

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

20-154/S-032, S-033

20-155/S-023, S-024

20-156/S-024, S-025

STATISTICAL REVIEW(S)

AUG 22 2000

Statistical Review and Evaluation

NDA#: 20154 SE08
APPLICANT: Bristol-Myers Squibb Company
NAME OF DRUG: Videx® (didanosine) Chewable/Dispersible Buffered Tablets
INDICATION: Treatment of HIV infection
DOCUMENTS REVIEWED: Submission dated 3/21/2000
Clinical Reviewer: HFD-530: Russ Fleischer, P.A.C, MPH

A: Background

Didanosine (ddI) is a nucleoside analogue available in tablets in strengths of 25, 50, 100, 150 and 200 mg of didanosine. 200mg QD was approved based on 24 weeks or shorter studies. This submission provide 48 weeks update for one of the pivotal studies, AI454-148.

Protocol

AI454-148

Title: "A Randomized Study of the Long-Term Suppression of Plasma HIV RNA Levels by Triple Combination Regimens in Treatment Naïve Subjects"

This is a multinational open-label study in anti-retroviral naïve HIV-infected subjects 12 (18 in Europe and Canada) years of age or older whose screening plasma HIV RNA level is ≥ 2000 copies/mL and whose CD4 cell count is ≥ 100 cells/mm³. Seven hundred subjects were to be randomized 2:1 to the following two treatment groups:

- Group 1: ddI (400mg QD) + d4T (40mg BID) + NLF (750mg TID) if weigh ≥ 60 Kg
ddI (250mg QD) + d4T (30mg BID) + NLF (750mg TID) if weigh < 60 Kg
- Group 2: ZDV (300mg BID) + 3TC (150mg BID) + NLF (750mg TID)

The randomization was to be balanced by plasma HIV RNA level ($< 30,000$ vs. $> 30,000$ copies/mL) and investigative site using the method proposed by Pocock and Simon with a centralized randomization code.

Dosing was planned for a period of at least 48 weeks after enrollment of the last subject. An early analysis was planned to occur when approximately 200 subjects have been treated for 6 months after randomization.

Participation in this trial may be discontinued for the following reasons:

- Increase in viral load to detectable levels (confirmed in a repeat assay, at least one value ≥ 1000 copies/mL).

- Major toxicity or pregnancy or use of prohibited medication.

Plasma HIV RNA level and CD4 cell counts were to be determined at screening (twice), day 1, Week 4, Week 8 and every 8 weeks thereafter, and at the final or early termination. Subjects who discontinued the randomized treatment were to be followed in the same manner.

The primary efficacy endpoint is the proportion of patients with HIV RNA < 400 copies/mL at Week 48. For the planned early submission the Week 24 data will be used instead. The secondary analysis includes time to viral load failure. Failure is defined as

- Failure to reach viral load <400 copies/mL by Week 24, or
- Experiencing AIDS-defining events or death, or
- Discontinued the randomized treatment except for switching between ZDV and d4T, or
- Confirmed rebound to detectable levels with at least one of which is >1000 copies/mL).

For the analysis of proportions below 400 copies/mL, subjects who discontinued the randomized treatment or lost to follow-up are regarded as failures. Missing values are regarded as failures unless bracketed by two values <400 copies/mL. The analysis uses an observation window of 8 weeks. The estimates and 95% confidence interval for the difference of proportions is adjusted for the HIV RNA strata with weights inversely proportional to within strata variance. A non-inferiority limit of 12% is used.

For the analysis of time to viral load failure, plots based on Kaplan-Meier estimates and Cox proportional hazards models are used to assess the treatment differences. The time averaged difference between the two groups in change from baseline (using \log_{10} scale for HIV RNA level) is analyzed using repeated measures model with a compound symmetry covariance structure for plasma HIV RNA levels and CD4 cell counts.

All treated subjects are included in the analysis.

The trial design provides at least 90% power to demonstrate the equivalence of the two arms when the common response rate is assumed to be 75% in both arms.

B. Results of the Applicant's Analyses

Due to a modification of the primary endpoint by FDA, the reviewer will not discuss in depth on the original primary endpoint. The sponsor results on the original endpoint are summarized below.

Table 1. HIV RNA <400 copies/mL

All subjects randomized, Missing as failures unless bracketed by two <400 copies/mL

Subset	HIV RNA < 400 copies/mL / Total (%)	
	Treatment Regimen	
	ddI/d4T/NFV N=503	ZDV/3TC/NFV N=253
All	263/503 (52)	143/253 (57)
Qualifying HIV RNA subsets		
<30,000 copies/mL	105/181 (58)	58/92 (63)
≥30,000 copies/mL	158/322 (49)	85/161 (53)

The treatment difference of the two arms is -4% with 95% confidence interval (-11.7%, 3.3%). Both the estimate and the 95% confidence interval for the difference of proportions are adjusted for the HIV RNA strata with weights inversely proportional to within strata variance.

Table 2. HIV RNA <50 copies/mL

All subjects randomized, Missing as failures unless bracketed by two <50 copies/mL

Subset	HIV RNA < 50 copies/mL / Total (%)	
	Treatment Regimen	
	ddI/d4T/NFV N=503	ZDV/3TC/NFV N=253
All	184/503 (37)	111/253 (44)
Qualifying HIV RNA subsets		
<30,000 copies/mL	74/181 (41)	47/92 (51)
≥30,000 copies/mL	110/322 (34)	64/161 (40)

The treatment difference of the two arms is -7% with 95% confidence interval (-14.6%, 0.1%).

Based on this endpoint, we see that ddI/d4T/NFV arm can be 12% worse than ZDV/3TC/NFV arm for proportions <400 copies/mL and approximately 15% worse for the proportions < 50 copies/mL using ultra sensitive assay.

The response rates are consistently higher among subjects with lower screening viral load, but the treatment differences are consistent across the HIV RNA strata.

Other analyses will be combined with the reviewer's analyses.

C. Statistical Reviewer's Comments

Modification of the primary endpoint

Traditionally time to failure is used for testing superiority while crude proportions below LOQ is used for testing equivalence in HIV studies, and failures are defined differently in these two analyses. DAVDP has recently modified the definition for the proportion analysis and the algorithm for the time to virologic failure analysis, and the new proportion analysis incorporates failures defined by time to virologic failure analysis. See Appendix for the new definition. The reviewer's analyses will be based on these new definitions.

Subjects who did not initiate treatment

There was an imbalance in proportions of subjects who did not initiate the assigned treatment: 4% in the ddI-containing arm and 2% in the control arm (p-value=0.15 for the difference). Since this is an open-label trial, it is possible that the decision of not initiating study drugs relates to the treatment assigned. Both analyses treating these subjects as failures and analyses excluding them are provided. The conclusions based on the two analyses are consistent. The analysis of treating not treated as failures will be considered primary.

Reasons for failures

Week 48 failures are classified according to the primary reason for the earliest cause of failure. The treatment difference between the two treatment arms is mostly accounted by virologic rebound (subjects who achieved confirmed plasma HIV RNA <LOQ and then have two consecutive measurements >LOQ).

Worsening of treatment difference over time

There is a worsening of the efficacy results over time. For example, the treatment difference increased from 5% at Week 24 to 9% at Week 48, both favoring the control arm. Further curves for proportions of responders over time showed a steady trend of increased gap between the two arms favoring the control arm.

D. Statistical Reviewer's Analyses

Baseline Characteristics

Seven hundred fifty six subjects were randomized and 730 of them initiated the treatment. Of all subjects randomized, 71% are male, 26% are black, 56% are white and 14% are Hispanic. The mean age is 34.7 years old (range 17 – 70). The median baseline HIV₁ RNA level is 4.69 log₁₀ copies/mL and the median baseline CD4 is 340 cells/mm³. 36% of the subjects have baseline HIV RNA values <30,000 copies/mL while the remaining 64% have values ≥ 30,000 copies/mL.

Subject Accountability

The following table presents the disposition for all subjects based on all data collected, including post week 48 information.

Table 3: Subject Status and Reason Discontinued by Treatment Group and Study

Treatment	ddI/d4T/NLF	ZDV/3TC/NFL
Total Randomized	503	253
No. Never Started Treatment	21	5
No. Started Treatment	482	248
No. discontinued randomized treatment	190	95
Lost to Follow-up	51	23
Subject Withdrew	18	13
Adverse Event	59	32
Non-compliance	11	8
Disease Progression or Relapse	24	11
Death	3	2
Pregnancy	2	1

Based on tables on pages 78 of Vol.1 and dataset submitted.

Among all 756 subjects randomized, 26 subjects (3.4%) never started the randomized treatment. This rate is higher in the ddI/d4T/NLF arm (4%) than in the ZDV/3TC/NLF arm (2%). The rates and pattern of discontinuation are similar between the two arms.

The Table below summaries reasons for discontinuation or disease progression for subjects who discontinued or had disease progression before Week 48.

Appears This Way
On Original

**Table 4: Discontinuations and Disease Progression Before Week 48
All Randomized Subjects**

Treatment	Ddl/d4T/NLF	ZDV/3TC/NFL
Number of subjects (%)	503	253
No. never started treatment	21 (4)	4 (2)
Discontinued with last RNA value before Week 48	136 (27)	59 (23)
Lost to Follow-up	38 (8)	16 (6)
Subject Withdrew	13 (3)	10 (4)
Adverse Event	37 (7)	12 (5)
Non-compliance	8 (2)	8 (3)
Disease Progression or Relapse	14 (3)	6 (2)
Death	3 (<1)	1 (<1)
Pregnancy	2 (<1)	1 (<1)
Disease progression before Week 48	7 (1)	1 (<1)
Completed 48 weeks	5 (1)	0 (0)
Discontinued due to AE prior to disease progression	1 (<1)	0 (0)
Discontinued due to disease progression or relapse	1 (<1)	1 (<1)

Reviewer's calculation based on data submitted. Before Week 48 means before day 274, the first day for Week 48 window.

A higher proportion of subjects discontinued in the ddI-containing arm (27%) compared to the control (23%). This difference is consistent across the reasons for discontinuation. However, more subjects had disease progression in ddI-containing regimen than the control (p-value=0.14 using Cochran-Mantel-Haenszel test).

Efficacy Endpoints

Instead of crude proportions < 400 copies/mL as the primary endpoint as planned in the protocol, FDA has modified the endpoint to regard early virologic failures as Week 48 failures. Therefore, subjects who experienced death, disease progression, viral rebound (two consecutive measurement > limit of quantification (LOQ)) after viral response (two consecutive measurement < LOQ), premature discontinuation, or lost to follow up before Week 48 measurement will be regarded as failures. Only subjects who achieved confirmed viral load < LOQ without experiencing virologic failures will be regarded as responders. The tables below display the results based on this new endpoint.

**Table 5: Week 48 HIV RNA Status using <400 copies/mL
All Randomized Subjects**

Subset	Responders at Week 48	
	ddl regimen N = 503	3TC regimen N = 253
	Number of responders/ Total subjects (%)	Number of responders / Total subjects (%)
All Subjects	250/503 (50)	150/253 (59)
Qualifying HIV RNA subset		
< 30,000 c/mL	105/181 (58)	61/92 (66)
≥ 30,000 c/mL	145/322 (45)	89/161 (55)

The treatment differences and associated 95% confidence intervals are:

**Table 6: Difference in Proportion of Responders at Week 48 <400 copies/mL
All Randomized Subjects**

	Difference in proportions ddl regimen – 3TC regimen	
	Estimate	95% CI
Overall stratified by HIV RNA	-9.5%	(-17.0%, -2.1%)
Qualifying HIV RNA subset		
< 30,000 c/mL	-8.3%	(-20.5%, 4.0%)
≥ 30,000 c/mL	-10.3%	(-19.7%, -0.8%)

The treatment difference is 9.5% favoring the control arm. This difference is statistically significant with two-sided p-value 0.012. The subgroup analysis by HIV RNA strata showed this difference is consistent in the two strata, even though the overall response rate is higher in the stratum with lower viral load. In fact, The difference in response rate between the two strata is estimated to be 12% favoring the lower stratum with 95% confidence interval (5%, 20%) and p-value <0.001 (stratified by treatment arms).

The calculation above is based on a stratified analysis with weights for each stratum proportional to the within stratum harmonic mean of the sample sizes of the two arms. This differs from the sponsor analysis where weights are inversely proportional to within strata variance. The first method is more robust against outlier strata while the second method could be influenced more by centers whose variances are underestimated. However, due to the large sample sizes of the strata in this study these two approaches should produce similar results. In fact, the tables above are identical to the sponsor results.

The analysis based on the ultrasensitive assay, which has a lower LOQ of < 50 copies/mL, is summarized below.

**Table 7: Week 48 HIV RNA Status using < 50 copies/mL
All Randomized Subjects**

Subset	Responders	
	ddl regimen N = 503	3TC regimen N = 253
	Number of responders / Total subjects (%)	Number of responders / Total subjects (%)
All Subjects	169/503 (34)	119/253 (47)
Qualifying HIV RNA subset		
< 30,000 c/mL	71/181 (39)	48/92 (52)
\geq 30,000 c/mL	98/322 (30)	71/161 (44)

**Table 8: Difference in Proportion of Responders at Week 48 using < 50 copies/mL
All Randomized Subjects**

	Difference in proportions ddl regimen - 3TC regimen	
	Estimate	95% CI
Overall stratified by HIV RNA	-13.4%	(-20.8%, -6.0%)
Qualifying HIV RNA subset		
< 30,000 c/mL	-13.0%	(-25.4%, -0.5%)
\geq 30,000 c/mL	-13.7%	(-22.7%, -4.6%)

This tables is nearly identical to the sponsor's results, the difference comes from the aforementioned differences in the statistical methods used.

The treatment difference is even wider in this analysis than the previous analysis using < 400 copies/mL. The difference of 13.4% favoring the control arm is statistically significant (p-value < 0.001). Again, this difference is consistent across the two HIV RNA stratum, and the lower stratum has a significantly higher response rate (difference of 9% with p-value 0.02 and 95% confidence interval (1%, 16%)).

Failures at Week 48 may have experienced one or more of the following: death or disease progression, premature discontinuation, lost to follow-up, or viral relapse. The table below classifies these subjects according to the reason for the earliest event meets the definition for failure.

Table 9: Outcomes of Randomized Subjects Through Week 48

Week 48 Status	Percent of Patients Using <400 copies/mL (50)	
	ddI/d4T/NFV N=503	3TC/ZDV/NFV N=253
Responder ^a	50 (34)	59 (47)
Virologic failure ^b	36 (57)	32 (48)
Death or disease progression	<1 (<1)	1 (<1)
Discontinued due to AE	4 (2)	2 (<1)
Discontinued due to others ^c	6 (3)	4 (2)
Never initiated treatment	4 (4)	2 (2)

^a: Subjects achieved virologic response (two consecutive viral load <400 (<50) copies/mL) and maintained it to Week 48.

^b: Includes viral rebound and failing to achieved confirmed <400 (<50) copies/mL by Week 48.

^c: Includes lost to follow up, non-compliance, withdrawal and pregnancy.

From this table we see that of the 9% treatment difference using LOQ= —, 13% using LOQ= —, 4% (9%) are due to the differences in virologic failures, other differences comes from discontinuations and not initiating the treatment.

In the above table, subjects who never achieved confirmed <LOQ status but discontinued before Week 48 are regarded as virologic failures, even though their reason for discontinuation may not be viral relapse. The table below lists the reasons for discontinuation for these subjects, stratified by the time of discontinuation (before or at Week 24 vs. after Week 24).

**Table 9: Reason for Discontinuation by Time
Among Subjects Discontinued Before Week 48 and Never Achieved Virologic Response**

	Percent of Patients Using <400 copies/mL (50)			
	Discontinued Before or at Week 24		Discontinued Between Week 24 and Week 48	
	ddI/d4T/NFV N=503	3TC/ZDV/NFV N=253	ddI/d4T/NFV N=503	3TC/ZDV/NFV N=253
Subjects never achieved Virologic Response	13 (16)	16 (18)	2 (3)	1 (1)
Virologic Failure*	1 (1)	1 (1)	1 (2)	1 (<1)
Death or Disease Progression	<1 (<1)	0 (0)	0 (0)	0 (0)
Discontinued due to AE	4 (5)	3 (4)	<1 (<1)	0 (<1)
Discontinued due to others	8 (10)	11 (13)	<1 (<1)	0 (0)

* By protocol definition where no confirmation is required for virologic response.

We see that around 90% of the subjects who discontinued before Week 48 and have not achieved virologic response did so before or at Week 24. For subjects discontinued before or at Week 24, most did so because of AE, noncompliance, lost to follow up, withdrawal or pregnancy. This is in contrast to the small number of subjects who discontinued after week 24, where majority of

them discontinued because of virologic rebound. Based on this observation, and the consideration that subjects who discontinued before or at Week 24 may not have had long enough exposure to the treatment to reach viral response, it is reasonable to classify subjects discontinued after Week 24 without achieving virologic response as virologic failures, but classify subjects who discontinued before or at Week 24 according to the reasons for discontinuation. The table below is based on such a classification.

Table 10: Outcomes of Randomized Subjects Through Week 48

Week 48 Status	Percent of Patients Using <400 copies/mL (50)	
	ddI/d4T/NFV N=503	3TC/ZDV/NFV N=253
Responder ^b	50 (34)	59 (47)
Virologic failure ^c	24 (42)	17 (31)
Death or disease progression	1 (1)	1 (1)
Discontinued due to AE	8 (7)	5 (5)
Discontinued due to others ^d	13 (13)	16 (14)
Never initiated treatment	4 (4)	2 (2)

^a: Subjects achieved virologic response (two consecutive viral load <400 (<50) copies/mL) and maintained it to Week 48.

^b: Includes viral rebound and failing to achieved confirmed <400 (<50) copies/mL through Week 48.

^c: Includes lost to follow up, non-compliance, withdrawal and pregnancy.

It is clear from this table that virologic failures accounted for most of the failures (7 out 9% using <400 or 11 out of 13% using <50). The frequency of discontinuations due to AE is higher for the ddI-containing arm, while discontinuation due to other reasons is lower for this arm by a similar margin (3%). The other 2% in the treatment difference come from the subjects never initiated treatment.

How the subjects who did not initiated treatment should be handled is debatable. If these subjects are excluded, then the primary efficacy results can be summarized below.

Table 11: Week 48 HIV RNA Status using <400 copies/mL Treated Subjects

	Responders	
	ddI regimen N = 482	3TC regimen N = 248
Response rate	250/482 (52)	150/248 (60)
Difference	-8.6%	
95% CI	(-16.2%, -1.1%)	

Table 12: Week 48 HIV RNA Status using <50 copies/mL Treated Subjects

	Responders	
	ddl regimen N = 482	3TC regimen N = 248
Response rate	169/482 (35)	119/248 (48)
Difference	-12.9%	
95% CI	(-20.4%, -5.5%)	

We see the treatment differences are slightly less in this analysis compared to the analysis where subjects never initiated treatment are regarded as failures. However, the magnitudes of the lower bounds are very close, and the effects on inference should be minimal.

The CD4 changes are very similar between the two treatment arms, either measured by the median of Week 48 measurement, or by the time-averaged difference. For example, the median CD4 cell counts increases at Week 48 are 189 cells/mm³ for the ddl-containing regimen and 186 cells/mm³ for the control arm.

Virologic Response Over Time

To examine the relative treatment benefits over the course of the study, the proportion of responders are calculated for each study visit.

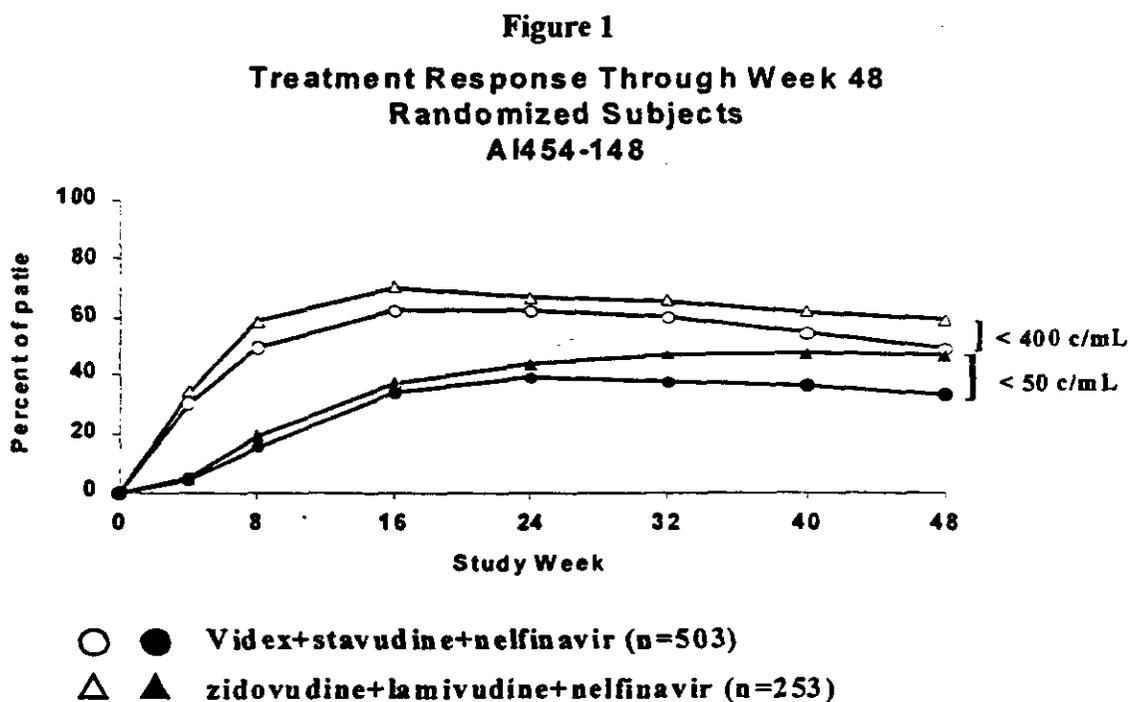
Table 13: Treatment Response Through Week 48 Using HIV RNA < 400 c/mL All Randomized Subjects

Time point	All Randomized Subjects	
	ddl regimen N responders/N at risk (%)	3TC regimen N responders/N at risk (%)
Week 4	154/503 (31)	87/253 (34)
Week 8	251/503 (50)	149/253 (59)
Week 16	314/503 (62)	178/253 (70)
Week 24	314/503 (62)	170/253 (67)
Week 32	303/503 (60)	167/253 (66)
Week 40	276/503 (55)	157/253 (62)
Week 48	250/503 (50)	150/253 (59)

Table 13: Treatment Response Through Week 48 Using HIV RNA \leq 50 c/mL
All Randomized Subjects

Time point	ddl regimen	3TC regimen
	N responders/N at risk (%)	N responders/N at risk (%)
Week 4	22/503 (4)	12/253 (5)
Week 8	78/503 (16)	49/253 (19)
Week 16	173/503 (34)	95/253 (38)
Week 24	198/503 (39)	111/253 (44)
Week 32	190/503 (38)	120/253 (47)
Week 40	185/503 (37)	121/253 (48)
Week 48	169/503 (34)	119/253 (47)

The data can be illustrated below:



It is clear from the plot that the treatment differences increased over time after Week 24, favoring the control arm.

Subgroup Analysis

The reviewer conducted the subgroup analyses by age, gender, race and baseline HIV RNA level (<30,000 vs. \geq 30,000 copies/mL) for the proportions of responders at Week 48. The homogeneity of treatment differences among the treatment groups across strata defined by these factors is tested using the Breslow-Day test. Gender, race and screening HIV RNA appear to have no interaction with the treatment (p-values > 0.2). For age, there appears to be an interaction (p-value=0.02).

At week 24 review, the HIV RNA by treatment interaction was considered possibly (p-value=0.031) and no evidence of interaction was found for age by treatment. Considering the number of subgroup analysis conducted here, and inconsistency of the results between Week 24 and Week 48, we do not have firm evidence to conclude interaction of treatment with any of these variables.

F. Overall Assessment

Study 148 demonstrated that ddI+d4T+NLF could be as much as 17% worse than ZDV+3TC+NLF in achieving responder status at Week 48 when standard assay is used, or as much as 20% worse if ultrasensitive assay is used. These differences are statistically significant and worsened over time. The differences come primarily from the differences in virologic rebounds. The CD4 changes over time are similar for the two groups. The double substitution in the design does not allow valid inference on the contribution of ddI.

 8/22/00
Greg Soon, Ph.D.
Mathematical Statistician

Appears This Way
On Original