

Statistical Review and Evaluation

NDA Number: 20-886
Applicant: Ligand Pharmaceuticals
Name of Drug: Panretin (9-*cis*-retinoic acid) gel (0.1%)
Indication: First-line topical treatment of cutaneous lesions in patients with
acquired AIDS-related Kaposi's sarcoma.
Documents Reviewed: Vols. 1.120-1.123, 1.139-1.141, 1.157-1.164 dated 28 May 1998
Medical Reviewer: Robert White, M.D.
Statistical Reviewer: David Smith, Ph.D.

NOV 17 1998

APPEARS THIS WAY
ON ORIGINAL

1. Background and Overview

In order to support labeling for the indication of first-line treatment of cutaneous lesions in patients with acquired AIDS-related Kaposi's sarcoma (heretofore abbreviated KS), the sponsor submitted an NDA which is comprised of two Phase III trials and nine Phase I/II trials. The sponsor's submission included the reports of the two pivotal Phase III studies and the combined results of the nine supportive Phase I/II studies. We will only consider the pivotal studies in this review.

A brief summary of the pivotal studies appears below.

Study	Type	N	Arms
L1057T-31 (Study 31)	Randomized Ph III	134	Panretin gel
		134	Vehicle gel
AGN 192013/ALRT1057-503 (Study 503)	Randomized Ph III	36	Panretin gel
		46	Vehicle gel

The next section includes relevant statistical issues for these studies. The following sections will discuss these studies in more detail and will follow the following format:

1. General description of study
2. Efficacy endpoints and results
3. Summary and conclusions

The last two sections will include overall conclusions and recommendations for the submission.

References will follow the review.

2. Statistical Issues

The Study 503 protocol specified that 78 patients (39 per arm) were to be enrolled for the interim analysis. However, 82 patients were enrolled.

Study 503 was a trial that was stopped as a result of an interim analysis. However, there was an imbalance in the randomization which prevented the sponsor from attaining the number of patients specified in the protocol. If the study's enrollment agreed with the protocol, the trial may not have been stopped. If one considers the patients that were accrued after the trial was stopped to supplement the interim analysis, the statistically significant interim result does not remain.

It is likely that the treatment blind was broken since patients on Panretin experienced redness and swelling on treated lesions, whereas vehicle patients did not experience either of these symptoms.

In the quality of life analysis of Study 31, the small amount of data available among dropouts or patients who did not comply makes interpretation of the QOL questionnaire results in this sub-population particularly susceptible to biased conclusions (*i.e.*, it is difficult to discuss "trends" in such a small population). There is also little inferential power in this sub-population, and so in the QOL section below, we will concentrate on interpreting the results of the questionnaires filled out by those with substantial follow-up information.

For the quality of life questionnaire in Study 31, the question "How have you been feeling about your job, work?" was often answered as "not applicable" due to a substantial number of patients who were not employed. The sponsor subsequently asked the study staff to instruct patients to interpret "work" as "any routine daily activity", although a large proportion of patients continued to answer this question as "not applicable". Due to the misinterpretation of this question, we will not consider the

results of this question in the QOL section below. Additionally, the sponsor constructed a "sum of all questions" endpoint. This included the results of this problematic "job/work" question, and so we will not consider this endpoint either.

3. Pivotal Phase III Trials

3.1 Description of Study 31

Study Objective: To evaluate the efficacy of Panretin for the treatment of cutaneous KS lesions in HIV-positive patients.

Study Enrollment Period: Enrollment began April 1996 and last clinical evaluation for the present submission was October 1997.

Study Design: Double-blind, multi-center randomized vehicle-controlled parallel group Phase III study. Patients were randomized in a blinded fashion in a 1:1 allocation ratio to either Panretin or vehicle gel treatment at an initial application frequency of TID, escalated to QID as tolerated. The planned blinded treatment period was twelve weeks. Patients were required to have a follow-up evaluation at least four weeks following the last dose, and two further follow-up evaluations at three-month intervals.

No stratification factors were specified in the protocol.

Sample Size: The sponsor assumed a response rate of 25% for the Panretin group and 5% for the vehicle gel group. Under a two-sided Type I error of 0.05 and power of 0.80, 230 patients (115 in the Panretin arm and 115 in the placebo arm) would detect a significant difference between the response rates. The sponsor recalculated the study's sample size from the interim vehicle response result and adjusted enrollment upward to 268 patients (134 on each arm). Out of 268 patients accrued, 134 were assigned to the Panretin arm and 134 were assigned to the placebo arm.

Interim Analysis: The protocol specified a single interim analysis when 100 patients were accrued (50 on each arm). Based on a two-sided Type I Error level of 0.05, the interim significance level was 0.005 from O'Brien-Fleming. This plan was disregarded, when, at the interim analysis, it was decided that no comparison between the arms should take place. Instead, only the response rate of the vehicle arm was calculated for sample size re-estimation.

Dose: The Panretin arm received 0.1% 9-*cis*-retinoic acid applied TID to cutaneous KS lesions. The placebo arm received a vehicle gel applied TID to cutaneous KS lesions.

Criteria for Evaluation: The primary efficacy endpoint is the patient's cutaneous KS tumor response rate at 12 weeks of study drug treatment as determined by evaluating the group of KS index lesion assessments in the context of ACTG response criteria for KS index lesions. Specifically, patients meeting ACTG criteria for response up to and including 12 weeks of treatment but requiring up to 16 weeks or more of treatment to confirm the four week or more persistence of response will be included in the primary efficacy endpoint. Aspects of the ACTG criteria not pertaining to the designated group of cutaneous KS index lesions *per se* (e.g., other preexisting lesions, new lesions, non-index lesion-associated edema and pain, tumor-associated effusion and visceral disease) will not be a factor in the patient response assessments used for the primary efficacy endpoint.

In a protocol amendment dated 14 March 1996, the following covariates were to be considered as potential prognostic factors: study center, CD4 counts, Karnofsky Performance Status, anti-retroviral therapy, and baseline index lesion area and height. All prognostic factors were studied retrospectively.

Patients were generally well-balanced at baseline with respect to prognostic factors and laboratory values. There were imbalances in baseline chloride values ($p = 0.024$) and LDH ($p = 0.056$), but there were no imbalances with respect to systemic anti-KS therapy and prior antiretroviral therapy.

Efficacy Endpoints

The sponsor's assessment of response after the initial blinded phase of the study appears in Table 3.1. According to this assessment, 35.1% of patients treated with Panretin and 17.9% of patients treated with vehicle gel had a complete or partial response. The Fisher's exact p-value for response was 0.002. A secondary Mantel-Haenszel analysis was performed that adjusted simultaneously for baseline CD4+ counts and number of raised index lesions at baseline. The p-value for this analysis was 0.0019.

Interim results of the open-label phase of Study 31 also appear in Table 3.1. Out of the 134 patients initially assigned to Panretin, 90 patients continued treatment with Panretin, and 85 patients out of 134 initially assigned to the vehicle gel continued treatment with vehicle gel. One patient crossed over from Panretin to vehicle gel and 8 patients crossed over from vehicle gel to Panretin. In the open-label phase, the response rate of those continuing on Panretin was 66.7% and the response rate of vehicle gel was 29.4%.

Table 3.1. Response frequencies of patients in Study 31. Legend: P = Panretin; V = Vehicle.

Response	Initial 12 wks		Open-label Phase (Baseline Trt \Rightarrow Open-label Trt)*			
	P	V	P \Rightarrow P	V \Rightarrow V	P \Rightarrow V	V \Rightarrow P
Complete Rsp.	1	0	1	1	0	0
Partial Rsp.	46	24	59	24	0	5
Stable Dis.	67	79	16	54	1	2
Prog. Dis.	20	31	14	6	0	1
Total	134	134	90	85	1	8

*Interim results at last clinic evaluation (6 Oct 97).

This reviewer performed a secondary analysis on response rate while adjusting for use of protease inhibitors. Using a Mantel-Haenszel analysis on the sponsor's response data set, the p-value for response was 0.00144, favoring the Panretin arm.

Time to event endpoints for the blinded phase of Study 31 appear in Table 3.2. Comparisons among the open-label phase have been omitted due to sparse information.

Table 3.2. Time to event endpoints (median and interquartile range) for the blinded phase of Study 31 (intent-to-treat population). Units are in days.

Event	Time to Response	Durability of Response	Time to Progression
Panretin	34 (29 - 62)	55 (31 - 67)	64 (23 - 94)
Vehicle	33 (29 - 75)	57 (36 - 68)	47 (28 - 63)

FDA Medical Reviewer's Assessment of Response

The FDA medical reviewer also performed an assessment of response in the intent-to-treat population. The medical reviewer's assessment appears in Table 3.3.

Table 3.3. Comparison of medical reviewer's assessment of response versus the sponsor's assessment of response in Study 31.

Response	Sponsor's assessment		FDA assessment	
	Pan.	Veh.	Pan.	Veh.
Complete Rsp.	1	0	1	Not reviewed
Partial Rsp.	46	24	43	Not reviewed
Total per arm	134	134	134	134
Response rate	35.1%		33%	

Fisher's exact p-value for comparing response rates between Panretin and vehicle gel is 0.0074, assuming no differences between the sponsor's assessment of response on the vehicle arm and the FDA assessment of the vehicle arm.

Quality of Life

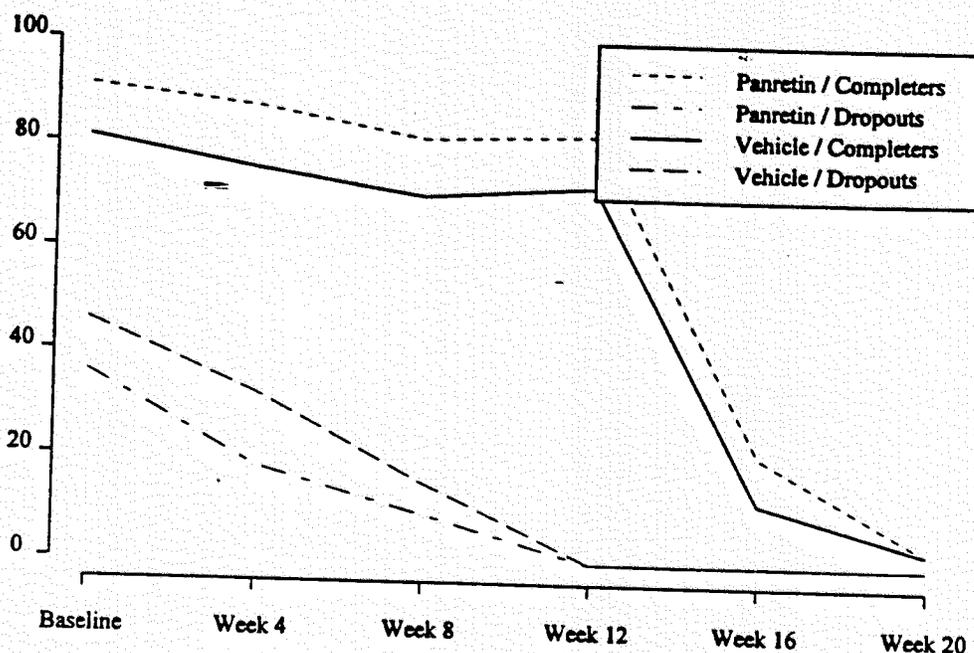
To assess quality of life (QOL), the sponsor used portions of the NIAID ACTG 286 questionnaire to collect each patient's emotions regarding their treatment. The questionnaire included 9 questions, and a patient responded by circling a number from either 1 to 10 (with 10 being "the very best I ever felt") or a number from 1 to 5 (with 5 being the most positive response on each question). The questions included "How have you been feeling overall on average in the past 4 weeks?" and "With respect to those KS lesions being treated with study drug, what is your level of satisfaction/dissatisfaction with your physical appearance over the past 4 weeks?"

In a classical univariate repeated ANOVA, a particular correlation structure known as compound symmetry must be assumed for a valid F-test of interaction of treatment and time. A multivariate approach may be considered when a compound symmetry assumption fails. However, in a multivariate approach, a distribution must be explicitly specified with the "correct" mean and covariance matrix.

The generalized estimating equation (GEE) approach was developed to cope with the potential problem of informative correlation among observations per subject. An advantage of a GEE approach is that it is not necessary to specify the correct correlation structure in advance. Using the idea of M-estimation theory (Huber, 1967; White, 1982; Liang and Zeger, 1986), the solution to the (potentially mis-specified) covariance matrix is consistent. Also, M-estimation protects the under-estimation of the covariance matrix by introducing "sandwich" estimators. Therefore, we have some assurance of a variance estimate that is robust.

The following analyses of the QOL data is based on a GEE linear model and derived a robust covariance estimator based on M-estimation theory. To deal with the problem of potentially informative dropout, the sponsor based the dropout analyses on the concept of a pattern-mixture model (Little, 1993 and 1995). Patients are grouped into two classes: those who drop out of the study before completing their 75th day of treatment, and those who remain on-study at least 75 days of treatment. We will call the former group "Dropouts" and the latter group "Completers". Compliance to the QOL questionnaire for Study 31 appears in Figure 1.

Figure 1. Compliance to the QOL questionnaire in Study 31. The units on the y-axis are number of questionnaires completed.



The small amount of data available among the Dropout group makes interpretation of the QOL questionnaire results particularly susceptible to biased conclusions (*i.e.*, it is difficult to discuss "trends" in such a small population). There is also little inferential power in this sub-population and so we will concentrate on interpreting the results of the questionnaires filled out by the Completer sub-population.

The question "How have you been feeling about your job, work?" was often answered as "not applicable" due to a substantial number of patients who were not employed. The sponsor subsequently asked the study staff to instruct patients to interpret "work" as "any routine daily activity", although a large proportion of patients continued to answer this question as "not applicable". Due to the misinterpretation of this question, we will not consider the results of this question further. Questions on employment rarely appear on QOL questionnaires for the aforementioned reasons, and this reviewer would not have considered the results of this question relevant regardless of any special instructions that the sponsor had given to the study staff on this question.

The p-values of the longitudinal analyses on the completer group appear in Table 3.4. There was no p-value adjustments for the multiplicity of subscales considered, although for most multiple comparisons adjustment procedures, the "physical appearance", "change in treated lesions" and "satisfaction of treatment of lesions" will remain statistically significant. Graphs for the three statistically significant subscales appear in Figures 2, 3, and 4.

From the graphs, it appears as if nearly all of the patient benefit from the Panretin arm occurred within the first four weeks of treatment. This may be due to the fact that patients were likely became unblinded to study treatment. Patients on Panretin experienced redness and swelling on treated lesions, whereas vehicle patients did not experience either of these symptoms. The redness and swelling appeared within the first 4 weeks of treatment in many cases, and so the QOL effects that were significant may reflect the patients' perceptions that they were being treated with an active agent.

Table 3.4. P-values for eight questions on the QOL questionnaire (the question on “feelings about work” has been omitted from this table).

Question	p-value
Overall feeling	0.154
Physical feeling	0.377
Emotional feeling	0.814
Personal life feeling	0.164
Physical appearance	0.0004
Lesions' interference with daily activities	0.211
Change in lesions since enrollment	0.0001
Satisfaction with treatment of lesions	0.0001

Figure 2. Mean scores to the question “For those lesions treated, what is your level of satisfaction with your physical appearance over the last four weeks?”.

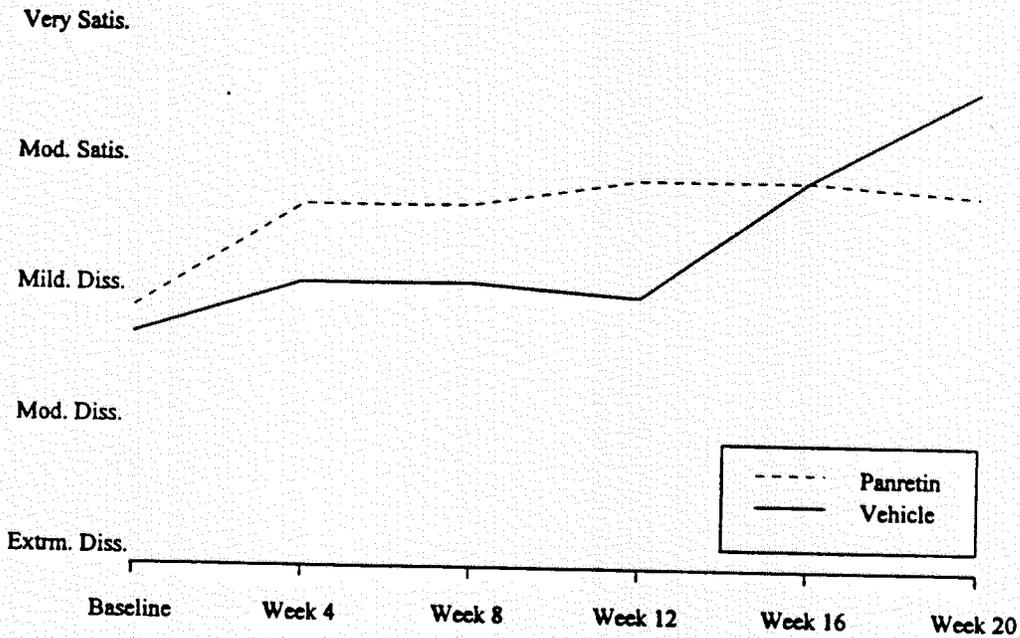


Figure 3. Mean scores to the question "For those lesions treated, rate the change in these lesions as compared to before your participation in this study."

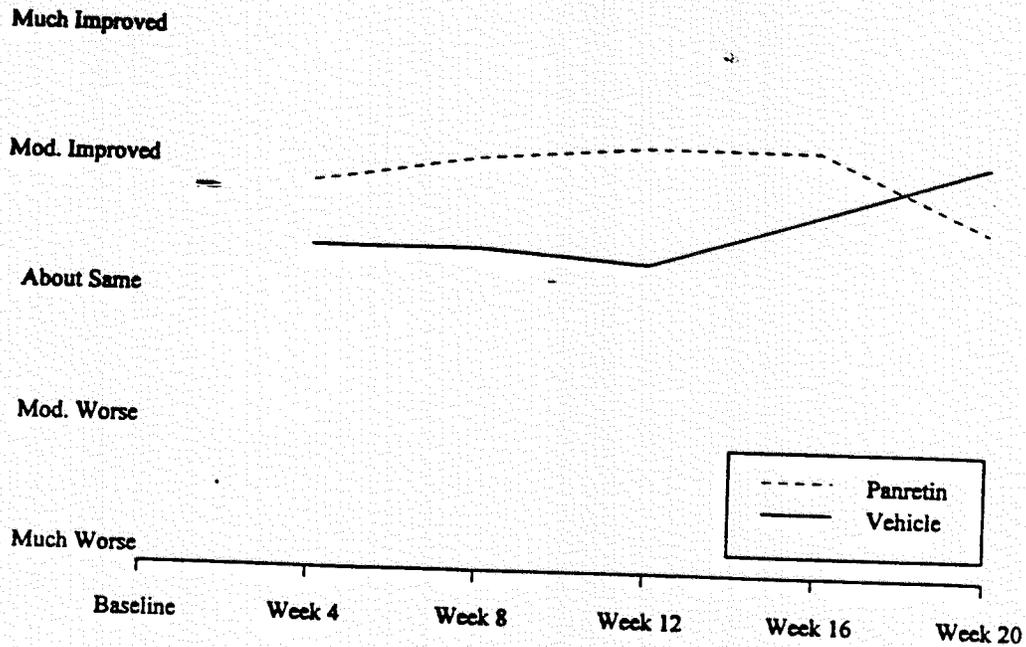
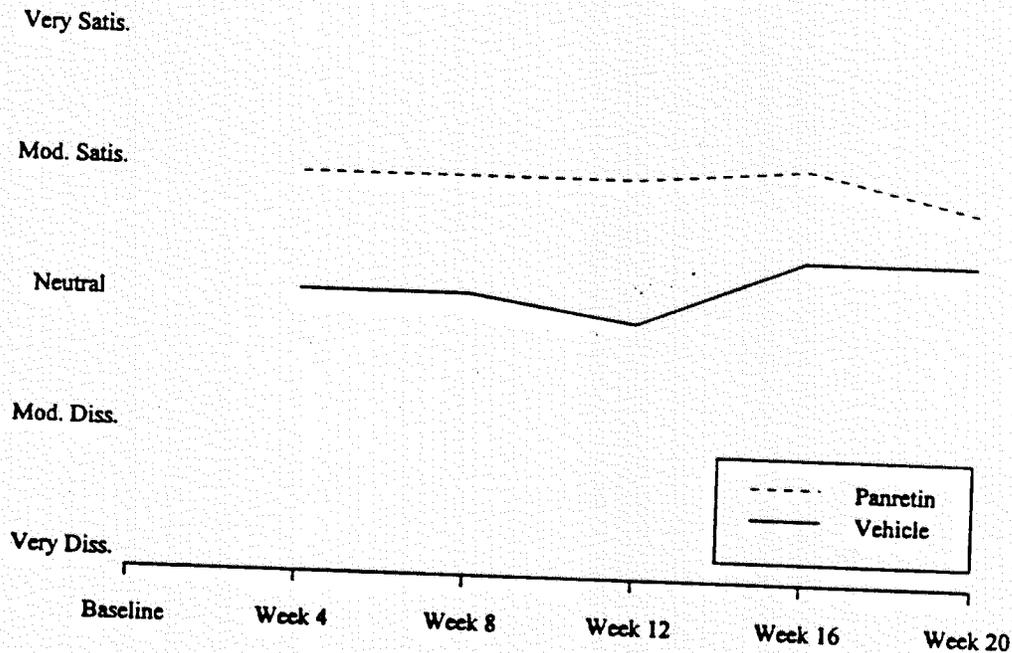


Figure 4. Mean scores to the question "For those lesions treated, what is your level of satisfaction with the study drug treatment over the last four weeks?"



Summary and Conclusions

Study 31 was a blinded, randomized, placebo-controlled Phase III trial. The initial blinded period of Study 31 lasted 12 weeks, and this period was followed by an open-label period. The patients were well-balanced with respect to the prognostic factors specified in the protocol. The sponsor's assessment of response showed a statistically significant advantage of response on the Panretin arm (35.1%) compared to the vehicle arm response (17.9%). The unadjusted Fisher's exact p-value for response was 0.002. When adjusted for prognostic factors such as use of protease inhibitors, prior systemic anti-KS therapy, prior antiretroviral therapy, and CD4+ counts at baseline, the statistically significant result remained.

The FDA medical officer also assessed response and found a 33% response rate in Panretin patients. For details on those patients for whom the FDA's assessment did not agree with the sponsor's assessment, please see the FDA Assessment of Efficacy: Study 31 section of the medical officer's review.

On the time to event endpoints (time to response, durability of response, and time to progression), there was little difference between the Panretin and vehicle arms, except that median time to progression was appreciably longer for the Panretin arm.

The patients filled out quality of life (QOL) questionnaires with specific questions regarding their physical appearance and other subjective aspects of their treatment. Although one question was poorly phrased and there was evidence of informative drop-out, there were statistically significant findings on "Satisfaction of physical appearance", "Change in treated lesions" and "Satisfaction with study drug" questions. Most of the QOL benefit for these three subscales were realized within the first 4 weeks of treatment, and this may be a reflection of the fact that the treatment blind was broken.

3.2 Description of Study 503

Study Objective: To evaluate the efficacy of Panretin for the treatment of cutaneous KS lesions in HIV-positive patients.

Study Enrollment Period: September 1996 to September 1997

Study Design: Double-blind, multi-center randomized vehicle-controlled parallel group Phase III study. Patients were randomized in a blinded fashion in a 1:1 allocation ratio to either Panretin or vehicle gel treatment at an initial application frequency of BID. The planned blinded treatment period was twelve weeks. After the initial blinded period, patients were given the option of continuing on Panretin on an open-label basis.

Sample Size: The sponsor assumed a response rate of 25% for the Panretin group and 5% for the vehicle gel group. Under a two-sided Type I error of 0.05 and power of 0.80, 270 patients (135 in the Panretin arm and 135 in the placebo arm) would detect a significant difference between the response rates. The study was terminated after 82 patients were accrued. Out of the 82 patients, 36 were assigned to the Panretin arm and 46 were assigned to the placebo arm. See the special section on the **Interim Analysis** below.

Dose: The Panretin arm received 0.1% 9-*cis*-retinoic acid applied BID to cutaneous KS lesions. The placebo arm received a vehicle gel applied twice-daily to cutaneous KS lesions.

Criteria for Evaluation: The primary efficacy endpoint was KS lesion response rate in the intent-to-treat population. Index lesions were selected at the beginning of the study and response was sequentially measured, based on the AIDS Clinical Trial Group criteria. The protocol specified comparing the overall response rate of the treatment arm against the placebo arm using a chi-square

test. The prognostic factors specified in the protocol were resistance to prior 5FU therapy, duration of prior 5FU therapy, age, performance status, visceral involvement, number of metastatic sites by organ, intent of and response to prior chemotherapy.

Patients were well-balanced at baseline with respect to prognostic factors and laboratory values. There were no imbalances in baseline systemic anti-KS therapy, number and size of index KS lesions, presence of visceral KS, or prior antiretroviral therapy.

Efficacy Endpoints

The sponsor's assessment of response after the initial blinded phase of the study appears in Table 3.5. According to this assessment, 41.7% of patients treated with Panretin and 6.5% of patients treated with vehicle gel had a complete or partial response. The interim significance level was 0.00025. Fisher's exact p-value for response was 0.00027, which is very close to the interim significance level. Fisher's exact test is a preferable test for the small number of responses in the vehicle arm, as opposed to the chi-square test, which was the test specified in the protocol. A secondary Mantel-Haenszel analysis was performed that adjusted for baseline CD4+ counts. The p-value for this analysis was 0.00006.

Table 3.5. Response frequencies of patients in Study 503.

Response	Panretin	Vehicle
Complete Rsp.	1	0
Partial Rsp.	14	3
Stable Dis.	16	27
Prog. Dis.	5	16
Total	36	46

This reviewer performed a secondary analysis on response rate while adjusting for use of protease inhibitors. Using a Mantel-Haenszel analysis on the sponsor's response data set, the p-value for response was less than 0.0001, favoring the Panretin arm.

Time to event endpoints for the blinded phase of Study 503 appear in Table 3.6. It is interesting to note that the median time to progression for the Panretin arm is two weeks shorter than the time to progression of the vehicle arm. The exact opposite trend occurred for time to progression in study 31.

Table 3.6. Time to event endpoints (median and interquartile range) for Study 503 (intent-to-treat population). Units are in days.

Event	Time to Response	Durability of Response	Time to Progression
Panretin	28 (14 - 35)	64 (57 - 71)	30 (16 - 85)
Vehicle	24 (14 - 39)	57 (48 - 85)	45 (17 - 57)

Quality of Life

QOL information was not collected for this study.

Interim Analysis

The protocol specified a total sample size of 270 patients. The interim analysis was based on a Panretin response rate of 50% and 10% response in the placebo group at 80% power. The protocol also specified that a total of 78 patients were to be enrolled for the interim analysis (39 patients per arm). This trial was stopped after 82 patients were enrolled because of superior efficacy results in favor of the Panretin arm. This reviewer had several concerns about the interim analysis.

The protocol stated that the interim analysis would occur after 39 patients per arm were accrued (78 patients total). The protocol incorrectly reported the significance level of the interim analysis as 0.005; the correct O'Brien-Fleming interim p-value should have been 0.00025 based on the amount of information collected at the interim. The 0.005 interim significance level is only correct when half of the total number of patients are analyzed at the interim. In this case, the sponsor intended to analyze 30.4% of the total number of patients at the interim and this was why the significance level should have been smaller than what was reported in the protocol.

Aside from this error in the protocol, there is an unanswered question as to the intended sample size on each arm at the interim. The protocol stated that 39 patients per arm would be enrolled for the interim analysis and that sample size would not be adjusted downward. At the interim analysis, a total of 82 patients were analyzed (36 in the Panretin arm and 46 in the vehicle arm). The departure from the randomization specified in the protocol is explained by the fact that there was a tendency to allocate more patients to the vehicle arm in the first treatment assignment of each randomization block. The randomization block size for this study was four patients, and 10 out of 17 centers enrolled fewer than four patients. This reviewer could not find explicit details on the randomization scheme (whether central or otherwise) among the reviewed volumes. See Table 3.7 for treatment assignment by center.

Table 3.7. Randomization of patients to treatment arm by center.

Center	Pan. N	Veh. N
0101	4	4
0105	1	2
0106	0	1
0202	1	2
0203	0	1
0207	2	1
0401	2	2
0402	4	6
0403	14	14
0404	0	1
0405	2	3
0501	1	2
0502	2	2
0503	0	1
0505	0	1
0509	1	1
0510	2	2

The sponsor examined the randomization code and found no imbalance, although the randomization realization of this trial is unusual. A randomization imbalance was not present in Study 31.

Lan and Zucker (1993) discussed the role of statistical information and Brownian motion for sequentially monitoring clinical trials. Specifically, they developed a general framework for sequential monitoring when the numbers of patients are unequal among two treatment groups at an interim analysis. Under the Lan and Zucker statistical information calculation and the O'Brien-Fleming spending function, the interim significance level for an allocation of 36 and 46 patients in two arms is 0.00033. One important assumption about this approach is that this must be the very first interim analysis performed thus far in the trial. When this significance level is compared to Fisher's exact p-value of 0.00027, we see that there is a statistically significant difference between the Panretin and vehicle arms. See Table 3.8 for this result and other analyses described below.

In addition to the 82 patients enrolled at the interim analysis, a further 52 patients were enrolled before enrollment was stopped in August 1997. When one compares the sequence of patient entry dates of the 52 patient cohort to the 82 patient cohort, one finds that some patients on the 52 patient cohort were enrolled before the last patient on the 82 patient cohort was enrolled. This implies that follow-up on the 52 patient cohort was going on before enrollment on the 82 patient cohort was completed.

In an exploratory analysis, this reviewer followed the interim analysis plan specified in the protocol and calculated Fisher's exact p-value at the interim when one considers the responses of the first 39 Panretin patients versus the responses of the first 39 vehicle patients (in order of enrollment, inclusive of the 52 patient cohort). The p-value for this analysis is 0.0024, which is not significant at an alpha level of 0.00025.

If one compares the sponsor's response rates based on the first 78 patients enrolled, 13 out of 33 Panretin patients responded and 3 out of 45 vehicle patients responded. Fisher's exact p-value for this comparison is 0.00056. We again must adjust the significance level for the imbalance between treatment and placebo allocation since the alpha level of 0.00025 is only applicable to equal numbers of patients in both arms. Using the scheme outlined by Lan and Zucker (1993) for interim significance levels and assuming that this is the first interim look, the alpha level is 0.00022. Note that this significance level is slightly smaller than 0.00025, which assumes equal sample sizes. Under this analysis, there is no statistically significant difference between treatment arms and the trial should not have been stopped early.

The first three Panretin patients enrolled in the 52-patient cohort were not responders. An analysis based on the first 39 (36 + 3) Panretin patients and the first 46 patients enrolled on the vehicle arm would result in a Fisher's exact p-value of 0.00041, based on the sponsor's analysis of response. Assuming that this was the first interim analysis, the significance level based on Lan and Zucker (1993) for this sample allocation is 0.00045. For this analysis, there is a statistically significant difference between arms.

Table 3.8. P-values and significance levels with number of responders and sample sizes for various interim analyses. The significance levels are based on the O'Brien-Fleming spending function and the Lan-Zucker information calculation.

Rationale for analysis	Resp / N		Fisher's exact		
	Pan.	Veh.	p-value	Signif. level	Conclusion
Sponsor's report	15 / 36	3 / 46	0.00027	0.00033	Significant
First 78 pts in equal allocation	15 / 39	3 / 39	0.00244	0.00025	Not Significant
First 78 pts in unequal allocation	13 / 33	3 / 45	0.00056	0.00022	Not Significant
Minimum of 39 pts per arm	15 / 39	3 / 46	0.00041	0.00045	Significant

FDA Medical Reviewer's Assessment of Response

The FDA medical reviewer also performed an assessment of response in the intent-to-treat population. The medical reviewer's assessment appears in Table 3.9.

Table 3.9. Comparison of medical reviewer's assessment of response versus the sponsor's assessment of response in Study 503.

Response	Sponsor's assessment		FDA assessment	
	Pan.	Veh.	Pan.	Veh.
Complete Rsp.	1	0	0	0
Partial Rsp.	14	3	14	3
Total per arm	36	46	36	46
Response rate	39%		6.5%	

Fisher's exact p-value for comparing response rates between Panretin and vehicle gel is 0.00062, which is not statistically significant under a O'Brien-Fleming spending function and Lan-Zucker's information calculation (nominal significance level = 0.00033).

Summary and Conclusions

Study 503 was a blinded, randomized, placebo-controlled Phase III trial. This trial was stopped early on the basis of superior efficacy of the Panretin arm. The patients were well-balanced with respect to the prognostic factors specified in the protocol. The sponsor's assessment of response showed a statistically significant advantage of response on the Panretin arm (41.7%) compared to the vehicle arm response (6.5%). The unadjusted Fisher's exact p-value for response was 0.00027 compared with an interim significance level of 0.00033.

This reviewer had some concerns with respect to the interim analysis. Namely, more patients in the placebo arm and fewer patients in the treatment arm than the numbers specified in the protocol were enrolled and there was an imbalance in the randomization which prevented the sponsor from attaining the number of patients specified in the protocol. If the study's enrollment agreed with the protocol, the trial may not have been stopped; this seems to be the case in light of the fact that when one uses the patients that were accrued after the trial was stopped to supplement the interim analysis, the statistically significant interim result does not remain.

The FDA medical officer's assessment of response did not completely agree with the sponsor's assessment. Specifically, the medical officer concluded that a complete responder should have been categorized as a partial responder and a partial responder was disqualified as a responder. This discrepancy between the sponsor's assessment of response and the medical officer's assessment of response changes the conclusion of whether there is a statistically significant difference between the Panretin arm and the vehicle arm at the time of the interim analysis.

On the time to event endpoints (time to response, durability of response, and time to progression), there was little difference between the Panretin and vehicle arms, except that median time to progression was appreciably longer for the vehicle arm. This is somewhat unexpected considering the superiority of response rate in the sponsor's assessment. This is also inconsistent with the findings of Study 31.

4. Summary and Conclusions

These studies were designed to test the superiority in the primary endpoint of response rate of Panretin gel versus a placebo control of vehicle gel. In the intent-to-treat population of the 31 Phase III trial, the sponsor's assessment of response showed a statistically significant advantage of response on the Panretin arm (35.1%) compared to the vehicle arm response (17.9%). The unadjusted Fisher's exact p-value for response was 0.002. In the sponsor's assessment of response in Study 503, there was a statistically significant advantage of response on the Panretin arm (41.7%) compared to the vehicle

arm response (6.5%). The unadjusted Fisher's exact p-value for response was 0.00027 compared with an interim significance level of 0.00033.

This reviewer had concerns with respect to the interim analysis of Study 503. There was a randomization imbalance which prevented the sponsor from attaining the number of patients specified in the protocol. The protocol specified enrolling 39 patients per arm, and at the interim analysis, there were 36 and 46 enrolled on the Panretin and vehicle arms, respectively. If the study's enrollment agreed with the protocol, the trial may not have been stopped.

There were differences between the medical reviewer's assessments of response and the sponsor's assessments for both studies. In both studies, the medical reviewer's assessment found fewer responders than the sponsor. In Study 31, there was no change in the overall conclusion of a statistically significant Panretin effect. However, in Study 503, if the medical reviewer's assessment would have been used as evidence for stopping the trial early, the trial would not have been stopped and one would conclude that there was no statistically significant difference between the arms.

There was a stronger trend toward a Panretin response rate advantage in Study 503 compared to Study 31, although Panretin was applied BID in 503 and TID in 31. There is reason to doubt the genuineness of the responders in 503 based on the medical officer's review and inconsistencies between response and some of the secondary endpoints. However, there is more consistency between response and the secondary endpoints in Study 31.

5. Overall Recommendations and Conclusions

In the two Phase III trials included in this submission, response rate was the primary endpoint. The results of Study 31 support the conclusion that Panretin offers a response benefit in patients with AIDS-related KS. This reviewer has questions as to the robustness of the interim result and concerns about the authenticity of the high response rate for Study 503, in light of the medical reviewer's analyses.

It is this reviewer's opinion that Panretin has demonstrated efficacy for the proposed indication in one Phase III trial, but doubts remain as to the appropriateness of the interim analysis and robustness of the response rates in the trial that was stopped early.

DS

David Smith, Ph.D.
Mathematical Statistician