.2.1.1 Sponsor's Results and Conclusions

The following table displays, for each treatment group, the number of natural deaths or moribund sacrifices, the number of animals at risk at the beginning of each interval, and the Kaplan-Meier Product Limit estimate of the proportion of survivors at the end of the interval. Note these totals do not include accidental deaths, which were treated as censored. In addition the statistical significance levels of different tests comparing survival curves are presented. Under the entry labeled "vehicle," the p-values for the tests of homogeneity and trend (labeled "Homog" and "Trend"), across the four Dapsone treatment groups are presented. The other p-values listed below correspond to tests of differences in survival between the vehicle and the DGME, Dapsone 1 mg/kg, 5 mg/kg, and 15 mg/kg treatment groups, respectively.

Table 4. ATLS-123: Kaplan-Meier Estimates: Deaths / At Risk (KM)

Gender	Weeks	Dapsone				
		Vehicle	DGME	1 mg/kg	5 mg/kg	15 mg/kg
M	0-52	8/50 (84%)	4/50 (92%)	10/50 (80%)	5/50 (90%)	8/50 (84%)
	53-78	16/42 (52%)	6/46 (80%)	9/40 (62%)	16/45 (58%)	12/41 (59%)
	79-92	9/25 (33%)	10/38 (59%)	10/30 (41%)	6/29 (46%)	10/29 (39%)
	93-EOS	1/16 (31%)	4/28 (50%)	3/20 (35%)	3/23 (40%)	5/19 (29%)
	Terminal	14	24	17	19	14
	p-value	Homog 0.0499	0.0036	0.5038	0.1965	0.6001
	Trend	0.7090				
F	0-52	5/50 (90%)	1/50 (98%)	3/50 (94%)	3/50 (94%)	2/50 (96%)
	53-78	16/45 (57%)	13/49 (72%)	11/47 (72%)	16/47 (62%)	19/48 (58%)
	79-EOS	15/28 (27%)	9/36 (54%)	15/36 (42%)	10/31 (42%)	7/29 (44%)
	Terminal	13	27	21	21	22
	p-value	Homog 0.0918	NA	NA	NA	NA
		Trend	0.1393			

"Among males, there was a significant difference in the homogeneity of survival rates across groups. In a pairwise comparison, there was a significant difference between the vehicle control and DGME groups, that is, survival was increased in the DGME group. There was no significant dose-response trend in survival rates in Dapsone treated males. Among females, there were no significant survival findings related to DGME or Dapsone-treated groups." (final report page 20)

For the analysis of tumors, the Sponsor reported that: "None of the microscopic findings were statistically significant by the Peto analysis, and the lack of statistical significance was supported by the histopathology conclusions." (final report page 20) This conclusion was not completely consistent with this reviewer's conclusions nor with the results submitted in response to the FDA analysis request.

3.2.1.2 FDA Reviewer's Results

Survival analysis:

Kaplan-Meier plots comparing treatment groups are presented in Appendix 1. Estimated quartiles from the Kaplan-Meier distributions and mean times to death among the uncensored times are as given in the following table:

Table 5. ATLS-123: Percentiles from Kaplan-Meier Estimates of Survival

Males	1 - Vehicle	2 - DGME	3-Dapsone 1 mg/kg	4-Dapsone 5 mg/kg	5-Dapsone 15 mg/kg
P25*	404	586	388	469	452
P50*	559	---	615	617.5	590
Mean	516.5	606.3	543.3	572.0	544.4
P75*	---	---	---	---	---

Females	1 - Vehicle	2 - DGME	3-Dapsone 1 mg/kg	4-Dapsone 5 mg/kg	5-Dapsone 15 mg/kg
P25*	460	539	532	490	495
P50*	572	---	632	609	616
Mean	525.0	575.5	575.1	549.8	560.4
P75*	---	---	---	---	---

* P25, P50, and P75 denote the 25th percentile (1st quartile), the median, and the 75th percentile (the 3rd quartile), respectively.

Logrank tests were used to test homogeneity of survival among the five treatment groups, including the DGME group (Group 2). Furthermore, the test of homogeneity among the four Dapsone treatment groups with vehicle (Groups 1, 3-5) also was conducted, followed by a somewhat more powerful test of trend among these four groups (using a partial likelihood model). The significance levels of these tests are given in Table 6, below:

Table 6. Results of tests of homogeneity of treatment groups

Test	Test	Males	Females
Groups 1-5	Log-Rank	0.0860	0.0795
	Wilcoxon	0.0499	0.0918
Groups 1, 3-5	Log-Rank	0.6337	0.1926
	Wilcoxon	0.5979	0.1961
	Trend over dose	0.7095	0.1424

This involves five different tests per gender, and thus from the point of view of noting differences should be conservative. Note there are some slight discrepancies between the significance levels reported in the Sponsor's analysis and the FDA analysis, however, none of these discrepancies have any effect on conclusions.

As with the Sponsor's analysis, note that after deleting the DGME group, there was no particularly strong evidence of a statistically significant difference in survival among the Dapsone groups. Strictly speaking, lack of evidence of heterogeneity in survival should not be treated as proof of homogeneity in survival. However, it does seem indicative of homogeneity, particularly for males.

Tumorigenicity analysis:

Appendix 4 shows the results of Peto mortality adjusted tests of trend over the four Dapsone groups, a test of differences between the high dose Dapsone group and the Dapsone control, and a test of differences between the Dapsone and DGME controls. Both exact (i.e., permutation) and asymptotic tests are provided. Due to programming difficulties, several of the FDA analyses of the pooled organs use the Cochran-Armitage test of trend and Fisher exact tests instead of Peto tests. That is, the tests are not adjusted for differences in mortality. While in principle this is a weakness, in practice it should cause no problem as there is no strong evidence of differences in mortality among the Dapsone groups. However, in response to a later FDA request, the Sponsor sent Peto adjusted tests for the requested combinations of organs and neoplasms. These results are also included.

The only statistically significant trends were with respect to skin papillomas, hemangiosarcomas, and pooled hemangiosarcomas/hemangiomas in male rats (trend: $p = 0.0018$, $p = 0.0037$, and $p = 0.0145$, respectively). Note that whether one classifies the first two listed neoplasms as rare tumors (as suggested by the incidence in the control group) or as common tumors, there is statistically significant evidence of a trend in tumorigenicity in the first two neoplasms. If one considers hemangiosarcomas as rare tumors, the difference between the high dose group and Dapsone vehicle is almost statistically significant ($p = 0.0587$).

3.2.2 Study ATLS-163: 26[33]-Week Dermal Carcinogenicity Study in Tg.AC Mice

According to the Sponsor: "Three groups of twenty five Tg.AC mice/sex (Groups 3-5) were treated with the test article, Dapsone (in 25% DGME), by daily dermal application at dose levels of 3%, 5% and 10%, which provided dose levels of approximately 150, 250, and 500 mg/kg/day, respectively. A positive control group (Group 2) received 20 μ g Tetradeconyl phorbol acetate (TPA) in 0.1 mL 25% DGME gel three times per week. A group consisting of the same number and strain of animals (Group 1) was treated daily with 25% DGME and served as a DGME vehicle control group. In addition, one group of 25 Tg.AC mice/sex (Group 8) was treated daily with the test article, Dapsone, at a concentration of 5% delivered in acetone. A second positive control group (Group 7), consisting of ten Tg.AC mice/sex, received 1.25 μ g TPA in 0.1 mL acetone three times per week. A group consisting of 25 Tg.AC mice/sex (Group 6) was treated daily with acetone alone and served as the acetone vehicle control group. At the recommendation of the Carcinogenicity Assessment Committee (FDA), the dose volume of the test article and vehicle control groups was increased from 2 mL/kg to 5 mL/kg. This change was implemented at Week 7 for the 25% DGME groups and at Week 4 for the acetone groups.

Additionally, with the exception of the positive control treatment groups, the study was extended in order for the animals to be treated continuously at the new dose volume for at least 26 weeks. All animals were dosed based on the most recent mean animal body weight for each sex and group." (final report, page 10)

The following table reiterates information on the treatment groups:

Table 7. ATLS-163: Dose Groups for 26[33]-week Carcinogenicity Study in Tg.AC mice

Group	# animals/ gender	Dose	Group	# animals/ gender	Dose
1	25	25% DGME vehicle	5	25	10% Dapsone / 25% DGME
2	25	20µg TPA / 25% DGME	6	25	Acetone
3	25	3% Dapsone / 25% DGME	7	10	1.25µg TPA / Acetone
4	25	5% Dapsone / 25% DGME	8	25	5% Dapsone / Acetone

Note that the positive controls, Groups 2 and 7, served as qualitative and quantitative indicators of the test system's response to a known tumor promoter. Animals were approximately 7-8 weeks old at the start of dosing. Dosing of the 25% DGME/test vehicle groups was initiated on June 10 and 11, for males and females, respectively. Dosing of the three Acetone/test groups was initiated on May 19 and 21, for males and females, respectively, totaling 31-33 weeks of treatment. All groups, except the two positive control groups, received a dose volume of 2 mL/kg for Weeks 1-6 (Groups 1 and 3-5) and Weeks 1-3 (Groups 6 and 8). The Sponsor reported that at the recommendation of the Carcinogenicity Assessment Committee (FDA) for the remainder of the study these doses increased to 5 mL/kg/day.

The site of each observed or latent tumor was recorded at each weekly evaluation. The Sponsor reported that skin papillomas were not counted as actual papillomas until they had been observed for three consecutive weeks. Tumor incidence was defined as the number of animals with tumor. Latent tumors and tumors that later disappeared were accumulated separately. Prior to the end of the study, the Sponsor's definition of tumor burden was the total of these actual site of application (SOA) tumors and the site of application tumors that later disappeared (DSOA). The definition of the tumor burden at the end of the study also included the latent papillomas (LSOA). Non site of application (NSOA) tumors were accumulated similarly, but were not included in the definition of tumor burden. Note that the definition of tumor burden given in Dunson et al. (2000) does not include the disappeared tumors and apparently not the latent tumors.

The protocol specified that all positive control group animals were to be sacrificed when the tumor burden reached 20 or more. Several animals were sacrificed before reaching this level and several were sacrificed after reaching this level. Since the positive controls were used only to demonstrate the sensitivity of the mouse model, this deviation was not considered to be serious. At Week 20, October 24, 2003, 17 Group 6, acetone vehicle, animals were inadvertently treated with 5% Dapsone in acetone. These animals were washed and re-treated with the correct treatment. The Sponsor claims that this should not affect the integrity of the study.

3.2.2.1 Sponsor's Results and Conclusions

Tables 2 and 3 below summarize the Sponsor's statistical analyses of the final survival and final tumor incidence in male and female Tg.AC mice, respectively. For each treatment group, the Sponsor tested tumor incidence against the appropriate DGME or acetone vehicle.

Table 8. ATLS-163 Mortality

Group Gender	1- DGME	2-TPA / DGME	3-3%Dap. /DGME	4-5%Dap /DGME	5-10% Dap /DGME	6-Acetone	7-TPA / Acetone	8-5% Dap /Acetone
Males	1/25	1/25	3/25	7/25*	25/25	1/25	0/10	10/25
Females	4/25	2/25	6/25	25/25	25/25	3/25	2/10	22/25

* Note the FDA analysis found 4 natural or moribund sacrifices among the male groups 4 animals.

The Sponsor stated that : "Statistical analysis (Fisher's Exact Test) of mortality data in the male mice revealed statistically significant increases in Groups 4 and 5 (5 and 10% Dapsone in 25% DGME) when compared to the 25% DGME vehicle control group (Group 1) . A statistically significant increase in mortality was noted in Group 8 (5% Dapsone in acetone) when compared to the acetone vehicle control group (Group 6). Similar statistical analysis results were noted in the female animals." (Final report, page 28) Note that except for the Group 4 Males, the FDA analysis agrees with the Sponsor's results.

Table 9. ATLS-163 Summary of SOA Tumors

Group Gender	1- DGME	2-TPA / DGME	3-3%Dap. /DGME	4-5%Dap /DGME	5-10% Dap /DGME	6-Acetone	7-TPA / Acetone	8-5% Dap /Acetone
Males	1	24	0	1	0	1	10	0
Females	1	24	2	1	0	0	9	0

According to the Sponsor: "Statistical analysis (Fisher's Exact Test) of the number of animals bearing latent or actual or either tumors at the final scoring did not reveal any statistically significant (p<0.05) differences in any of the test article treatment groups when compared to the vehicle control groups (Groups 1 and 6) . The positive control groups (groups 2 and 7) were both statistically significantly (p=0.0000) different in both sexes when compared to the vehicle control groups." (Final Report, page 33)

3.2.2.2 FDA Reviewer's Results

Survival analysis:

Tables 10 (Males) and 11 (Females) below summarize overall mortality for the different dosages (Groups 1-6). Kaplan-Meier plots of the survival distributions among the eight treatment groups are given in Appendix 6 for both genders.

Table 10. ATLS-163 Males: Percentiles from Kaplan-Meier Estimates of Survival

	1-DGME Vehicle	2 - TPA / DGME	3-3% Daps. / DGME	4-5% Daps. / DGME	5-10% Daps. / DGME	6-Acetone Vehicle	7-TPA / Acetone	8-Daps./ Acetone
P25	225	79	225	217	169	211	119	191
P50	225	87	225	225	180	211	133.5	211
Mean	224.9	98.2	218.7	200.8	180.7	210.6	145.9	196.7
P75	225	103	225	225	200	211	183	211

For males the hypotheses of homogeneity in survival over the four DGME/Dapsone groups (Groups 1, 3-5) was rejected (both logrank and Wilcoxon $p < 0.0001$). The test of linear trend and non-linearity in trend were also statistically highly significant (both $p < 0.0001$). The hypotheses of homogeneity in survival over the two Acetone/Dapsone groups (Groups 6 & 8) was also rejected with high statistical significance ($p < 0.0019$).

Table 11. ATLS-163 Females: Percentiles from Kaplan-Meier Estimates of Survival

	1-DGME Vehicle	2 - TPA / DGME	3-3% Daps. / DGME	4-5% Daps. / DGME	5-10% Daps. / DGME	6-Acetone Vehicle	7-TPA / Acetone	8-Daps./ Acetone
P25	230	94	230	164	137	211	129	146
P50	230	101	230	177	146	211	129	152
Mean	211.3	114.5	220.4	169.6	143.7	194.8	133.5	155.3
P75	230	122	230	185	158	211	155	156

For females, tests of the hypotheses of homogeneity in survival over the four DGME/Dapsone groups (Groups 1, 3-5) was rejected, as was the hypotheses of homogeneity in survival over the two Acetone/Dapsone groups (Groups 6 & 8), (all $p < 0.0001$).

These results are summarized in Table 12 below:

Table 12. Results of tests of homogeneity in survival of the treatment groups

Test	Test	Males	Females
Groups 1, 3-5	Log-Rank	< 0.0001	< 0.0001
	Wilcoxon	< 0.0001	< 0.0001
	Trend over dose	< 0.0001	< 0.0001
	Nonlinearity	< 0.0001	< 0.0001
Groups 6 & 8	Log-Rank	0.0019	< 0.0001
	Wilcoxon	0.0018	< 0.0001

Tumorigenicity analysis:

Results for survival of the animals are presented first, followed by analyses of the following endpoints assessing papilloma incidence and burden:

- (1) percentage (or count) of animals with tumors, at the end of the study
- (2) average number of tumors per animal at risk

The following tables display the number of animals alive or dead (natural or moribund sacrifice) at the end of the study or at time of final sacrifice, the number of animals with site of application (SOA) papillomas, and the number of animals with non site of application papillomas (NSOA). Tumor incidence, as recorded below, was defined as the number of animals with any of these tumors. The mean number of SOA and NSOA papillomas is defined as the mean number of papillomas in those animals with tumors, so animals with no tumors do not contribute to the denominator when computing the mean number of tumors. (see Final Report, page 20)

Table 13. ATLS-163 Males: Summary of Papillomas at the End of Study

Group Statistic	1- DGME	2-TPA/ DGME	3-3%Dap. /DGME	4-5%Dap /DGME	5-10% Dap /DGME	6-Acetone	7-TPA / Acetone	8-5% Dap /Acetone
# alive	24	0	22	20	0	24	0	15
# dead	1	1	3	5	25	1	0	10
# sacrificed	0	24	0	0	0	0	10	0
# w/ SOA tumors	1	24	0	1	0	1	10	0
# w/NSOA tumors	2	13	0	0	0	3	3	3
Mean # SOA tumors	1.0	24.8	0	1.0	0.	2.0	22.2	0
Mean NSOA tumors	1.0	2.3	0	0	0.	1.3	2.7	4.3

Table 14. ATLS-163 Females: Summary of Papillomas at the End of Study

Group Statistic	1- DGME	2-TPA/ DGME	3-3%Dap. /DGME	4-5%Dap /DGME	5-10% Dap /DGME	6-Acetone	7-TPA / Acetone	8-5% Dap /Acetone
# alive	21	0	19	0	0	22	0	3
# dead	4	2	6	25	25	3	2	22
# sacrificed	0	23	0	0	0	0	8	0
# w/ SOA tumors	1	24	2	1	0	0	9	0
# w/NSOA tumors	6	10	8	1	3	2	2	0
Mean # SOA tumors	1.0	20.3	1.5	2.0	0.	0	18.6	0
Mean NSOA tumors	1.0	1.7	1.5	1.0	1.3.	1.0	1.5	0

In both tables note the small number of animals with SOA papillomas and the small number of papillomas in all non-TPA treatment groups. It is clear that within the four Dapsone/DGME and the two Dapsone/Acetone treatment groups there are no dose related trends or differences in neoplasms (e.g., all pairwise comparisons have $p = 0.2449$).

Note that the Sponsor reported that skin papillomas were not counted as actual papillomas until they had been observed for three consecutive weeks. Latent tumors and tumors that later disappeared were also included. (see Final Report, page 20)

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

NA

5 SUMMARY AND CONCLUSIONS

5.1. Statistical Issues and Collective Evidence

Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats

There are several issues typical of such analyses:

1. Logrank tests were used to test homogeneity of survival among the five treatment groups, including the DGME group. Further, the test of homogeneity among the four Dapsone treatment groups also was conducted, followed by a somewhat more powerful test of trend among these four groups. This involved testing multiple hypotheses, but from the point of view of finding differences among treatment groups, i.e., minimizing Type II error, would have been conservative. Appendices 1-3 review the animal survival.

2. Tests in Neoplasms:

Only the organs of animals in the vehicle group, the DGME group, and the high dose group were exhaustively checked by the pathologist for tumors. That is, animal organs in the low and medium dose groups were not exhaustively inspected. Thus, following the Sponsor, one might argue that tests for trend in tumorigenicity are not generally interpretable, and only the pairwise tests between the high dose and vehicle groups, and the vehicle groups and DGME groups were appropriate. However, generally about 60% or more of the low and medium dose group animals were checked by the pathologist, and this may be sufficient to suggest trends. A problem with the FDA Peto tumorigenicity analysis was that the software used assumes that all animals (50/group) were checked for tumors. A problem with the Sponsor's original statistical analyses of tumors was that neoplasms that are usually analyzed systemically were assigned to specific organs, thus reducing incidence. Further, several neoplasms were assigned to suborgans, potentially masking trends and differences in neoplasms. This was particularly true for hemangiosarcomas.

Finally, due to programming difficulties, the FDA analyzed some of the pooled results using Cochran-Armitage tests of trend and Fisher exact tests instead of Peto tests. These tests were not adjusted for differences in mortality. However, this should have caused no problem, as the only statistically significant evidence of tumorigenicity comes from male rats, and, for male rats, there was no evidence of differences in mortality. Note that unlike the FDA Peto type

analysis, these tests used the actual number of animals analyzed in the low and medium dose groups. In those groups where the organs were not exhaustively analyzed for neoplasms, it is not clear if one should base analysis on all animals at risk (i.e., 50 animals), or just those analyzed. Basing the analysis on all 50 animals would tend to underestimate the true number of neoplasms, since some animals with neoplasms would be missed. However, basing the analysis on animals chosen since they were dead or moribund, might tend to inflate the proportion of animals with apparent neoplasms. In addition, at the request of the FDA, on 22 April 2005, the Sponsor provided Peto mortality adjusted analyses for certain requested combinations of organs and neoplasms. These results are reviewed in Appendix 4.

3. Multiplicity of Tests on Neoplasms:

Testing the various neoplasms involved a large number of statistical tests, one for each organ tumor combination. This implies that an adjustment for experiment-wise Type I error is needed to correct for the multiple comparisons involved in such tests. Based on his experience with such analyses, Haseman (1983), proposed a p-value adjustment rule that is applicable to such comparisons. That is, for a roughly 0.10 (10%) overall false positive error rate in tests of differences, rare tumors (with a historical control incidence 1% or below) should be tested at a 0.05 (5%) level, and common tumors (with a historical control incidence greater than 1%) at a 0.01 level. Note that the corresponding rules for trend developed by Lin and Rahman (1996) are that rare tumors should be tested at a 0.025 (2.5%) level and common tumors at a 0.005 (0.5%) level.

Study ATLS-163: 26[33]-Week Dermal Carcinogenicity Study in Tg.AC Mice

1. Survival Analysis:

Logrank tests were used to test homogeneity of survival among i) the four Dapsone/DGME treatment groups, and ii) the two acetone groups. Further, there was a test of trend among the four Dapsone dose groups. Note that a test of trend for the two acetone groups would have been superfluous.

2. Tests in Neoplasms:

According to Dunson, et al. (2000), the current standard for statistical analysis of skin papillomas separately tests for differences between the experimental group and the control group with respect to the following endpoints (in a slightly different order):

- (1) percentage (or count) of animals with tumors, presumably at the end of the study,
- (2) average number of tumors per animal at risk,
- (3) average number of tumors per tumor-bearing animal,
- (4) average latency time to appearance of the first skin tumor,
- (5) average latency time to the appearance of the maximal number of tumors.

However, as can be seen in Appendices 3 and 5, except for the positive controls, very few treatment groups displayed site of application (SOA) tumors. Thus, since there are so few tumors, it is apparent that latency times would show no statistically significant differences, and no detailed modeling was attempted. Furthermore, an analysis based only on the number of tumor-bearing animals would utilize too few animals for any conclusions to be drawn. Similarly, there are insufficient tumors for any analysis using the Dunson et al (2000) Poisson mixed effects model.

5.2. Conclusions and Recommendations

The submission involved two studies entitled, respectively:

Study ATLS-123: 104-Week Carcinogenicity Study of Diethylene Glycol Monoethyl Ether and Dapsone Administered via Oral Gavage to Sprague-Dawley Rats

Study ATLS-163: 26[33]-Week Dermal Carcinogenicity Study in Tg.AC Mice (Note study actually lasted for 33 weeks)

Study ATLS-123 involved three Dapsone Gel treatment groups (1, 5, and 15 mg/kg/day nominal dose), a Dapsone vehicle group, and a Diethylene Glycol Monoethyl Ether (DGME) treatment group (540 mg/kg/day). Mortality was high in all treatment groups. Interestingly, for both genders, among the five treatment groups, survival was generally lowest in the vehicle group and highest in the DGME group, though most of the contributions to any apparent differences in survival are due to the high survival in the DGME group (see Appendix 1).

Due to the low survival rates in the vehicle treatment groups, all study treatment groups were terminated early. Females were sacrificed in Week 93, males in Week 100. Histopathology was performed on tissues from animals in the Dapsone high dose (15 mg/kg/day) group, the Dapsone vehicle group, and the DGME treatment group, plus all animals dead or moribund sacrifice and all gross lesions in the low (1 mg/kg/day) and medium (5 mg/kg/day) dose groups. That is, the neoplasms for the low and medium dose groups were not exhaustively analyzed. Thus, for tumorigenicity, the tests of trend over all Dapsone doses are somewhat questionable. Strictly, only the assumptions of the pairwise tests between the high dose and vehicle groups, and the vehicle groups and DGME groups are completely satisfied. With that caveat, systemic hemangiosarcomas and pooled hemangiomas and hemangiosarcomas showed highly statistically significant dose related trends (Hemangiosarcomas Peto test: $p = 0.0037$, Cochran-Armitage test: $p = 0.0036$ and Pooled hemangiomas and hemangiosarcomas Peto test: $p = 0.0164$, Cochran-Armitage test: $p = 0.0786$) and an almost statistically significant difference between the high dose group and the Dapsone control ($p = 0.0750$). Note the Sponsor's submission received 22 April 2005 did describe these as rare tumors, so following the Haseman-Lin-Rahman rules, discussed below, these tests were statistically significant or close to it. The only other statistically significant dose related trend was in skin papillomas ($p = 0.0018$).

Study ATLS-163 involved three Dapsone Gel treatment groups (3%, 5%, and 10% in 25% DGME), vehicle control #1 (25% DGME), vehicle control #2 (Acetone), 5% Dapsone in

Acetone, and two dose groups using tetradecanoyl phorbol acetate (TPA). These latter groups were positive control #1, 20 µg TPA in 25% DGME 3 times/week, and positive control #2, 1.25µg TPA in 0.1mL Acetone 3 times/week. At the recommendation of the Carcinogenicity Assessment Committee, dosing was increased and the study extended to allow for at least 26 weeks of dosing at the higher dose. For both genders, there was an increasing trend in mortality (i.e., decreasing survivability) among the Dapsone treatment groups. As a proof of the model, papilloma incidence was high in the two TPA treatment groups. However, among the surviving animals, papilloma incidence was quite low for all non-TPA treatment groups, with no evidence of a dose related trend.

**APPEARS THIS WAY
ON ORIGINAL**

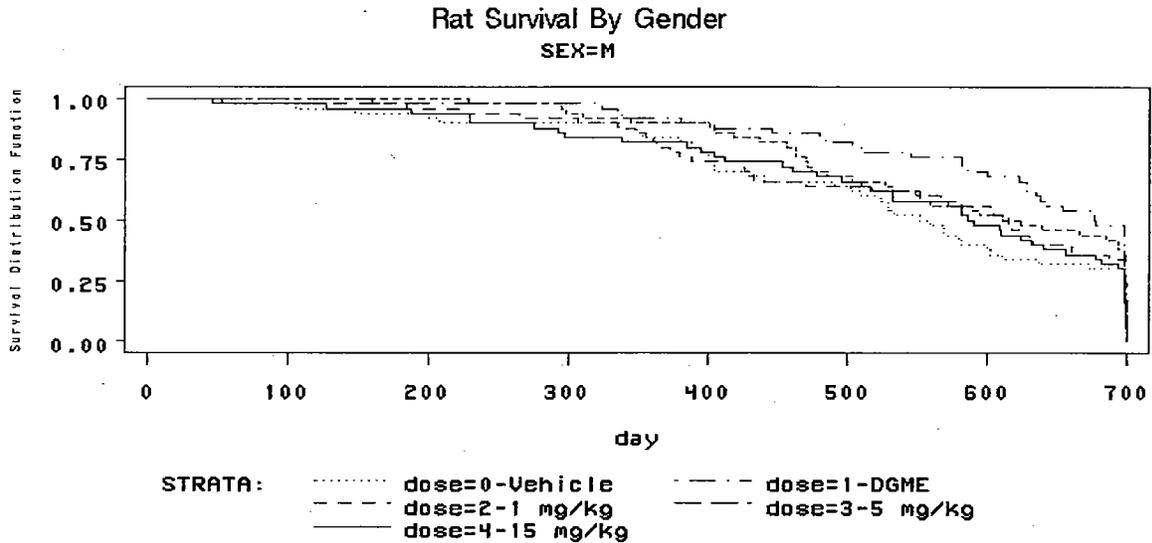
APPENDICES:

Appendix 1. ATLS-123 Two Year Rat Study: Survival Analysis

The plots below display the overall Kaplan-Meier curves for each gender among the four Dapsone groups and the single DGME treatment group. For males the test of homogeneity in survival over the five treatment groups was close to statistically significant or just barely statistically significant (logrank $p = 0.0860$ and Wilcoxon $p = 0.0499$). However, among the four Dapsone groups the test of homogeneity in survival was not statistically significant (logrank $p = 0.6337$ and Wilcoxon $p = 0.5979$). So differences are due to the discrepant DGME vehicle group. The test of trend in these four groups also was not statistically significant ($p = 0.7095$).

Figure A.1.1 Male Sprague-Dawley Rats

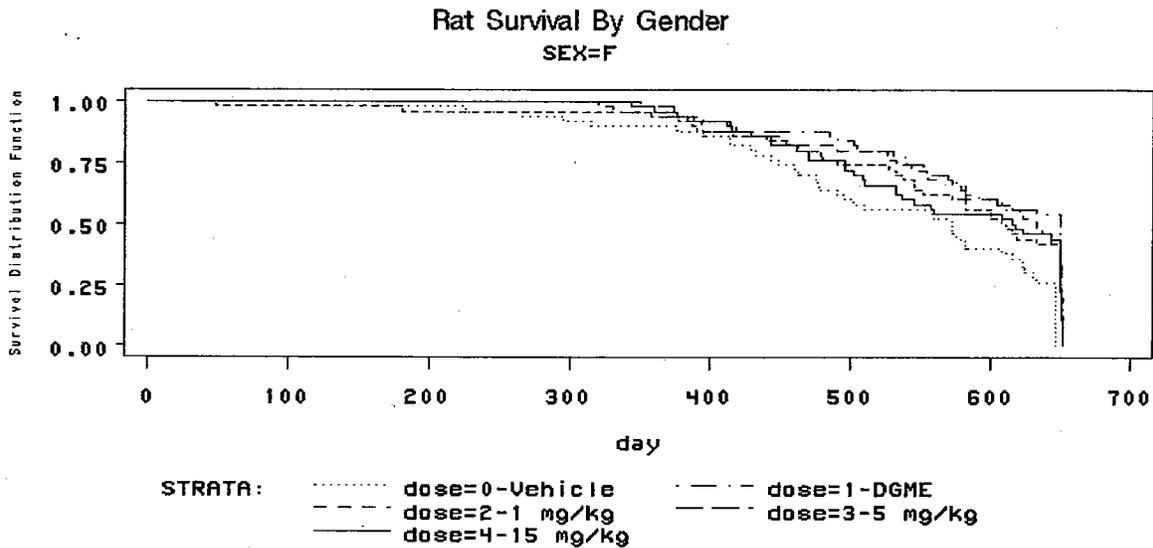
Best Available Copy



The analysis of the survival curves above provides no statistically significant evidence of any difference in survival among the four Dapsone groups. Strictly, failure to demonstrate differences does not imply no difference, but the large p -values here are consistent with the notion of little to no difference in survival. Interestingly, the DGME group seems to have higher survivability.

For females the test of homogeneity in survival over the five treatment groups was close to statistically significant (logrank $p = 0.0795$ and Wilcoxon $p = 0.0918$). Among the four Dapsone groups the test of homogeneity in survival was not statistically significant (logrank $p = 0.1926$ and Wilcoxon $p = 0.1961$). The test of trend in these four groups also was not significant ($p = 0.1424$). The Kaplan-Meier estimates curves for the time to first tumor are given below:

Figure A.1.2. Female Sprague-Dawley Rats



Although, differences among the four Dapsone groups are not statistically significant, most of the apparent differences are due to the lower survivability in the vehicle group. Note again that the DGME group seems to be associated with the lowest mortality (i.e., highest survival).

APPEARS THIS WAY
ON ORIGINAL

Appendix 2. ATLS-123 Two Year Rat Study: Survival to Specified Time

The FDA toxicologist wanted to investigate the high mortality in the vehicle group, and requested information on the survival in the control groups in other studies. Note that in the current study only 30% of the male rats survived to Week 100, and 26% of the female rats survived to Week 92. The Sponsor provided the following data for four supposedly similar studies:

Percentage of Male Control Rats Still Alive During Week 100

	Study 1	Study 2	Study 3	Study 4
Week 100	0*	0*	46.7%	47.1%

* The Sponsor indicated these studies were prematurely stopped due to excess mortality in a test drug group.

Percentage of Female Control Rats Still Alive During Week 92

	Study 1	Study 2	Study 3	Study 4
Week 92	65%	56.7%	60%	71.4%

One approach to the analysis is to note from the Central Limit theorem that the proportion surviving approximately follows a normal distribution. One way to model this relation is to express the within study proportion as constant across studies with an additive noise proportion. If we denote the mean over the studies as M and the standard deviation as s_d , then ignoring the prematurely stopped studies as following a different model, we can approximately test the homogeneity of the ATLS-123 values, denoted Y , with a simple two-sided t-test, as with:

$$t = (Y - M) / s_d.$$

Note that for male rats the corresponding t-statistic has one degree of freedom, while the t-statistic for the female rats has three degrees of freedom. Both tests were statistically significant ($p = 0.0106$ and $p = 0.0101$, respectively), indicating lack of homogeneity. However, the degrees of freedom for error are small and the corresponding t-tests are not very robust.

Although the Sponsor did not provide the number of rats in each study, usually these studies involve 50 rats of each gender for each treatment group. Note that this implies that the percentage still alive would be an even whole number, not satisfied by most of the groups above. Still assuming this is true, for each study i , we can model each proportion as having a Binomial($p_i, 50$) distribution. Then we can compute the likelihood ratio:

$$? = \sup_{\text{restricted}} ? p_i^{x_i} (1 - p_i)^{50-x_i} / \sup ? p_i^{x_i} (1 - p_i)^{50-x_i},$$

where the supremum of the numerator is restricted to reflect the hypotheses being tested. Then $-2 \log ?$ is approximately chi-square. The unrestricted $\arg \sup ? p_i^{x_i} (1 - p_i)^{50-x_i} = x_i/50$, and the restricted values are computed similarly over the pooled studies. Note that with these specifications $-2 \log ?$ is often called the deviance.

The relevant hypotheses can be written as follows, for $k=3$ or 5 , males and females, respectively:

$$H_0: p_1, p_2, \dots, p_k,$$

$$H_1: p_1, p_2 = \dots = p_k,$$

$$H_2: p_1 = p_2 = \dots = p_k,$$

where equality restrictions on the k parameters are applied as indicated. The corresponding likelihood ratio tests are given below:

Hypothesis	Males			Females		
	df	χ^2	p-value	df	χ^2	p-value
? : $H_1 H_0$	1	0.0016	0.9680	3	45.06	<0.0001
? : $H_2 H_1$	1	4.01	0.0451	1	22.88	<0.0001
? : $H_2 H_0$	2	4.02	0.1343	4	67.94	<0.0001

So the test of homogeneity of the proportions over the earlier studies referenced by the Sponsor, not counting the current study, is very nonsignificant for males. That is, for males, we accept the hypothesis that p_1 for the ATLS-123 study is unrestricted, while the p_2, p_3 referenced by the Sponsor satisfy $p_2 = p_3$ (significance level: $p = 0.9680$). Further, we reject the hypothesis that given $p_2 = p_3, p_1 = p_2 = p_3$ ($p = 0.0451$). For females we strongly reject the various hypotheses of uniformity in survival.

One alternative approach is Bayesian. This allows one to model the probability of exceeding the respective bound using a logit model, for example:

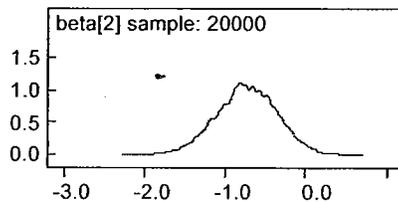
$$\text{logit}(p_i) = \beta_1 + \beta_2 d(i;1),$$

where $d(i;1)$ is one when $i = 1$, otherwise zero. Note that if $i=1$ is used to denote the current study, the β_2 parameter can be used to assess the consistency of the exceeding the bound in the current study with the studies provided by the Sponsor. Using a Normal(0.0, 1000²) prior for β_1, β_2 , we can estimate the parameters using the WinBUGS 1.4 program given below:

With level0 denoting the probability that $\beta_2 < 0$, for males this gives the following MCMC estimates:

Node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	-0.1382	0.2111	0.002077	-0.5512	-0.1357	0.2717	5001	20000
beta[2]	-0.7276	0.3764	0.00371	-1.488	-0.724	-0.006894	5001	20000
level0	0.9761	0.1527	0.001183	1.0	1.0	1.0	5001	20000

Thus the estimated probability of a differential effect due to the current study is estimated to be 0.9761. A plot of the β_2 is as follows:



For females we find the corresponding estimates:

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	0.5556	0.1461	0.001263	0.2717	0.5553	0.8418	5001	20000
beta[2]	-0.4737	0.3225	0.002891	-1.111	-0.4738	0.1603	5001	20000
level0	0.929	0.2569	0.002102	0.0	1.0	1.0	5001	20000

So for females, the estimated probability of a differential effect due to the current study is estimated to be 0.929.

Thus both for both genders the probability of a differential effect for the current study is estimated to be about 0.93 or higher.

These are estimated by programs similar to the following (for males):

```
Model{
  for (i in 1:N) {
    y[i] ~ dbin(p[i],n[i]);
    logit(p[i]) <- beta[1] + beta[2]*equals(i,1)
  }
  for (j in 1:2) {
    beta[j] ~ dnorm(0.0, 0.001) ;
  }
  level0 <- 1 - step(beta[2]);
}

data
  list( y=c(15,23,19), n=c(50,50,40), N=3)
inits
  list(beta=c(0.1,0.1))
```

**APPEARS THIS WAY
ON ORIGINAL**

Appendix 3. ATLS-123 Two Year Rat Study: Bayesian Analysis of Survival

A commonly used distribution for event times is a Weibull distribution:

$$f(t_j) = r e^{\beta x_j} t_j^{r-1} \exp(-e^{\beta x_j} t_j^r)$$

where t_j is the failure time of an individual with covariate vector x_j and β is the corresponding vector of unknown regression coefficients. This leads to a baseline hazard function:

$$h_0(t_j) = r t_j^{r-1}$$

The parameterization used in BUGS is based on exponentiating the linear predictor, i.e., defining $m_j = e^{\beta z_j}$ gives:

$$t_j \sim \text{Weibull}(r, m_j)$$

For censored observations, the survival distribution is a truncated Weibull, with lower bound corresponding to the censoring time and assumes independent (i.e., noninformative) censoring. The regression β coefficients are assumed a priori to follow independent Normal distributions with zero mean and "vague" precision 0.0001. The shape parameter r for the survival distribution was given a $\text{Gamma}(1, 0.0001)$ prior, which is slowly decreasing on the positive real line.

One model specifies a single parameter β_i for each treatment i , $i=1, \dots, 5$. An alternative model would be to specify a linear trend in dose due in the four metronodazole treatment groups and a separate parameter for the DGME control group. Thus we model the linear predictor as

$$\log(m_j) = \beta_1 \delta_{1j} + \beta_2 \delta_{2j} + \beta_3 \delta_{3j} + \beta_4 \delta_{4j} + \beta_5 \delta_{5j},$$

or

$$\log(m_j) = \beta_1 \delta_{2j} + (1 - \delta_{2j}) (\beta_2 + \beta_3 d_j),$$

where $\delta_{ki} = 1$ if observation is from the k^{th} treatment group, 0 otherwise and d_j denotes the dose in the i^{th} Dapsone treatment group, i.e., 0, 1, 5, or 15. We assume priors $\beta_i \sim N(0.0, 1000^2)$.

One approach to model selection in Bayesian models is to use the Deviance Information Criterion (DIC). Effectively, for $D(\theta)$ denoting the usual deviance, $\text{DIC} \approx E(D(\theta)) + 1/2 (\text{Var}(D(\theta)))$. For a given data set the model with the smallest DIC would be preferred.

Deviance Information Criterion	Males	Females
Model with different parameter for each group	2281.73	2066.6
Model with trend in Dapsone groups, separate DGME parameter	4036.32	3898.87

Thus for both genders the model with a different treatment group for each parameter would be preferred to the model with linear trend. These parameters were estimated using WinBUGS 1.4 programs similar to the one presented below. The programs with a different parameter for each treatment group lead to the following estimates:

Parameter estimates for Male Rats

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	-18.28	1.455	0.0965	-21.34	-18.17	-15.56	10001	40000
beta[2]	-18.94	1.478	0.09776	-22.05	-18.85	-16.19	10001	40000
beta[3]	-18.44	1.462	0.09685	-21.5	-18.33	-15.72	10001	40000
beta[4]	-18.58	1.466	0.0971	-21.66	-18.48	-15.83	10001	40000
beta[5]	-18.29	1.46	0.09677	-21.38	-18.19	-15.58	10001	40000
dgme.control	-0.6663	0.2694	0.002687	-1.194	-0.6651	-0.1371	10001	40000
low	-0.1589	0.2488	0.001828	-0.6511	-0.1579	0.3284	10001	40000
medium	-0.3035	0.2523	0.002133	-0.7947	-0.3038	0.1913	10001	40000
high	-0.01494	0.2425	0.001736	-0.4879	-0.0160	0.4639	10001	40000

Parameter estimates for Female Rats

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	-26.07	1.779	0.1213	-29.41	-25.96	-22.78	10001	40000
beta[2]	-26.82	1.795	0.1221	-30.21	-26.71	-23.49	10001	40000
beta[3]	-26.56	1.792	0.1221	-29.94	-26.44	-23.2	10001	40000
beta[4]	-26.48	1.789	0.1218	-29.86	-26.37	-23.17	10001	40000
beta[5]	-26.47	1.787	0.1217	-29.82	-26.37	-23.17	10001	40000
dgme.control	-0.751	0.2675	0.002358	-1.283	-0.7489	-0.232	10001	40000
low	-0.4834	0.2498	0.001837	-0.9807	-0.481	0.007564	10001	40000
medium	-0.4114	0.2515	0.00168	-0.9069	-0.4108	0.08329	10001	40000
high	-0.402	0.253	0.00181	-0.9036	-0.3992	0.09313	10001	40000

The terms for dgme.control, low, medium, and high correspond correspond to the difference between the named dose group and Dapsone vehicle. Since the credible intervals for the difference between Dapsone vehicle and DGME both exclude zero we can conclude there is strong evidence of this difference. For males there is little evidence of a difference between the other dose groups and control, while for females there is some evidence for a difference.

These estimates were derived from models similar to the following for trend:

```

model{
  for(i in 1 : N) {
    t[i] ~ dweib(r, mu[i])I(cens[i],);
    term[i] <- -beta[2] + beta[3] * (equals(dosegp[i], 3) + 5 * equals(dosegp[i], 4) +
      15 * equals(dosegp[i], 5));
    mu[i] <- exp(beta[1] * equals(dosegp[i], 2) + (1 - equals(dosegp[i], 2)) * term[i])
  }
  for (j in 1:nb){
    beta[j] ~ dnorm(0.0, 0.001)
  }
  r ~ dexp(0.001)
}

inits
  list(beta=c(1,1,0), r=3)
data
  list(N=249, nb=3)
dosegp[ ] t[ ] cens[ ]
  1 NA 698
  1 569 0
- data -
  5 532 0
  5 NA 700
END

```

Appendix 4. ATLS-123 Two Year Rat Study: Tumorigenicity Analysis

Note that of the five treatment groups, only the vehicle, DGME, and high dose groups were exhaustively analyzed over most neoplasms. Then, strictly speaking, an argument can be made that test of trend over the Dapsone doses is not appropriate. However, since generally at least 60% of the animals in the low and medium dose groups were analyzed, it was felt that was close enough to being exhaustively analyzed so that tests of trend could be useful and interpretable.

The following table displays the number of neoplasms in each organ and tumor combination. For the first reference to an organ, the number of animals with a histopathological analysis is included, denoted "# neoplasms/# animals", for other neoplasms only the number of neoplasms. Then the significance levels of tests of trend are presented, followed by pairwise tests of the high dose group (15 mg/kg/day) versus vehicle, and finally tests of then the DGME group versus the vehicle group. For each of these three comparisons, i.e., trend, high vs. vehicle, and DGME vs. vehicle, both exact (i.e. permutation) and asymptotic Peto tests are presented. For the small number of neoplasms generally found here, the exact tests would usually be recommended. Tumor incidence is given, but when there is only one neoplasm, no tests would be statistically significant, and the associated test statistics and significance levels are not presented.

The Sponsor did not analyze some neoplasms or combinations that are usually analyzed systemically or with grosser organ definition. Due to programming difficulties, the FDA analysis of some of these pooled results uses the Cochran-Armitage test of trend and Fisher exact tests instead of Peto mortality-adjusted tests. The fact that these tests are not adjusted for differences in mortality should cause no problem as there was no evidence of differences in survival. However, in response to a specific request the Sponsor provided Peto mortality adjusted tests for some combinations of organs and neoplasms. The results of these tests are given in the second row of significance levels for these combinations.

Table A.4.1. ATLS-123 Tumorigenicity in Male Sprague-Dawley Rats

Tumor						Trend		Hi vs. Control		Control vs DGME	
	Vehicle	DGME	Low	Medium	High	Exact2	Asymp2	Exact3	Asymp3	Exact4	Asymp4
Adrenal glands											
Adenocarcinoma	0/50	0/50	0/33	0/31	1/50						
Adenoma	0	0	1	0	0						
Ganglioneuroma	1	0	0	0	0						
Pheochromocytoma, benign	3	2	1	2	1	0.8994	0.8966	0.9776	0.9380	0.9334	0.9316
Pheochromocytoma, malignant	1	1	0	2	0	0.8214	0.8161	1.0000	0.8944	0.8632	0.8672
Bone											
Liposarcoma	0/1	0/1	0/0	0/2	1/2						

Table A.4.1.(cont.) ATLS-123 Tumorigenicity in Male Sprague-Dawley Rats

Tumor	Trend					Hi vs. Control		Control vs DGME			
	Vehicle	DGME	Low	Medium	High	Exact2	Asymp2	Exact3	Asymp3	Exact4	Asymp4
Brain (cerebellum)											
Astrocytoma malignant	0/50	0/50	1/33	0/31	0/50						
Glioma, malignant	1	0	0	0	0						
Brain (cerebrum)											
Astrocytoma malignant	0/50	1/50	0/33	0/31	1/50						
Astrocytoma, benign	0	0	1	1	0	0.7131	0.7824				
Glioma, malignant	1	0	0	0	0						
Cavities											
Hemangiosarcoma	0/0	0/1	0/1	0/0	1/1						
Duodenum											
Adenocarcinoma	0/50	1/50	0/33	0/31	0/50						
Ear											
Fibroma	1/1	0/0	0/0	0/1	0/0						
Harderian Gland											
Fibrosarcoma	1/50	0/50	0/33	0/31	0/50						
Heart											
Fibrosarcoma	1/50	0/50	1/33	0/31	0/50						
Jejunum											
Adenocarcinoma	0/50	1/50	0/33	0/31	0/50						
Kidneys											
Lipoma	0/50	1/50	0/35	0/34	2/50	0.0911	0.0243	0.2143	0.0893	0.4737	0.4790
Mesenchymal tumor	1	0	0	0	0						
Liver											
Hepatocellular adenoma	0/50	1/50	0/38	2/38	0/50	0.6148	0.6517			0.3333	0.3618
Hepatocellular carcinoma	1	0	0	1	0	0.7667	0.7959	1.0000	0.8944	1.0000	0.9891
Histiocytic Sarcoma	0	0	0	0	1						
Hepato. Adenoma/Carcinoma**	1	1	0	3	0	0.2743	0.2783	0.3149	1.0000	0.7525	1.0000
						0.8446		1.000		0.7495	
Lungs (with bronchi)											
Alveolar bronchiolar adenoma	0/50	0/50	0/37	1/37	0/50						
Malignant lymphoma	0	0	1	0	0						
Mesothelioma, malignant	0	0	1	0	0						

** Tests on 1st row not mortality adjusted, tests on 2nd row are mortality adjusted, and are from Sponsor's 22 April 2005 submission.

Table A.4.1. (cont.) ATLS-123 Tumorigenicity in Male Sprague-Dawley Rats

Tumor						Trend		Hi vs. Control		Control vs DGME	
	Vehicle	DGME	Low	Medium	High	Exact2	Asymp2	Exact3	Asymp3	Exact4	Asymp4
Lymph node											
Adenocarcinoma	0/1	0/3	0/4	1/6	0/7						
Hemangiosarcoma	0	1	0	0	0						
Lymph node (Mandibular)											
Hemangiosarcoma	0/48	0/48	0/35	0/31	1/50						
Mammary gland or area											
Adenocarcinoma	0/50	0/50	0/34	1/32	0/50						
Adenoma	0	0	1	0	0						
Fibroadenoma	0	0	2	1	0	0.7746	0.8479				
Lipoma	1	0	0	0	0						
Neurofibroma	0	0	0	0	1						
Squamous cell carcinoma	0	1	0	0	0						
Mediastinum											
Liposarcoma	0/0	0/0	1/3	0/0	0/1						
Pancreas											
Islet cell adenoma	0/50	4/50	2/34	2/33	4/50	0.1074	0.1098	0.0572	0.0276	0.1846	0.1812
Islet cell carcinoma	1	0	0	0	0						
Islet cell Adenoma/Carcinoma**	1	4	2	2	4	0.1299	0.1341	0.1811	0.1687	0.1811	0.1687
Parathyroid glands (when in section)											
Adenoma	0/42	0/44	0/31	0/25	1/45						
Pituitary gland											
Adenoma	26/50	23/50	20/36	23/36	25/50	0.8255	0.8262	0.7153	0.6696	0.9947	0.9945
Skin (other)											
Basal cell tumor, malignant	0/11	1/13	0/21	0/21	0/12						
Fibrosarcoma	0	0	2	2	0	0.6732	0.7391				
Hemangioma	0	0	1	1	0	0.6556	0.6758				
Keratoacanthoma	1	0	1	1	0	0.8862	0.8463	1.0000	0.8716	1.0000	0.9772
Lipoma	0	0	0	1	0						
Osteoma	1	0	0	0	0						
Papilloma	0	3	0	2	4	0.0018	0.0005	0.0625	0.0254	0.0784	0.0805

** Tests not mortality adjusted

Table A.4.1. (cont.) ATLS-123 Tumorigenicity in Male Sprague-Dawley Rats

Tumor	Vehicle	DGME	Low	Medium	High	Trend Exact2	Asymp2	Hi vs. Control Exact3	Control vs DGME Asymp3	Exact4	Asymp4
Skin (cont.)											
Schwannoma, malignant	0	0	1	0	0						
Sebaceous adenoma	0	0	1	0	0						
Squamous cell carcinoma	1	1	0	0	0					0.7647	0.7670
Trichoepithelioma	0	0	0	1	0						
Kera./Papil./Squamous**	2	4	1	3	4	0.0436	0.0296	0.3652	0.4085	0.4101	0.478
						0.0459		0.2884		0.3898	
Spleen											
Hemangioma	0/50	1/50	0/33	0/31	0/50						
Hemangiosarcoma	0	1	0	0	2	0.1989	0.0646	0.2414	0.1034	0.6154	0.5937
Leukemia, NOS	0	0	0	0	1						
Testes											
Adenoma, Interstitial Cell	0/50	1/50	2/35	1/36	2/50	0.3556	0.4086	0.4000	0.1846	0.6154	0.5937
Thymus or remnant											
Thymoma	0/42	0/41	1/27	0/27	1/41	0.3228	0.3212	0.5385	0.2498		
Thyroid gland											
Adenoma	0/50	0/50	1/33	0/31	0/50						
Cystadenoma	1	1	0	0	0					0.9123	0.9024
Follicular adenoma	0	3	0	1	2	0.1543	0.1053	0.4000	0.1846	0.3104	0.2868
Follicular carcinoma	1	0	0	0	0						
Parafollicular cell adenoma	2	3	0	1	4	0.1869	0.1538	0.4083	0.2855	0.6157	0.6110
Follicular Adenoma/carcinoma**	1	3	0	1	2	0.2129	0.1684	0.5000	0.5577	0.3087	0.3074
						0.3179		0.5640		0.4056	
Urinary bladder											
Transitional cell carcinoma	0/50	1/50	0/33	0/32	0/50						
Zymbal's Gland											
Squamous cell carcinoma	0/50	0/50	0/33	0/30	1/50						
Systemic*,**											
Hemangioma	0/50	1/50	1/50	1/50	0/50	0.5000	0.2949			0.5000	0.3143
						0.6425				0.6316	
Hemangiosarcoma	0	2	0	0	4	0.0036	0.0004	0.0587	0.0412	0.2475	0.1531
						0.0037		0.0750		0.3926	

* Assumes all animals at risk

** Tests on 1st row not mortality adjusted, tests on 2nd row are mortality adjusted, and are from Sponsor's 22 April 2005 submission.

Table A.4.1. (cont.) ATLS-123 Tumorigenicity in Male Sprague-Dawley Rats

Tumor						Trend		Hi vs. Control Control vs DGME			
	Vehicle	DGME	Low	Medium	High	Exact2	Asymp2	Exact3	Asymp3	Exact4	Asymp4
Systemic*,**											
Hemangioma/ Hemangiosarcoma**											
0	3	1	1	4	0.0145	0.0080	0.0587	0.0412	0.1212	0.0786	
					0.0164		0.0750		0.2399		

* Assumes all animals at risk

** Tests on 1st row not mortality adjusted, tests on 2nd row are mortality adjusted, and are from Sponsor's 22 April 2005 submission.**Table A.4.2. ATLS-123 Tumorigenicity in Female Sprague-Dawley Rats**

Tumor						Trend		Hi vs. Control Control vs DGME			
	Vehicle	DGME	Low	Medium	High	Exact2	Asymp2	Exact3	Asymp3	Exact4	Asymp4
Adrenal glands											
Pheochromocytoma, benign											
1/50	0/50	0/35	0/32	0/50							
Pheochromocytoma, malignant											
0	0	0	1	1	0.1553	0.1236	0.5778	0.2737			
Brain (cerebrum)											
Astrocytoma malignant											
0/50	0/50	1/29	0/292	0/50							
Kidneys											
Transitional cell carcinoma											
0/50	0/50	0/30	1/30	0/50							
Liver											
Hepatocellular adenoma											
1/50	0/50	0/36	0/33	0/50							
Hepatocellular carcinoma											
0	0	1	0	0							
Histiocytic Sarcoma											
1	0	0	0	0							
Leukemia myelogenous											
1	0	0	0	0							
Malignant lymphoma											
1	0	0	0	0							
Lungs (with bronchi)											
Leukemia, mononuclear cell											
0/49	0/50	1/30	0/31	0/50							
Lymph node											
Malignant lymphoma											
0/1	0/1	0/2	1/3	0/3							
Lymph node (mediastinal)											
Malignant lymphoma											
0/1	0/0	1/1	0/0	0/0							
Lymph nodes (mandibular)											
Malignant lymphoma											
0/50	0/50	1/29	0/29	0/50							
Mammary gland or area											
Adenocarcinoma											
9/50	8/50	15/41	5/42	4/50	0.9933	0.9910	0.8488	0.7831	0.7072	0.7061	
Adenoma											
11	5	6	9	8	0.8352	0.8360	0.9135	0.8759	0.9825	0.9806	
Fibroadenoma											
17	17	14	11	14	0.7243	0.7264	0.8002	0.7445	0.6997	0.6988	

Table A.4.2. (cont.) ATLS-123 Tumorigenicity in Female Sprague-Dawley Rats

Tumor						Trend		Hi vs. Control		Control vs DGME	
	Vehicle	DGME	Low	Medium	High	Exact2	Asymp2	Exact3	Asymp3	Exact4	Asymp4
Mammary Gland (cont.)											
Fibroma	0	0	0	1	0						
Pancreas											
Islet cell adenoma	1/50	2/50	0/29	0/29	1/50	0.6444	0.4109	0.7421	0.5300	0.6300	0.6202
Parathyroid glands											
Adenoma	0/44	1/40	0/24	0/21	0/43						
Pituitary gland											
Adenocarcinoma	0/50	0/50	0/48	0/45	1/50						
Adenoma	43	46	42	37	36	0.9895	0.9881	0.9984	0.9976	0.9946	0.9947
Skin (abdominal)											
Keratoacanthoma	0/50	0/50	0/30	0/29	1/50						
Skin (other)											
Histiocytic Sarcoma	0/8	0/7	0/15	0/17	1/6						
Squamous cell carcinoma	1	0	1	0	0	1.0000	0.9263	1.0000	0.7917		
Spleen											
Malignant lymphoma	0/50	0/50	0/30	0/29	1/50						
Subcutaneous tissue											
Malignant fibrous histiocytoma	0/0	0/0	1/1	0/0	0/0						
Thyroid gland											
Follicular adenoma	0/50	0/50	0/29	1/29	0/50						
Parafollicular cell adenoma	1	2	2	1	1	0.6442	0.7065	0.7421	0.5300	0.6300	0.6202
Parafollicular cell carcinoma	0	1	0	0	0						
Uterus											
Leiomyoma	1/50	0/50	0/29	0/29	0/50						
Polyp	1	0	0	1	1	0.3844	0.3449	0.6627	0.4636	1.0000	0.9669
Vagina											
Fibroma	1/50	0/50	0/30	0/29	0/50						
Squamous cell carcinoma	0	0	1	0	0						
Systemic											
Malignant lymphoma*,**	1/50	0/50	2/50	1/50	1/50	0.4499	0.3728	0.7525	1.0000	0.3149	0.5000
						0.4739		0.7667		1.000	

* Assumes all animals at risk

** Tests on 1st row not mortality adjusted, tests on 2nd row are mortality adjusted, and are from Sponsor's 22 April 2005 submission.

Appendix 5. ATLS-163 Tg.AC Study: Observed Individual Skin Lesions At End of Study

The following tables display lesion counts for each animal with lesions. The Sponsor defined papilloma codes were as follows:

SOA = Site of application

NSOA = Non-site of application

L = Latent papilloma (Not yet observed for 3 consecutive weeks)

D = Observed for 3 consecutive weeks but subsequently diminished

(Note DNSOA papillomas are indicated by parentheses in NSOA column, e.g., (1))

Table A.5.1 Male Tg.AC Papilloma Incidence at End of Study, by Animal

Group 1 - 25% DGME Vehicle						Group 3 - 3% Dapsone / 25% DGME					
Animal	SOA	DSOA	LSOA	NSOA	LNSOA	Animal	SOA	DSOA	LSOA	NSOA	LNSOA
5				1							
22			1								
25					1						
Group 4 - 5% Dapsone / 25% DGME						Group 5 - 10% Dapsone / 25% DGME					
Animal	SOA	DSOA	LSOA	NSOA	LNSOA	Animal	SOA	DSOA	LSOA	NSOA	LNSOA
84		1	1								
Group 6 - Acetone Vehicle						Group 8 - 5% Dapsone in Acetone					
Animal	SOA	DSOA	LSOA	NSOA	LNSOA	Animal	SOA	DSOA	LSOA	NSOA	LNSOA
251					1	291				4 (1)	3
254				1	1	299				2 (2)	
256				(1)		309				1	
271	2										
Group 2 - 20µg TPA / 25% DGME						Group 7 - 1.25µg TPA in 0.1ml Acetone					
Animal	SOA	DSOA	LSOA	NSOA	LNSOA	Animal	SOA	DSOA	LSOA	NSOA	LNSOA
26	27			2		276	22	1		1 (1)	
27	5	17		1		277	24				
28	23	1		3		278	22	1	2		
29	46					279	22	1		1	
30	26			1		280	21				
31	40					281	13	7			
32	23			1		282	23	2	9	4	1
33	18			4		283	3	1	3		
34	32	2				284	19	1			
35		2		(1)		285	21	4			
36	29			2							
37	23										
38	11	10		2							
39	10										
40	14	10		4							
42	24										
43	35										
44	19			3							
45	16			4							
46	16	10									
47	25										
48	14	7		2							
49	28										
50	34	1									

These assessments were made at Week 27 for the Positive Control animals (Groups 2 and 7), Week 31 for the acetone groups (Groups 6 and 8), and Week 33 for the 25% DGME and Dapsone groups (Groups 1, 3, 4, and 5). Note that for animals sacrificed or died earlier in the study, the last number of tumors is carried forward.

Table A.5.2 Female Tg.AC Papilloma Incidence at End of Study, by Animal

Group 1 - 25% DGME						Group 3 - 3% Dapsone / 25% DGME					
Animal	SOA	DSOA	LSOA	NSOA	LNSOA	Animal	SOA	DSOA	LSOA	NSOA	LNSOA
126				2		182				1 (1)	
135				1		185				(2)	
136					1	187				1	1
137				1		188				1	
143				1		191			2	2	
146	1					193			1		
150				1		196				1	
						199				2	
						200				1	
Group 4 - 5% Dapsone / 25% DGME						Group 5 - 10% Dapsone / 25% DGME					
Animal	SOA	DSOA	LSOA	NSOA	LNSOA	Animal	SOA	DSOA	LSOA	NSOA	LNSOA
214				1		231				1	
224	2					237				2	
						239				(1)	
Group 6 - Acetone Vehicle 0 ml/kg						Group 8 - 5% Dapsone in Acetone					
Animal	SOA	DSOA	LSOA	NSOA	LNSOA	Animal	SOA	DSOA	LSOA	NSOA	LNSOA
312				1							
330					1						
Group 2 - 20µg TPA / 25% DGME						Group 7 - 1.25µg TPA in 0.1ml Acetone					
Animal	SOA	DSOA	LSOA	NSOA	LNSOA	Animal	SOA	DSOA	LSOA	NSOA	LNSOA
151	25			1		336	21			1 (1)	
152	7	2				337	16	9			
154	28			1		339	20	1			
155	26	2				340	18	1		1	
156	24	1		1		341	19	1			
157	19					342	6				
158	9	12		1		343	12	12			
159		1				344	9	1	3		
160	20	1				345	15	3			
161	24	2	4	1							
162	24	1									
163	21	1									
164	26	2		1							
165	17	3		3							
166	21	1									
167	24										
168	4	1		3							
169	17										
170	22	1									
171	5										
172	25										
173	7	9		2 (1)							
174	27	2		2							
175	21										

Note that results from the active controls (Groups 2 & 7) demonstrate the sensitivity of the model, while results in Groups 1, 3, 4, and 5 show no difference in tumor burden (SOA or NSOA).

Appendix 6. ATLS-163 Tg.AC Study: Survival Analysis

For convenience the eight treatment groups in each gender are summarized below:

Table A.6. 1 ATLS-163: Dose Groups in Tg.AC mice

Group	# animals/ gender	Dose	Group	# animals/ gender	Dose
1	25	25% DGME vehicle	5	25	10% Dapsone / 25% DGME
2	25	20µg TPA / 25% DGME	6	25	Acetone
3	25	3% Dapsone / 25% DGME	7	10	1.25µg TPA / Acetone
4	25	5% Dapsone / 25% DGME	8	25	5% Dapsone / Acetone

For males the hypotheses of homogeneity in survival over the four DGME/Dapsone groups (Groups 1, 3-5) was rejected with high statistical significance (both logrank and Wilcoxon $p < 0.0001$). The test of linear trend and non-linearity in trend were also statistically highly significant (both $p < 0.0001$). The hypotheses of homogeneity in survival over the two Acetone/Dapsone groups (Groups 6 & 8) was also rejected with high statistical significance (logrank $p = 0.0019$ and Wilcoxon $p = 0.0018$). The Kaplan-Meier estimated curves of the time to first tumor are given below:

Figure A.6.1. Male Tg.AC Mice (All groups)

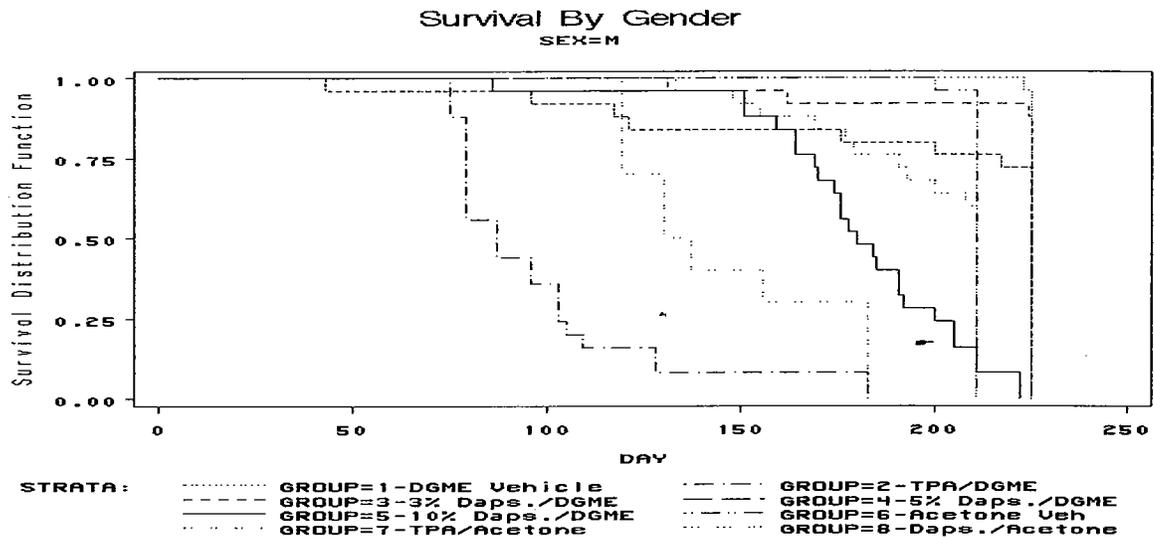
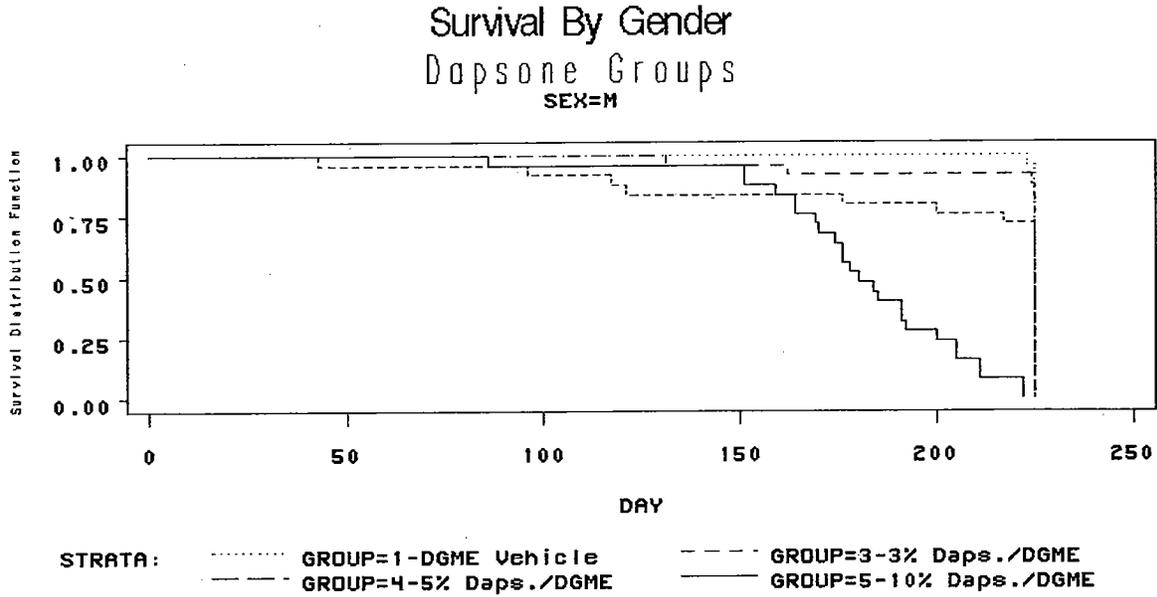


Figure A.6.2. Male Tg.AC Mice (Four Dapsone Groups)

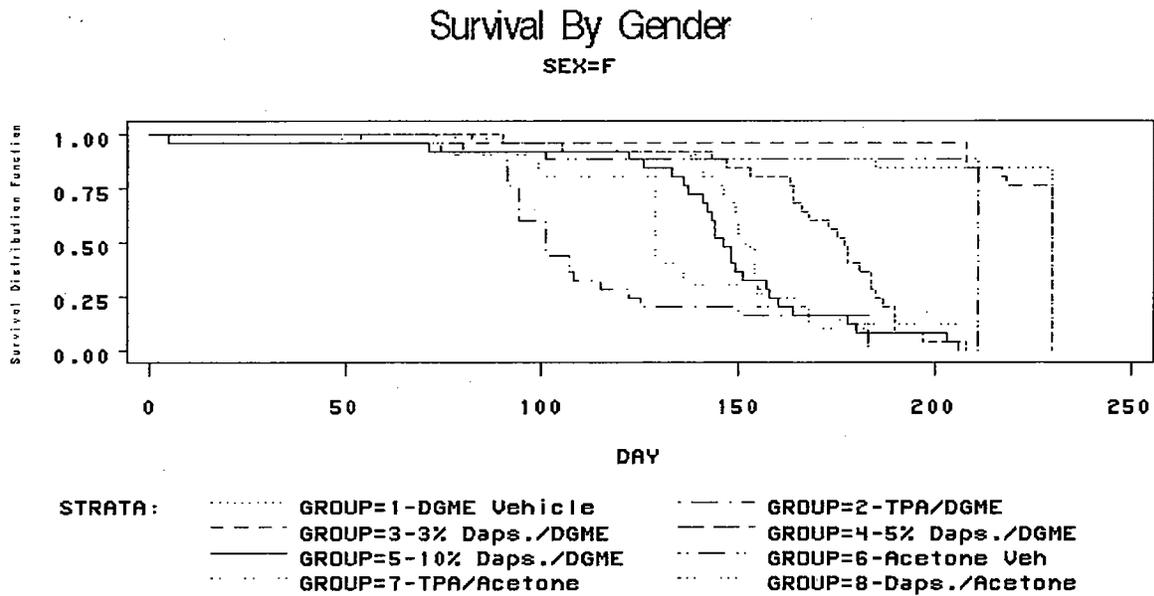


Note from the table above, and summarized in Table 10, there is little difference in survivability between the DGME vehicle, the 3% Dapsone group, and the 5% Dapsone group. However, the 10% Dapsone group shows a clear increase in mortality.

**APPEARS THIS WAY
ON ORIGINAL**

For females the hypothesis of homogeneity in survival over the four DGME/Dapsone groups (Groups 1, 3-5) and the Acetone/Dapsone groups (Groups 6 & 8) was rejected with high statistical significance (for each test & group, both logrank and Wilcoxon $p < 0.0001$). The test of linear trend and nonlinearity in trend were also statistically highly significant in the four DGME/Dapsone groups (both $p < 0.0001$).

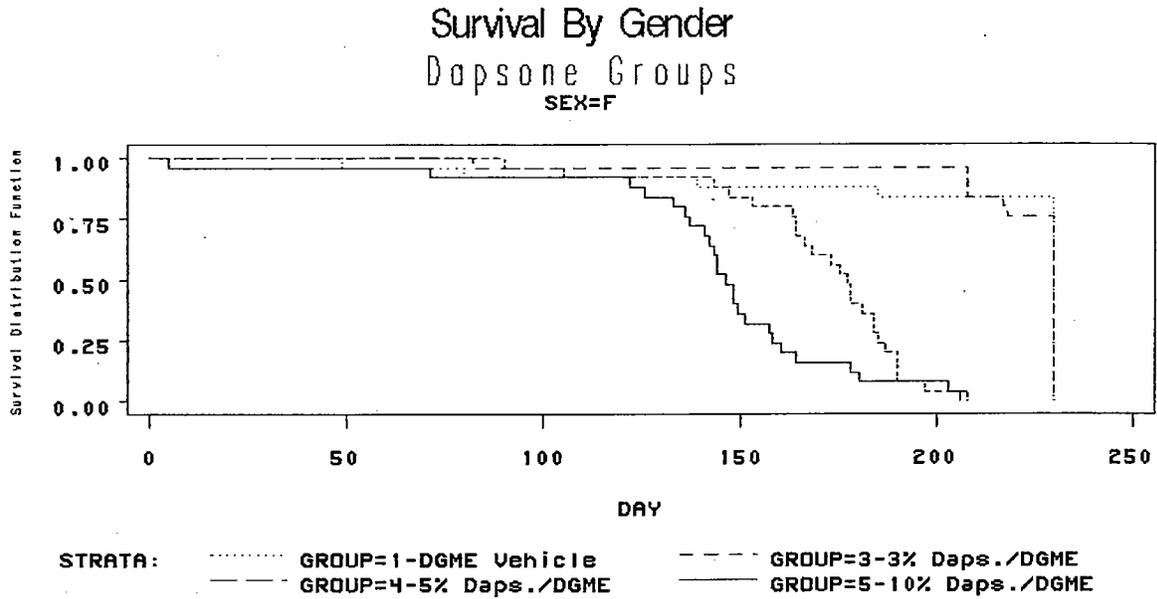
Figure A.6.3. Female Tg.AC Mice



To clarify results Kaplan-Meier curves of the four Dapsone groups are presented next.

**APPEARS THIS WAY
ON ORIGINAL**

Figure A.6.4. Female Tg.AC Mice (Four Dapsone Groups)



Note from the table above, and summarized in Table 11, there seems to be little difference between the vehicle and the 3% Dapsone group, but thereafter an apparently dose related increasing trend in mortality for the 5% and 10% Dapsone groups.

APPEARS THIS WAY
ON ORIGINAL

Appendix 7. ATLS-163 Tg.AC Study: Weekly SOA Tumor Incidence

The following tables summarize the number of animals that died (moribund sacrifice, natural or accidental death, but not planned sacrifices) and the number alive during the study by week. The tables also show the number of animals with active or latent papillomas, and active, diminished or disappeared, and latent tumors prior to death, either at the site of application (SOA) or non-site of application (NSOA). Note that for animals sacrificed or died earlier, the last number of tumors is carried forward. The mean number of tumors is presented among mice with SOA or NSOA tumors. The minimum and maximum numbers of tumors are also displayed.

For both genders the following dose groups were specified

Group	# animals	Dose	Group	# animals	Dose
1	25	25% DGME vehicle	5	25	10% Dapsone / 25% DGME
2	25	20µg TPA / 25% DGME	6	25	Acetone
3	25	3% Dapsone / 25% DGME	7	10	1.25µg TPA / Acetone
4	25	5% Dapsone / 25% DGME	8	25	5% Dapsone / Acetone

Table A.5.1. Male Papilloma Tumor Incidence

	Week										
	1-4	5	6	7	8	9	10	11	12	13	14
1-25% DGME											
# dead	0	0	0	0	0	0	0	0	0	0	0
# alive	25	25	25	25	25	25	25	25	25	25	25
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOAtumors	0	0	0	0	0	0	0	0	0	0	0
2- 20ug TPA/ 25% DGME											
# dead	0	0	0	0	0	0	0	0	0	0	0
# alive	25	25	25	25	25	25	25	25	22	14	11
# w/SOA tumors	0	4	10	18	21	22	24	24	24	24	24
# w/NSOAtumors	0	0	0	1	1	4	7	13	15	13	13
SOA tumors mean	.	1.0	1.9	3.4	5.8	14.1	22.5	34.2	37.2	30.7	27.3
min-max	.	1	1-3	1-12	1-15	2-40	2-55	2-76	3-79	2-70	2-51
3- 3% Dapsone / 25% DGME											
# dead	0	0	0	0	0	0	0	0	0	0	0
# alive	25	25	25	25	25	25	25	25	25	25	25
# w/SOAtumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOAtumors	0	0	0	0	0	0	0	0	0	0	0
4- 5% Dapsone / 25% DGME											
# dead	0	0	0	1	1	1	1	1	1	1	1
# alive	25	25	25	24	24	24	24	24	24	24	24
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
5- 10% Dapsone / 25% DGME											
# dead	0	0	0	0	0	0	0	0	0	0	1
# alive	25	25	25	25	25	25	25	25	25	25	24
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOAtumors	0	0	0	0	0	0	0	0	0	0	0

Table A.7.1. (cont.) Male Papilloma Tumor Incidence

	Week 1-4	5	6	7	8	9	10	11	12	13	14
6- Acetone 0 ml/kg											
# dead	0	0	0	0	0	0	0	0	0	0	0
# alive	25	25	25	25	25	25	25	25	25	25	25
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	1	1	1	1	1	1	2	2	2	2	2
7- 1:25ug TPA/Acetone											
# dead	0	0	0	0	0	0	0	0	0	0	0
# alive	10	10	10	10	10	10	10	10	10	10	10
# w/SOA tumors	0	0	0	1	2	5	8	8	8	9	10
# w/NSOA tumors	0	0	0	0	0	0	0	1	1	1	1
SOA tumor mean	.	.	.	1.0	1.0	1.2	1.8	2.3	4.4	6.2	8.5
min-max	.	.	.	1	1	1-2	1-2	1-4	1-9	1-16	1-22
8- 5% Dapsone / Acetone											
# dead	0	0	0	0	0	0	0	0	0	0	0
# alive	25	25	25	25	25	25	25	25	25	25	25
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
Week											
	15	16	17	18	19	20	21	22	23	24	25
1- 25% DGME											
# dead	0	0	0	0	0	0	0	0	0	0	0
# alive	25	25	25	25	25	25	25	25	25	25	25
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
2- 20ug TPA/25% DGME											
# dead	0	1	1	1	1	1	1	1	1	1	1
# alive	9	5	4	4	4	2	2	2	2	2	2
# w/SOA tumors	24	24	24	24	24	24	24	24	24	24	24
# w/NSOA tumors	13	12	12	12	12	13	13	13	13	13	13
SOA tumors mean	26.3	25.7	24.9	25.2	25.5	24.8	24.8	24.8	24.8	24.8	24.8
min-max	2-46	3-46	3-46	2-46	2-46	2-46	2-46	2-46	2-46	2-46	2-46
3- 3% Dapsone / 25% DGME											
# dead	0	0	0	0	0	1	1	1	1	2	2
# alive	25	25	25	25	25	24	24	24	24	23	23
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
4- 5% Dapsone / 25% DGME											
# dead	1	1	1	1	2	2	2	2	2	2	2
# alive	24	24	24	24	23	23	23	23	23	23	23
# w/SOA	0	0	0	0	1	1	0	0	1	3	2
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
SOA tumors mean	1.0	1.0	.	.	1.0	1.0	1.0
Min-max	1	1	.	.	1	1	1

Table A.7.1. (cont.) Male Papilloma Tumor Incidence

	Week											
	15	16	17	18	19	20	21	22	23	24	25	
5- 10% Dapsone / 25% DGME												
# dead	1	1	1	1	1	1	1	1	3	4	7	
# alive	24	24	24	24	24	24	24	24	22	21	18	
# w/SOA tumors	0	0	1	0	0	0	0	0	1	1	0	
# w/NSOA	0	0	1	1	1	0	0	0	1	1	0	
SOA tumors mean	.	.	1.0	1.0	1.0	.	
Min-max	.	.	1	1	1	.	
6- Acetone 0 ml/kg												
# dead	0	0	0	0	0	0	0	0	0	0	0	
# alive	25	25	25	25	25	25	25	25	25	25	25	
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0	
# w/NSOA tumors	2	2	2	2	2	2	2	2	2	2	2	
7- 1.25ug TPA/Acetone												
# dead	0	0	0	0	0	0	0	0	0	0	0	
# alive	10	10	10	7	7	5	4	4	4	3	3	
# w/SOA tumors	10	10	10	10	10	10	10	10	10	10	10	
# w/NSOA tumors	1	2	2	1	2	2	2	2	2	3	3	
SOA tumor mean	12.0	15.2	18.6	16.5	18.3	18.5	17.8	18.0	18.3	19.4	21.1	
min-max	1-28	1-30	1-35	1-32	1-34	1-40	2-25	2-26	3-26	3-25	4-28	
8- 5% Dapsone / Acetone												
# dead	0	0	0	0	0	1	1	2	3	3	4	
# alive	25	25	25	25	25	24	24	23	22	22	21	
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0	
# w/NSOA tumors	0	0	0	0	0	1	0	0	0	0	2	
	Week											
	26	27	28	29	30	31	32	33				
1- 25% DGME												
# dead	0	0	0	0	0	0	0	1				
# alive	25	25	25	25	25	25	25	24				
# w/SOA tumors	0	0	0	0	0	0	0	1				
# w/NSOA tumors	0	0	0	0	1	1	2	2				
SOA tumors mean	1.0				
Min-max	1				
2- 20ug TPA/ 25% DGME												
# dead	1	1				
# alive	2	0				
# w/SOA tumors	24	24				
# w/NSOA tumors	13	13				
SOA tumors mean	24.8	24.8				
Min-max	2-46	2-46				
3- 3% Dapsone / 25% DGME												
# dead	2	2	2	2	2	2	2	3				
# alive	23	23	23	23	23	23	23	22				
# w/SOA tumors	0	0	0	0	0	0	0	0				
# w/NSOA tumors	0	0	0	0	0	0	0	0				

Table A.7.1.(cont.) Male Papilloma Tumor Incidence

	Week							
	26	27	28	29	30	31	32	33
4- 5% Dapsone / 25% DGME								
# dead	3	3	3	3	4	4	5	5
# alive	22	22	22	22	21	21	20	20
# w/SOA tumors	1	1	1	1	2	1	1	1
# w/NSOA tumors	0	0	0	0	0	0	0	0
SOA tumor mean	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Min-max	1	1	1	1	1	1	1	1
5- 10% Dapsone / 25% DGME								
# dead	11	13	15	18	19	23	23	25
# alive	14	12	10	7	6	2	2	0
# w/SOA tumors	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0
6- Acetone 0 ml/kg								
# dead	0	0	0	0	1	1	.	.
# alive	25	25	25	25	24	24	.	.
# w/SOA tumors	1	1	1	1	1	1	.	.
# w/NSOA tumors	2	2	2	2	3	3	.	.
SOA tumor mean	1.0	1.0	1.0	2.0	2.0	2.0	.	.
Min-max	1	1	1	2	2	2	.	.
7- 1.25ug TPA/Acetone								
# dead	0	0
# alive	3	0
# w/SOA tumors	10	10
# w/NSOA tumors	3	3
SOA tumor mean	21.1	22.2
min-max	4-28	7-34
8- 5% Dapsone / Acetone								
# dead	4	6	6	8	9	10	.	.
# alive	21	19	19	17	16	15	.	.
# w/SOA tumors	0	0	0	0	0	0	.	.
# w/NSOA tumors	2	2	3	3	3	3	.	.

Note that except for the TPA positive controls there are almost no SOA tumors among the males, i.e., at the end of the study there was only one SOA papilloma in each of Group 1 (25% DGME vehicle), Group 4 (5% Dapsone / 25% DGME), and Group 6 (acetone vehicle). With only 25 animals per group one could argue that this is too small a response upon which to base a statistical analysis. However, within the Dapsone treatment groups and the Acetone group no Fisher exact test comparing the number of SOA tumors was statistically significant.

Table A.7.2. Female Papilloma Tumor Incidence

	Week										
	1-4	5	6	7	8	9	10	11	12	13	14
1- 25% DGME											
# dead	0	0	0	0	1	1	1	1	1	2	2
# alive	25	25	25	25	24	24	24	24	24	23	23
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
2- 20ug TPA/ 25% DGME											
# dead	0	0	0	0	0	0	0	0	0	0	0
# alive	25	25	25	25	25	25	25	25	24	23	19
# w/SOA tumors	0	2	4	8	13	18	20	20	21	21	21
# w/NSOA tumors	0	0	0	0	0	3	3	6	8	8	10
SOA tumors mean	.	1.0	1.0	2.0	4.0	7.9	9.3	18.7	22.2	26.9	27.1
Min-max	.	1	1	1-5	1-15	1-25	1-33	4-61	1-41	2-52	3-58
3- 3% Dapsone / 25% DGME											
# dead	0	0	0	0	0	0	0	0	0	1	1
# alive	25	25	25	25	25	25	25	25	25	24	24
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
4- 5% Dapsone / 25% DGME											
# dead	0	0	0	0	0	0	0	0	0	0	1
# alive	25	25	25	25	25	25	25	25	25	25	24
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	1	0	0	0
5- 10% Dapsone / 25% DGME											
# dead	1	1	1	1	1	1	1	2	2	2	2
# alive	24	24	24	24	24	24	24	23	23	23	23
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	1	2	2	2	2	2	2	2
6- Acetone 0 ml/kg											
# dead	0	0	0	0	0	1	1	1	2	2	2
# alive	25	25	25	25	25	24	24	24	23	23	23
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	1	0	0	0	0	0	0	0	0	0	0
7- 1.25ug TPA/Acetone											
# dead	0	0	0	0	0	0	0	0	1	1	1
# alive	10	10	10	10	10	10	10	10	9	9	9
# w/SOA tumors	0	0	0	0	1	4	8	9	9	9	9
# w/NSOA tumors	0	0	0	0	0	0	1	1	1	1	1
SOA tumors mean	1.0	1.3	2.3	4.0	6.7	7.9	9.6
Min-max	1	1-2	1-5	1-6	1-12	2-17	2-19
8- 5% Dapsone / Acetone											
# dead	0	0	0	0	0	0	0	0	0	0	1
# alive	25	25	25	25	25	25	25	25	25	25	24
# w/SOA tumors	0	0	0	0	0	1	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
SOA tumors mean	1.0
Min-max	1

Table A.7.2. (cont.) Female Papilloma Tumor Incidence

	Week										
	15	16	17	18	19	20	21	22	23	24	25
1- 25% DGME											
# dead	2	2	2	2	2	2	3	3	3	3	3
# alive	23	23	23	23	23	23	22	22	22	22	22
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	1	1	1
2- 20ug TPA/ 25% DGME											
# dead	0	0	1	1	2	2	2	2	2	2	2
# alive	15	11	8	7	5	5	5	5	4	4	4
# w/SOA tumors	22	22	23	24	24	24	24	24	24	24	24
# w/NSOA tumors	9	12	11	11	10	10	10	10	11	11	12
SOA tumors mean	25.5	23.7	22.9	21.4	19.8	19.8	20.0	20.4	20.0	20.1	21.3
Min-max	1-49	1-36	2-43	1-45	1-29	1-29	1-29	1-29	1-29	1-29	5-29
3- 3% Dapsone / 25% DGME											
# dead	1	1	1	1	1	1	1	1	1	1	1
# alive	24	24	24	24	24	24	24	24	24	24	24
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	1	1	1	1	1	3	3	4	4	5	5
4- 5% Dapsone / 25% DGME											
# dead	1	2	2	2	2	2	2	4	5	5	10
# alive	24	23	23	23	23	23	23	21	20	20	15
# w/SOA tumors	1	1	1	1	1	1	1	1	1	1	1
# w/NSOA tumors	0	0	0	0	0	1	1	1	1	1	1
SOA tumors mean	1.0	1.0	1.0	1.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0
Min-max	1	1	1	1	1	2	2	2	2	2	2
5- 10% Dapsone / 25% DGME											
# dead	2	2	2	2	4	5	8	15	17	20	21
# alive	23	23	23	23	21	20	17	10	8	5	4
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	2	2	3	3	2	3	4	4	3	3	3
6- Acetone 0 ml/kg											
# dead	2	3	3	3	3	3	3	3	3	3	3
# alive	23	22	22	22	22	22	22	22	22	22	22
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0
7- 1.25ug TPA/Acetone											
# dead	2	2	2	2	2	2	2	2	2	2	2
# alive	8	8	8	8	8	4	3	3	2	2	1
# w/SOA tumors	9	9	9	9	9	9	9	9	9	9	9
# w/NSOA tumors	1	1	2	3	4	3	2	2	2	3	2
SOA tumors mean	11.6	13.8	6.0	18.1	21.1	18.6	18.2	18.1	18.2	18.2	18.2
Min-max	3-22	3-22	3-23	3-28	4-37	5-29	6-25	5-25	6-28	6-27	6-25
8- 5% Dapsone / Acetone											
# dead	1	1	1	2	2	2	5	8	18	19	21
# alive	24	24	24	23	23	23	20	17	7	6	4
# w/SOA tumors	0	0	0	0	0	0	0	0	0	0	0
# w/NSOA tumors	0	0	0	0	0	0	0	0	0	0	0

Table A.7.2. (cont.) Female Papilloma Tumor Incidence

	week							
	26	27	28	29	30	31	32	33
1- 25% DGME								
# dead	3	3	4	4	4	4	4	4
# alive	22	22	21	21	21	21	21	21
# w/SOA tumors	0	0	1	1	1	1	1	1
# w/NSOA tumors	1	2	3	3	6	5	6	6
SOA tumors mean	.	.	1.0	1.0	1.0	1.0	1.0	1.0
Min-max	.	.	1	1	1	1	1	1
2- 20ug TPA/ 25% DGME								
# dead	2	2
# alive	4	0
# w/SOA tumors	24	24
# w/NSOA tumors	10	10
SOA tumors mean	20.3	20.3
Min-max	1-29	1-29
3- 3% Dapsone / 25% DGME								
# dead	1	1	1	1	1	4	6	6
# alive	24	24	24	24	24	21	19	19
# w/SOA tumors	0	0	3	3	0	1	1	2
# w/NSOA tumors	6	6	7	8	9	8	8	8
SOA tumors mean	.	.	1.0	1.0	.	1.0	1.0	1.5
Min-max	.	.	1	1	.	1	1	1-2
4- 5% Dapsone / 25% DGME								
# dead	12	16	23	24	24	25	25	25
# alive	13	9	2	1	1	0	0	0
# w/SOA tumors	1	1	1	1	1	1	1	1
# w/NSOA tumors	2	2	1	1	1	1	1	1
SOA tumors mean	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Min-max	2	2	2	2	2	2	2	2
5- 10% Dapsone / 25% DGME								
# dead	21	23	23	23	24	25	25	25
# alive	4	2	2	2	1	0	0	0
# w/SOA tumors	0	0	0	0	0	0	0	0
# w/NSOA tumors	3	3	3	3	3	3	3	3
6- 0 ml/kg/Acetone								
# dead	3	3	3	3	3	3	.	.
# alive	22	22	22	22	22	22	.	.
# w/SOA tumors	0	0	0	0	0	0	.	.
# w/NSOA tumors	0	1	1	1	1	2	.	.
7- 1.25ug TPA/Acetone								
# dead	2	2
# alive	1	0
# w/SOA tumors	9	9
# w/NSOA tumors	2	2
SOA tumors mean	18.4	18.6
Min-max	6-25	6-25
8- 5% Dapsone / Acetone								
# dead	22	22	22	22	22	22	.	.
# alive	3	3	3	3	3	3	.	.
# w/SOA tumors	0	0	0	0	0	0	.	.
# w/NSOA tumors	0	0	0	0	0	0	.	.

Similar to the male mice, except for the TPA positive controls, there were almost no SOA tumors among the females, i.e., at the end of the study there was only one SOA papilloma in each of Group 1 (25% DGME vehicle) and Group 4 (5% Dapsone / 25% DGME), and two SOA papillomas in Group 4 (3% Dapsone / 25% DGME). However, within the Dapsone treatment groups and the Acetone group no Fisher exact test comparing the number of SOA tumors was statistically significant.

Thus for both sexes there is no evidence of dose differences among either the four Dapsone/DGME treatment groups or the two Dapsone/acetone groups.

Appendix 8. References

- Dunson, DB, JK Haseman, APJM van Birgelen, S Stasiewicz, and RW Tennant, R.W., 2000, Statistical Analysis of Skin Tumor Data from Tg.AC Mouse Bioassays, *Toxicological Sciences*, 55:293-302.
- Haseman, J. K. (1983), A Reexamination of false-positive rates for carcinogenicity studies, *Fundamental and Applied Toxicology*, 3: 334-339.
- Lin, K. K. and Rahman, M.A. (1998), Overall False Positive Rates in Tests for Linear Trend in Tumor Incidence in Animal Carcinogenicity Studies of New Drugs, *Journal of Biopharmaceutical Statistics*, 8(1):1-15.
- Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., Parrish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980) Guidelines for sample sensitive significance tests for carcinogenic effects in long-term animal experiments, *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, supplement 2: Long term and Short term Screening Assays for Carcinogens: A Critical Appraisal*, International Agency for Research Against Cancer, 311-426.

APPEARS THIS WAY
ON ORIGINAL

SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer: Steve Thomson
Date: May 11, 2005

Concurring Reviewer(s):

Statistical Team Leader: Mohamed Alesh, Ph.D., HFD-725

cc:

HFD-540/Frank Cross, JR
HFD-540/Norman See, Ph.D.
HFD-540/Paul Brown, Ph.D.
HFD-540/Jonathan Wilkin, M.D., Ph.D.
HFD-725/Steve Thomson
HFD-725/ Mohamed Alesh, Ph.D.
HFD-725/ Mohammed Huque, Ph.D.,
HFD-725/ Charles Anello, Ph.D.

c:\Data\Profiles\MyDocuments~\carcNDA50803DapsoneGelConnectics.doc

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

/s/

Steven Thomson
5/16/05 11:46:30 AM
BIOMETRICS

Mohamed Alesh
5/16/05 12:15:55 PM
BIOMETRICS
Concur with review