CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-875

STATISTICAL REVIEW(S)

Addendum

NDA/Serial Number:	21-875 / N_000
Drug Name:	NUVIGIL™ Tablets(armodafinil)
Indication(s):	Excessive sleepiness associated with shift work sleep disorder (SWSD), obstructive sleep apnea/hypopnea syndrome (OSAHS), and narcolepsy
Applicant:	Cephalon
Biometrics Division:	Biometrics I, HFD-710
Statistical Reviewer:	Jialu Zhang, Ph.D.
Concurring Reviewers:	Kun Jin, Ph.D. Kooros Mahjoob, Ph. D.
Medical Division:	Division of Neuropharmacological Drug Products, HFD-120
Clinical Team:	Norman Hershkowitz, M.D.

In the four conservative analyses requested by the Agency, the reviewer verified that one of the four showed some significant results for studies 3020, 3021 and 3025 (Table 1). This conservative analysis (also referred to analysis 7.2 in the statistical review) is conducted as follows: in placebo group, replace each flawed session at baseline with the lowest sleep latency of the first four baseline sessions of each corresponding patient (take consideration of both flawed and unflawed sessions but count actual awakening time for sleep latency if it is a flawed session) and replace each flawed session at endpoint with highest sleep latency of unflawed sessions from the first four endpoint sessions of the corresponding patient. In treatment group(s), replace each flawed session at baseline with the highest sleep latency of unflawed sessions from the first four baseline session of the corresponding patient and replace each flawed session at endpoint with lowest sleep latency of the first four endpoint sessions (take consideration of both flawed and unflawed sessions but count actual awakening time for sleep latency if it is a flawed session). The derived dataset was provided by the sponsor and was not verified.

A similar conservative analysis which uses all six sessions instead of the first four sessions fails to show the similar significant results. This indicates that the drug may have more effect in the earlier sessions.

Table 1. Conservative Analysis Results in Mean Sleep Latency (Minutes)

Study	Armodafinil 250 mg N Mean (SD) p-value	Armodafinil 150 mg N Mean (SD) p-value	Armodafinil Combined N Mean (SD) p-value	Placebo N Mean (SD)	Treatment difference 95% CI
3020	60 2.1 (5.93) 0.004	58 -0.3 (5.16) 0.1206	118 0.9 (5.67) 0.0114	58 -1.6 (6.18)	(0.49, 3.86)
3021	121 1.6 (8.35) 0.025	120 1.3 (7.35) 0.0479	241 1.5 (7.85) 0.015	124 -0.7 (8.29)	(0.43, 3.90)
3025		116 2.1 (7.98) 0.0315		120 -0.1 (7.14)	(0.19, 4.00)

Jialu Zhang, Ph.D.
Mathematical Statistician

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

/s/

Jialu Zhang
4/12/2006 10:22:09 AM
BIOMETRICS

Kun Jin
4/13/2006 11:42:58 AM
BIOMETRICS

Kooros Mahjoob
4/14/2006 11:01:14 AM
BIOMETRICS

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21-875 / N_000
Drug Name:	NUVIGIL™ Tablets(armodafinil)
Indication(s):	Excessive sleepiness associated with shift work sleep disorder (SWSD), obstructive sleep apnea/hypopnea syndrome (OSAHS), and narcolepsy
Applicant:	Cephalon
Date(s):	Date of Document: March 31, 2005 PDUFA due date: January 31, 2006
Review Priority:	Standard
Biometrics Division:	Biometrics I, HFD-710
Statistical Reviewer:	Jialu Zhang, Ph.D.
Concurring Reviewers:	Kun Jin, Ph.D. Kooros Mahjoob, Ph. D.
Medical Division:	Division of Neuropharmacological Drug Products, HFD-120
Clinical Team:	Norman Hershkowitz, M.D.
Project Manager:	Jacqueline H Ware, Pharm.D.
Keywords:	Sleep latency, inconsistency, raw data, derived data, ad-hoc analysis

Table of Contents

List of Tables

List of Figures

1. EXECUTIVE SUMMARY

The submission has 4 efficacy studies targeting 3 indications, namely, narcolepsy, obstructive sleep apnea/hypopnea syndrome (OSAHS), and shift work sleep disorder (SWSD).

1.1 Conclusions and Recommendations

There are serious mishandling of raw data in these studies and therefore the original primary analyses from the sponsor based on the derived data **are likely to be invalid. Upon the agency's** requests, the sponsor conducted a few ad-hoc analyses. The ad-hoc analyses, except worst case analyses, show consistently significant treatment effects. However, it remains a question whether the post-hoc analyses accurately assess the treatment effects.

1.2 Brief Overview of Clinical Studies

The objective of this clinical program was to demonstrate efficacy and safety of armodafinil in the treatment of patients with excessive sleepiness associated with narcolepsy, obstructive sleep apnea/hypopnea syndrome (OSAHS), and shift work sleep disorder (SWSD).

This submission includes 4 completed Phase 3 double-blind, placebo-controlled, parallel-group efficacy/safety studies (2 studies in OSAHS, 1 study in SWSD, and 1 study in narcolepsy). The major features of these studies are summarized in Table 1.

Table 1: Phase 3 Double-Blind, Placebo-Controlled Studies of Armodafinil

Study No	Number of efficacy evaluable patients	Indication	primary endpoints	Dose/day
3020	176 / 3 arms	narcolepsy	mean sleep latency in 20-minute MWT, CGI-C	250mg/day, 150mg/day
3021	365 / 3 arms	OSAHS	mean sleep latency in 30-minute MWT, CGI-C	250mg/day, 150mg/day
3022	216 / 2 arms	SWSD	mean sleep latency in 20-minute MSLT, CGI-C	150mg/day
3025	236 / 2 arms	OSAHS	mean sleep latency in 30-minute MWT, CGI-C	150mg/day

OSAHS=obstructive sleep apnea/hypopnea syndrome; SWSD=shift work sleep disorder; MWT=Maintenance of Wakefulness Test; MSLT=Multiple Sleep Latency Test

1.3 Statistical Issues and Findings

The inconsistency between sleep technicians and the central readers artificially prolongs sleep latency as some sessions were interrupted prior to the termination of the sessions. Specifically, a patient can be disconnected from the monitoring device by a local technician before the session ends and yet the central reader may not find the protocol-defined sleep pattern in the pruned

tracing. When this happened, the central reader assigned full session length to that patient, indicating that the patient did not fall in sleep in that session, even though the patient may fall in sleep at any time between the time of disconnection and the termination of the session. The overall incidences of affected sessions range from 6.7% to 12.7% in studies 3020, 3021 and 3025 (Table 2). The number of patients affected in those three studies varies from 17% to 34.7% in either baseline visit or endpoint visit (Table 3). Study 3022 was much less affected. As a result of discrepancies, the original analyses in this NDA submission cannot reach a meaningful conclusion due to this serious problem on sleep latency derivation.

The sponsor was requested to provide the summary statistics of number of sessions affected and perform a serial of sensitivity analyses including worst case analysis to assess the potential impact.

2. INTRODUCTION

2.1 Overview

NUVIGIL™ (also referred to as CEP-10953, R-modafinil and armodafinil) is the levorotatory enantiomer of the racemic compound modafinil. Modafinil is a wakefulness-promoting agent **used as the active ingredient in Cephalon's** product PROVIGIL Tablets for the treatment of patients with excessive sleepiness associated with narcolepsy, obstructive sleep apnea/hypopnea syndrome (OSAHS), and shift work sleep disorder (SWSD).

The objective of this clinical program was to demonstrate efficacy and safety of armodafinil in the treatment of patients with excessive sleepiness associated with OSAHS, SWSD, or narcolepsy.

This submission includes 4 completed Phase 3 double-blind, placebo-controlled, parallel-group efficacy/safety studies (2 studies in OSAHS, 1 study in SWSD, and 1 study in narcolepsy).

2.2 Data Sources

The sponsor's electronic submission is stored under the directory of \\CDSESUB1\N21875\N_000\2005-03-31 **in the center's electronic document room. The** sponsor also submitted the SAS programs that produced all of the NUVIGIL efficacy results on June 13, 2005. This submission is stored under the directory of \\CDSESUB1\N21875\N_000\2005-06-13. The raw score data from the 30 randomly selected patient is stored under the directory of \\CDSESUB1\N21875\N_000\2005-08-12.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

Each study has two primary efficacy parameters. Change from baseline in the mean sleep latency in the Maintenance of Wakefulness Test (MWT) is one primary efficacy measure in Study 3020, 3021 and 3025. Study 3022 uses change from baseline in the mean sleep latency in the Multiple Sleep Latency Test (MSLT) as a primary efficacy measure. Clinical Global Impression of Change (CGI-C) was the co-primary efficacy endpoint in all four studies.

In MWT, Patients are instructed to try to remain awake. Sleep latency is defined as the time to onset of the first of 3 consecutive epochs of stage 1 sleep, or the time to onset of any epoch of stages 2, 3, and 4, or rapid eye movement (REM) sleep. The tracings obtained from MWT were divided into 30-second epochs. Epochs were assigned a sleep stage (wake, or sleep stages 1, 2, 3, or 4, or rapid eye movement [REM] sleep) individually using 50% (or 16-second) rule. If a patient does not fall sleep, their sleep latency will be assigned to full session length (20 or 30 minutes).

In MSLT, patients were instructed to lie quietly and attempt to sleep. Each MSLT nap continued until: (a) 3 consecutive 30-second epochs of stage 1 sleep are reached or (b) any single, 30 second epoch of stage 2, 3, 4, or REM sleep is reached. Each nap was terminated after 20 minutes if no sleep occurs. Sleep latency will be measured as the elapsed time from lights-out to the first epoch scored as sleep.

According to the protocol and clinical study report, if a patient falls asleep, he/she should be awakened immediately, and should stay in bed but be prevented from falling asleep again for the remainder of the testing time.

In the three studies using Maintanice of Wakefulness Test (MWT), the sponsor claimed that an error in the computer program that derives the primary efficacy variable from the raw scores of sleep stages for the MWT was discovered after the unblinding of the data. Specifically, the tracings from MWT were first sent to [redacted] and scored by a central reader. A data reduction program written at [redacted] to calculate the sleep latency from the raw scores was applied incorrectly, resulting in sleep latency values that were not consistent with the protocol definition of latency to unequivocal sleep. The sleep latency algorithm was corrected and the derived dataset was updated with the corrected sleep latency values. All the analyses in submitted NDA were based on derived datasets from corrected algorithm. b(4)

For verification purpose, the reviewer requested the raw score data from 30 randomly selected patients. There were some serious findings in the raw score data requested. In essence, sleep technicians who conducted MWT in study centers were instructed to wake up patients and disconnect equipment when the patients were considered to fall into sleep. The MWT tracings, however, were sent to [redacted] and scored by central readers. This creates a problem due to the inconsistency between clinical judgment from sleep technicians and the reading from central readers. A patient can be disconnected by a local technician before the session ends and yet the central reader may not find the protocol-defined sleep pattern in the pruned tracing. When this happened, the central reader assigned full session length to that patient, indicating that the b(4)

patient did not fall in sleep in that session. This artificially prolongs sleep latency as some sessions were interrupted within a few minutes. From the raw score data of 30 patients randomly selected, the reviewer found that 20 patients had at least one session affected and 7 patients had critical visits (baseline or the last visit) affected. Due to substantial changes to the observed data, the original analyses and results on the submitted data cannot be used for the regulatory decision.

Upon the Division's request, the sponsor conducted several ad-hoc analyses to understand the impact of this problem. The analyses requested include

Analysis 1. Use the latency to sleep as determined by local site: i.e. based upon time of awakening at local site.

Analysis 2. Use the last observation carried forward (LOCF) for endpoint visits with a flawed nap session. Thus, if there is a flawed score in any one of the four naps at the endpoint visit, use the next earliest visit where all four scores are unflawed (e.g., use week 8 if week 12 is incomplete because of a single flawed score or week four if week 12 and 8 are incomplete). If all experimental period naps are incomplete (weeks 4, 8, and 12), use the unflawed experimental median endpoint score for the endpoint visit for the combined placebo/drug groups to replace **that patient's endpoint value. If** one of the four baseline naps is flawed, use the median combined placebo/drug baseline value of unflawed data **to replace that patient's baseline value.**

Analysis 3. In a visit with a flawed nap session, replace each flawed nap score with the next unflawed nap score to obtain an average of 4 nap sessions (i.e. if one of the four naps is flawed, replace with the fifth and then sixth if there is a problem with the fifth). If two naps are flawed, replace with the next 2 unflawed naps. If there remains no unflawed nap to serve as a replacement, use the unflawed median of combined drug/placebo for that single nap score replacement (baseline median for baseline missing scores and endpoint experimental median for endpoint missing scores).

Analysis 4. Discard flawed nap sessions that occur in the first 4 naps and replace with the median of unflawed scores for the combined placebo/drug (first 4 naps) for that particular experimental stage. Baseline values should be replaced with baseline medians and experimental endpoint with experimental endpoint median.

Analysis 5. Discard data and do not analyze patients with any of the first 4 nap sessions flawed at baseline or final endpoint measure.

Analysis 6. Average only nap sessions that are unflawed in the first four naps.

Analysis 7. Worst case analysis

> **Analysis 7.1.** In placebo group, replace each flawed session at baseline with the lowest sleep latency of six baseline sessions of each corresponding patient (take consideration of both flawed and unflawed sessions but count actual awakening time for sleep latency if it is a flawed session) and replace each flawed session at endpoint with highest sleep latency of unflawed sessions from six endpoint sessions of the corresponding patient. In treatment

group(s), replace each flawed session at baseline with the highest sleep latency of unflawed sessions from six baseline session of the corresponding patient and replace each flawed session at endpoint with lowest sleep latency of six endpoint sessions (take consideration of both flawed and unflawed sessions but count actual awakening time for sleep latency if it is a flawed session). This would increase primary endpoint measures to placebo group and decrease the measures in treatment group(s).

Analysis 7.2. Repeat analysis 7.1), but instead of using all six sessions, use the first 4 sessions.

Analysis 8. Most conservative worst case analysis

Analysis 8.1. In placebo group, replace each flawed session at baseline with the lowest sleep latency of six baseline sessions of the corresponding patient (take consideration of both flawed and unflawed sessions but count actual awakening time for sleep latency if it is a flawed session) and replace each flawed session at endpoint with full length of study session (20 or 30 minutes). In treatment group(s), replace each flawed session at baseline with full session length (20 or 30 minutes) and replace each flawed session at endpoint with lowest sleep latency of six endpoint sessions (take consideration of both flawed and unflawed sessions but count actual awakening time for sleep latency if it is a flawed session).

Analysis 8.2. Repeat analysis 8.1), but instead of using all six sessions, use the first 4 sessions.

In addition, a number of summary statistics were requested in order to better understand the scope of the impact. The overall in**cidences of "flawed" sessions range from 6.7% to 12.7% in** studies 3020, 3021 and 3025 (Table 2). The number of patients affected in those three studies varies from 17% to 34.7% in either baseline visit or endpoint visit (Table 3). Study 3022 was much less impacted compared to the rest. This is because the sleep latency in study 3022 is defined as elapsed time from lights-out to the first 30-second epoch scored as sleep.

Shown in

Table 4, the average awakening time of "flawed" sessions in study 3020 was around 10 minutes, which is half of the full session length (20 minutes). The average actual awakening time in studies 3021 and 3025 were at two thirds of the full session length (30 minutes).

By looking at the summary statistics, the "flawed" sessions did not appear to bias towards a particular treatment arm or a visit.

Table 2: Frequency Count of Flawed Sessions

summary	Protocol	CEP-10953 250 MG	CEP-10953 150 MG	Placebo	Total Overall
Pooled counts in all visits	3020	7.9% (109/1388)	15.3% (203/1328)	14.7% (195/1330)	12.5% (507/4046)
	3021	7.3% (197/2706)	6.2% (168/2696)	6.6% (190/2880)	6.7% (555/8282)
	3025		10% (268/2676)	13.5% (367/2723)	11.8% (635/5399)
	3022		1.5% (30/2060)	0.6% (12/1930)	1.1% (42/3990)
pooled counts in critical visits	3020	9.2% (66/715)	14.1% (98/694)	14.9% (103/692)	12.7% (267/2101)
	3021	8.6% (124/1438)	6.8% (98/1437)	6.2% (92/1481)	7.2% (314/4356)
	3025		9.6% (133/1382)	14% (199/1422)	11.8% (332/2804)
	3022		0.8% (9/1120)	0.2% (2/1045)	0.5% (11/2165)

Table 3: Total Number of Patients with At Least One Flawed Session in the First 4 Nap Sessions

Protocol	Visit	CEP-10953 250 MG	CEP-10953 150 MG	Placebo	Total Overall
3020	Baseline	16.7% (10/60)	34.5% (20/58)	31% (18/58)	27.3% (48/176)
	Endpoint	28.3% (17/60)	34.5% (20/58)	41.4% (24/58)	34.7% (61/176)
3021	Baseline	18.2% (22/121)	22.5% (27/120)	18.5% (23/124)	19.7% (72/365)
	Endpoint	21.5% (26/121)	17.5% (21/120)	12.1% (15/124)	17% (62/365)
3025	Baseline		21.6% (25/116)	31.7% (38/120)	26.7% (63/236)
	Endpoint		22.4% (26/116)	30% (36/120)	26.3% (62/236)
3022	Baseline		0% (0/112)	0% (0/104)	0% (0/216)
	Endpoint		4.5% (5/112)	1% (1/104)	2.8% (6/216)

Table 4: Summary Statistics of the Flawed Session Length (minutes) in the First 4 Nap Sessions

Protocol	Visit	Statistic	CEP-10953 250 MG	CEP-10953 150 MG	Placebo	Total Overall
3020	Baseline	n	20	30	30	80
		Mean	11.0	9.4	10.0	10.0
		SD	4.97	4.86	4.58	4.76
		Median	9.3	8.5	8.8	9.0
	Endpoint	n	25	38	40	103
		Mean	10.8	11.7	10.1	10.9
		SD	5.08	4.40	4.95	4.79
		Median	10.5	12.0	9.0	10.5
3021	Baseline	n	42	42	41	125
		Mean	23.1	20.6	20.6	21.4
		SD	6.05	7.15	4.96	6.19
		Median	24.8	21.8	21.0	22.5
	Endpoint	n	38	26	22	86
		Mean	23.4	23.5	23.3	23.4
		SD	6.45	5.72	4.62	5.75
		Median	25.5	25.8	22.5	25.5
3025	Baseline	n		49	69	118
		Mean		17.4	19.3	18.5
		SD		7.75	7.50	7.63
		Median		15.5	21.0	19.3
	Endpoint	n		39	68	107
		Mean		20.1	17.4	18.4
		SD		7.20	6.74	7.00
		Median		20.5	17.5	19.0
3022	Baseline	n		0	0	0
		Mean		.	.	.
		SD		.	.	.
		Median		.	.	.
	Endpoint	n		6	1	7
		Mean		11.7	19.0	12.7
		SD		8.44	.	8.19
		Median		12.3	19.0	19.0

Individual subsections from 3.1.1 to 3.1.4 describe respectively in more detail about each of the four studies. The study descriptions in these **subsections are based on the sponsor's study reports.** The reviewer had identical or very similar results as reported in the following subsections.

3.1.1 STUDY 3020

3.1.1.1 Study Objective

The primary objective of this study was to determine whether treatment with CEP-10953 (armodafinil) is more effective than placebo treatment for patients with excessive sleepiness associated with narcolepsy.

3.1.1.2 Study Design

This study is a 12-week, multicenter, randomized, double-blind, placebo-controlled, parallel-group study to evaluate the efficacy and safety of Armodafinil at 150 and 250mg/day as treatment for adults with excessive sleepiness associated with narcolepsy. Patients were randomized (1:1:1) to receive 150 or 250mg/day of Armodafinil or placebo once daily for a 12-week double-blind treatment period. The 20-minute version MWT was administered 6 times (naps at 0900, 1100, 1300, 1500, 1700, and 1900) in each visit at weeks 4, 8, and 12.

3.1.1.3 Efficacy Measures

(1) Primary Efficacy Endpoint and Analysis

The study involves 2 primary efficacy variables. One is the change from baseline to endpoint in the mean sleep latency from the MWT (average of 4 tests at 0900, 1100, 1300 and 1500). The second is the proportion of patients with at least minimal improvement in the CGI-C rating at endpoint.

The primary comparison is between the combined armodafinil and the placebo group by ANCOVA. If the primary comparison is significant at α=0.05, then individual dose-group comparisons versus the placebo group will be made.

Same testing procedure is applied on the other primary efficacy variable CGI-C. Both primary endpoints were tested separately at α=0.05. No multiplicity adjustment was made.

(2) Secondary Efficacy Endpoints

The key secondary efficacy variable is change from baseline in the mean quality of episodic secondary memory from the tests of memory from the CDR system at endpoint. Other secondary efficacy variables include attention and memory assessed by CDR system, mean sleep latency from MWT at later time points, Epworth Sleepiness Scale (ESS) score, Brief Fatigue Inventory (BFI) score, CGI-C by visit and mean sleep latency by visit.

3.1.1.4 Patient Disposition, Demographic and Baseline Characteristics

Disposition of Patients

A total of 196 patients were randomized in the study at 47 centers in 6 countries. Of the 196 patients randomized, 194 (99%) patients received at least 1 dose of study drug and were evaluated for safety in the study. The population evaluable for efficacy comprised 176 (90%) patients who received at least 1 dose of study drug and had a baseline MWT assessment and at least 1 postbaseline MWT and CGI-C assessment. A total of 160 (82%) patients completed the study, 105 (80%) patients receiving armodafinil and 55 (86%) patients receiving placebo.

Table 5: Disposition of Patients in Study 3020

	Number (%) of patients				
Patient disposition	**Armodafinil 250 mg/day (N=67)**	**Armodafinil 150 mg/day (N=65)**	**Armodafinil combined (N=132)**	**Placebo (N=64)**	**Total (N=196)**
Screened	..	..	..	..	326
Randomized	67 (100)	65 (100)	132 (100)	64 (100)	196 (100)
Safety analysis set	67 (100)	64 (98)	131 (>99)	63 (98)	194 (99)
Full analysis set	60 (90)	58 (89)	118 (89)	58 (91)	176 (90)
Completed	56 (84)	49 (75)	105 (80)	55 (86)	160 (82)
Discontinued	11 (16)	16 (25)	27 (20)	9 (14)	36 (18)

SOURCE: **Sponsor's Clinical Study Report C10953/3020/NA/MN Table 5 on page 58.**

Patients' Demographic Characteristics

The treatment groups are well matched in regard to race and **gender. Patients' demographic** characteristics are summarized in Table 6. Of 194 patients in the safety analysis set, 73% patients in the study are white. It is noted that the 3 treatment groups (armodafinil 250 mg/day, armodafinil 150 mg/day, and placebo) differ with regard to age (p=0.0355).

Baseline Characteristics

Mean baseline sleep latencies in MSLT and CGI-S ratings at baseline are similar among patients in each treatment group in the evaluable population (Table 7).

Table 6: Demographic Information of Patients in Study 3020 (Safety Analysis Set)

Demographic variable Statistic	Armodafinil 250 mg (N=67)	Armodafinil 150 mg (N=64)	Armodafinil combined (N=131)	Placebo (N=63)	Total (N=194)	p-value
Age, years						
Mean	35.0	40.4	37.7	39.2	38.1	0.0355[a]
SD	12.52	12.52	12.76	11.98	12.50	
Sex, n (%)						
Male	25 (37)	28 (44)	53 (40)	32 (51)	85 (44)	0.3017[b]
Female	42 (63)	36 (56)	78 (60)	31 (49)	109 (56)	
Race, n (%)						
White	48 (72)	44 (69)	92 (70)	49 (78)	141 (73)	0.7737[c]
Black	9 (13)	13 (20)	22 (17)	10 (16)	32 (16)	
Other	7 (9)	5 (9)	12 (10)	1 (2)	13 (7)	
Missing	3 (4)	2 (3)	5 (4)	3 (5)	8 (4)	

Source: Sponsor's Clinical Study Report C10953/3020/NA/MN Table 8 on page 63.
a The p-value for the overall treatment comparison is from an analysis of variance (ANOVA) with treatment group and center as a factor.
b The p-value for the overall treatment comparison is from a Pearson's chi square test.
c The p-value for the overall treatment comparison is from a Fisher's exact test.

Table 7: Patients' Baseline Characters in Study 3020

Baseline variable Statistic	Armodafinil 250 mg (N=67)	Armodafinil 150 mg (N=64)	Armodafinil combined (N=131)	Placebo (N=63)	Total (N=194)	p-value
CGI-S ratings, n (%) (safety analysis set)						
Normal.not at all ill	0	0	0	0	0	0.7672
Borderline ill	0	0	0	0	0	
Slightly ill	0	0	0	0	0	
Moderately ill	25 (37)	19 (30)	44 (34)	18 (29)	62 (32)	
Markedly ill	29 (43)	32 (50)	61 (47)	34 (54)	95 (49)	
Severely ill	12 (18)	11 (17)	23 (18)	11 (17)	34 (18)	
Among the most extremely ill	1 (1)	2 (3)	3 (2)	0	3 (2)	
MSLT sleep latency, minutes (full analysis set)						
n	60	58	118	58		
Mean	2.6	2.5	2.5	2.6		
SD	1.61	1.72	1.66	1.55		
Min, max	0.1, 6.1	0.0, 6.6	0.0, 6.6	0.3, 6.0		

Source: Sponsor's Clinical Study Report C10953/3020/NA/MN Table 9 on page 67.

3.1.1.5 Sponsor's Primary Efficacy Results

The change from baseline in sleep latency from the MWT was tested using ANCOVA model with treatment and country as factors, and the corresponding baseline value as a covariate. The proportion of patients with at least minimal improvement in the CGI-C was tested using a Cochran-Mantel-Haenzel (CMH) chi-square test adjusted for country.

Table 8: Changes from Baseline to Endpoint in Mean Sleep Latency (Minutes) from the Maintenance of Wakefulness Test (MWT) in Study 3020

Method *	Armodafinil 250 mg N Mean (SD) p-value	Armodafinil 150 mg N Mean (SD) p-value	Armodafinil Combined N Mean (SD) p-value	Placebo N Mean (SD)	treatment difference 95% CI
Analysis 1	60 **2.3 (5.84)** 0.0002	58 **1.3 (5.11)** 0.0005	118 **1.8 (5.50)** <0.0001	58 **–2.3 (5.68)**	1.84, 5.04
Analysis 2	60 **2.0 (6.08)** 0.0007	58 **1.8 (5.30)** 0.0003	118 **1.9 (5.69)** <0.0001	58 **–2.3 (5.85)**	1.88, 5.36
Analysis 3	60 **2.7 (5.56)** 0.0001	58 **0.9 (5.32)** 0.0017	118 **1.8 (5.49)** <0.0001	58 **–2.2 (5.81)**	1.77, 5.04
Analysis 4	60 **2.3 (5.72)** 0.0001	58 **08 (4.92)** 0.0010	118 **1.6 (5.37)** <0.0001	58 **–2.4 (5.62)**	1.75, 4.87
Analysis 5	40 **2.6 (6.11)** 0.0085	26 **1.4 (4.64)** 0.0292	66 **2.1 (5.57)** 0.0072	25 **–2.5 (6.76)**	1.13, 6.53
Analysis 6	58 **2.8 (5.91)** 0.0003	54 **0.7 (5.12)** 0.0050	112 **1.8 (5.61)** 0.0002	55 **–2.2 (5.95)**	1.60, 5.06
Analysis 7.1	60 1.9 (5.94) 0.0279	58 -0.7 (5.56) 0.4064	118 0.6 (5.88) 0.0795	58 -1.2 (6.30)	-0.18, 3.24)
Analysis 7.2	60 2.1 (5.93) 0.0047	58 -0.3 (5.16) 0.1206	118 0.9 (5.67) 0.0114	58 -1.6 (6.18)	(0.49, 3.86)
Analysis 8.1	60 1.1 (6.52) 0.6716	58 -1.2 (5.56) 0.6176	118 0.0 (6.15) 0.9633	58 0.1 (6.91)	(-1.79, 1.73)
Analysis 8.2	60 1.2 (6.50) 0.5739	58 -1.0 (5.45) 0.7473	118 0.1 (6.08) 0.8911	58 -0.0 (6.89)	(-1.62, 1.88)
original NDA	60 **2.6 (6.24)** 0.0099	58 **1.3 (6.31)** 0.0068	118 **1.9 (6.28)** 0.0024	58 **–1.9 (6.87)**	1.02, 4.61

Source: Sponsor's 10-31-2005 and 12-16-2005 submissions.

* Please refer to the detailed analysis methods listed on page 8 in Section 3.1.

Most ad-hoc analyses and the original NDA analysis showed significant treatment difference for the primary comparison between the combined armodafinil and placebo treatment groups in favor of Armodafinil (Table 8). The worst case analyses, however, failed to differentiate among treatment groups.

The proportion of patients who had at least minimal improvement in the CGI-C rating from baseline to endpoint was statistically significant ($p<0.0001$) for the primary comparison between the combined armodafinil and placebo treatment groups (Table 9).

Table 9: Proportion of Patients with at Least Minimal Improvement in the Clinical Global Impression of Change (CGI-C) Rating at Endpoint in Study 3020

	Number (%) of patients			
CGI-C rating	Armodafinil 250 mg/day (N=60)	Armodafinil 150 mg/day (N=58)	Armodafinil combined (N=118)	Placebo (N=58)
At least minimal improvement	44 (73)	40 (69)	84 (71)	19 (33)
p-value	<0.0001	<0.0001	<0.0001	.

Source: Sponsor's Clinical Study Report C10953/3020/NA/MN Table 13 on page 76

3.1.1.6 Sponsor's Secondary Efficacy Results

The key secondary efficacy variable in the study was mean quality of episodic secondary memory from the CDR system testing. The result was significant with p-value=0.0032 comparing the combined armodafinil treatment groups with placebo group. Other secondary endpoints, including mean speed of memory, mean power of attention, ESS score, and average fatigue score were also significant.

3.1.2 STUDY 3021

3.1.2.1 Study Objectives

The primary objective of this study was to determine whether treatment with Armodafinil is more effective than placebo treatment for patients with residual excessive sleepiness associated with obstructive sleep apnea/hypopnea syndrome (OSAHS).

3.1.2.2 Study Design

This was a 12-week, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. Patients were randomized at ratio 1:1:1 to receive 150 or 250 mg/day of Armodafinil or a placebo once daily. Each MWT session was 30-minute long.

3.1.2.3 Efficacy Measures

(1) Primary Efficacy Endpoint

The primary efficacy variables for this study were the change from the baseline in mean sleep latency (average of 4 naps at 0900, 1100, 1300 and 1500) from 30-minute MWT and the proportion of patients with at least minimal improvement in the CGI-C ratings.

Both primary efficacy variables were tested separately at α=0.05 between the combined armodafinil group and the placebo group first. Provided that the test showed significance, tests comparing individual dose-group versus placebo group were performed. No multiplicity adjustment was made.

(2) Secondary Efficacy Endpoints

The key secondary efficacy variable was change from baseline in the mean quality of episodic secondary memory from the tests of memory from the CDR system at endpoint. Other secondary efficacy variables included attention and memory assessed by CDR system, mean sleep latency from MWT at later time points, ESS score, BFI score, CGI-C by visit and mean sleep latency by visit.

3.1.2.4 Patient Disposition, Demographic and Baseline Characteristics

Disposition of Patients

In this study, 638 patients were screened, of whom 395 (62%) patients met entrance criteria and were randomized in the study at 37 centers in 2 countries. The disposition of patients is summarized in Table 10.

Table 10: Disposition of patients in Study 3021

	Number (%) of patients				
Patient disposition	**Armodafinil 250 mg/day (N=131)**	**Armodafinil 150 mg/day (N=131)**	**Armodafinil combined (N=262)**	**Placebo (N=130)**	**Total (N=392)**
Screened	..	..	..	..	638
Randomized	131 (100)	133 (100)	264 (100)	131 (100)	395 (100)
Safety analysis set	131 (100)	131 (98)	262 (>99)	130 (>99)	392 (>99)
Full analysis set	121 (92)	120 (90)	241 (91)	124 (95)	365 (92)
Completed	110 (84)	114 (86)	224 (85)	120 (92)	344 (87)
Discontinued	21 (16)	19 (14)	40 (15)	11 (8)	51 (13)

SOURCE: Sponsor's Clinical Study Report C10953/3021/AP/MN Table 5 on page 58.

Patients' Demographic Characteristics

The treatment groups were well matched in regard to age, race and gender (summarized in Table 11). Of 392 patients in the safety analysis set, 85% patients in the study were white.

Table 11: Patients' Demographic Information in Study 3021

Demographic variable Statistic	Armodafinil 250 mg (N=131)	Armodafinil 150 mg (N=131)	Armodafinil combined (N=262)	Placebo (N=130)	Total (N=392)	p-value
Age, years						
Mean	49.1	49.3	49.2	50.1	49.5	0.6551[a]
SD	8.74	9.17	8.94	9.43	9.1	
Sex, n (%)						
Male	89 (68)	97 (74)	186 (71)	90 (69)	276 (70)	0.5216[b]
Female	42 (32)	34 (26)	76 (29)	40 (31)	116 (30)	
Race, n (%)						
White	111 (85)	109 (83)	220 (84)	113 (87)	333 (85)	0.7004[b]
Nonwhite	20 (15)	22 (17)	42 (16)	17 (13)	59 (15)	

Source: Sponsor's Clinical Study Report C10953/3021/AP/MN Table 8 on page 63.
a. The p-value for the overall treatment comparison is from an analysis of variance (ANOVA) with treatment group and center as a factor.
b. The p-value for the overall treatment comparison is from a Pearson's chi square test.

Baseline Characteristics

CGI-S ratings at baseline were similar among patients in each treatment group. Treatment **comparison at baseline is tested by Fisher's exact test.**

Table 12: Patients' Baseline Characteristics in Study 3021

Baseline variable Statistic	Armodafinil 250 mg (N=131)	Armodafinil 150 mg (N=131)	Armodafinil combined (N=262)	Placebo (N=130)	Total (N=392)	p-value
CGI-S ratings, n (%) (safety analysis set)						
Normal.not at all ill	0	0	0	0	0	0.6635
Borderline ill	0	0	0	0	0	
Slightly ill	0	0	0	0	0	
Moderately ill	69 (53)	78 (60)	147 (56)	63 (48)	210 (54)	
Markedly ill	42 (32)	34 (26)	76 (29)	43 (33)	119 (30)	
Severely ill	17 (13)	17 (13)	34 (13)	22 (17)	56 (14)	
Among the most extremely ill	3 (2)	2 (2)	5 (2)	2 (2)	7 (2)	

Source: Sponsor's Clinical Study Report C10953/3021/AP/MN Table 9 on page 66.

3.1.2.5 Sponsor's Primary Efficacy Results

Since there was evidence of the treatment-by-covariate interaction in the ANCOVA model, covariate was dropped from the ANCOVA model per study protocol. The change from baseline in sleep latency was analyzed by ANOVA with treatment and country as factors.

Table 13: Change From Baseline to Endpoint in Mean Sleep Latency (Minutes) from the Maintenance of Wakefulness Test (MWT) in Study 3021

Method *	Armodafinil 250 mg N Mean (SD) p-value	Armodafinil 150 mg N Mean (SD) p-value	Armodafinil Combined N Mean (SD) p-value	Placebo N Mean (SD)	treatment difference 95% CI
Analysis 1	121 **2.3 (8.02)** 0.0005	120 **2.1 (6.70)** 0.0007	241 **2.2 (7.38)** <0.0001	124 **–1.2(8.19)**	1.75, 5.08
Analysis 2	121 **1.8 (7.81)** 0.0015	120 **1.5 (6.49)** 0.0050	241 **1.6 (7.17)** 0.0006	124 **–1.3 (8.50)**	1.28, 4.61
Analysis 3	121 **2.5 (7.97)** 0.0001	120 **1.8 (6.90)** 0.0014	241 **2.1 (7.45)** <0.0001	124 **–1.3 (8.12)**	1.82, 5.16
Analysis 4	121 **2.4 (7.84)** 0.0001	120 **2.0 (6.39)** 0.0006	241 **2.2 (7.14)** <0.0001	124 **–1.4 (8.36)**	1.94, 5.23
Analysis 5	84 **2.6 (8.00)** 0.0011	80 **1.8 (6.05)** 0.0088	164 **2.2 (7.11)** 0.0006	92 **–1.2 (8.25)**	1.47, 5.34
Analysis 6	118 **2.6 (8.19)** <0.0001	120 **1.8 (7.41)** 0.0012	238 **2.2 (7.80)** <0.0001	122 **–1.6 (8.59)**	2.01, 5.55
Analysis 7.1	121 1.5 (8.35) 0.0612	120 1.0 (7.25) 0.1636	241 1.2 (7.81) 0.0590	124 -0.5 (8.44)	(-0.06, 3.43)
Analysis 7.2	121 1.6 (8.35) 0.0258	120 1.3 (7.35) 0.0479	241 1.5 (7.85) 0.0150	124 -0.7 (8.29)	(0.43, 3.90)
Analysis 8.1	121 1.1 (8.58) 0.1422	120 0.8 (7.33) 0.2551	241 0.9 (7.97) 0.1317	124 -0.4 (8.42)	(-0.41, 3.12)
Analysis 8.2	121 1.1 (8.58) 0.1178	120 0.8 (7.37) 0.2174	241 1.0 (7.99) 0.1056	124 -0.5 (8.38)	(-031, 3.22)
Original NDA	121 2.2 (8.07) 0.0001	120 1.7 (6.49) 0.0008	241 1.9 (7.32) <0.0001	124 **–1.7 (8.59)**	**1.93, 5.31**

Source: Sponsor's 10-31-2005 and 12-16-2005 submissions.

* Please refer to the detailed analysis methods listed on page 8 in Section 3.1.

Another primary efficacy measure CGI-C was tested using a Cochran-Mantel-Haenzel (CMH) chi-square test adjusted for country.

Most ad-hoc analyses and the original NDA analysis showed significant treatment difference for the primary comparison between the combined armodafinil and placebo treatment groups in favor of Armodafinil (Table 13). The worst case analyses, however, failed to differentiate among treatment groups.

The proportion of patients who had at least minimal improvement in the CGI-C rating from baseline to endpoint was statistically significant ($p<0.0001$) for the primary comparison between the combined armodafinil and placebo treatment groups (Table 14).

Table 14: Proportion of Patients with at Least Minimal Improvement in CGI-C Rating at Endpoint in Study 3021

	Number (%) of patients			
CGI-C rating	Armodafinil 250 mg/day (N=121)	Armodafinil 150 mg/day (N=120)	Armodafinil combined (N=241)	Placebo (N=124)
At least minimal improvement	89 (74)	85 (71)	174 (72)	46 (37)
p-value	<0.0001	<0.0001	<0.0001	.

Source: Sponsor's Clinical Study Report C10953/3021/AP/MN Table 13 on page 74

3.1.2.6 Sponsor's Secondary Efficacy Results

The ESS and BFI scores had significant results when comparing the combined Armodafinil to the placebo. The key secondary variable (mean quality of episodic secondary memory) and other secondary variables were not significant.

3.1.3 STUDY 3022

3.1.3.1 Study Objectives

The primary objective of this study was to determine whether treatment with armodafinil is more effective than placebo treatment for patients with excessive sleepiness associated with chronic shift work sleep disorder (SWSD).

3.1.3.2 Study Design

This was a 12-week, multicenter, randomized, double-blind, placebo-controlled, parallel-group study of the efficacy and safety of armodafinil as treatment for adults with excessive sleepiness associated with chronic SWSD. Patients were randomized in a 1:1 ratio to receive 150 mg of armodafinil or placebo for 12 weeks. Outcome assessments included the MSLT, the BFI, KSS, CDR system testing, and CGI-C.

3.1.3.3 Efficacy Measures

(1) Primary Efficacy Endpoint

The primary efficacy variables for this study were the mean change from the baseline assessment in 20-minute MSLT mean sleep latency (average of 4 naps at 0200, 0400, 0600, and 0800) and the proportion of patients with at least minimal improvement in the CGI-C ratings (for sleepiness during night shifts including the commute to and from work) as assessed at week 12 or the last postbaseline observation.

(2) Secondary Efficacy Endpoints

The key secondary variable in the study was the change from baseline in the mean quality of episodic secondary memory from the tests of memory from the CDR system. Other secondary variables included attention and memory assessed by CDR system, Karolinska Sleepiness Scale (KSS) scores, BFI scores, CGI-C ratings by visit and mean sleep latency in MSLT by visit.

3.1.3.4 Patient Disposition, Demographic and Baseline Characteristics

Disposition of Patients

A total of 747 patients were screened for the study and 254 of them were randomized in the study at 42 centers in USA and Canada. Disposition of patients is summarized in Table 15.

Table 15: Disposition of Patients in Study 3022

	Number (%) of patients		
Patient disposition	**Armodafinil 150 mg**	**Placebo**	**Total**
Screened	..	..	747
Randomized	127 (100%)	127 (100%)	254 (100%)
Safety analysis set	123 (97%)	122 (96%)	245 (96%)
Full analysis set	112 (88%)	104 (82%)	216 (85%)
Completed study	97 (76%)	89 (70%)	186 (73%)
Discontinued	30 (24%)	38 (30%)	68 (27%)

Source: Sponsor's Clinical Study Report C10953/3022/CM/MN table 5

Patients' Demographic Characteristics

The armodafinil and placebo groups were generally matched in regard to age, gender and race. The armodafinil group had less white patients (60%) than did the placebo group (70%). But the different is not significant when tested at alpha=0.05.

Table 16: Patients' Demographic Characteristics in Study 3022

Demographic variable Statistic, n (%)	Armodafinil 150 mg (N=123)	Placebo (N=122)	Total (N=245)
Age, years			
Mean	38.9	40.3	39.6
SD	10.75	10.76	10.76
Sex, n (%)			
Male	66 (54)	64 (52)	130 (53)
Female	57 (46)	58 (48)	115 (47)
Race group, n (%)			
White	74 (60)	86 (70)	160 (65)
Nonwhite	49 (40)	36 (30)	85 (35)

Source: Sponsor's Clinical Study Report C10953/3022/CM/MN Table 8

Baseline Characteristics

CGI-S ratings at baseline were similar between the two groups. 56% and 57% patients were moderately ill in armodafinil and placebo group, respectively.

3.1.3.5 **Sponsor's Primary Efficacy Results**

Since there was evidence of treatment*covariate interaction, the covariate was dropped from the ANCOVA model and ANOVA model was used instead. The testing result (see Table 17) from ANOVA was significant (p-value<0.0001). Since the **impact of "flawed" session is minimal in** study 3022, the ad-hoc analyses were consistent with the original NDA analysis and showed significant treatment effect. The sponsor did not perform worst case analysis for this study.

Table 17: Changes from Baseline to Endpoint in Mean Sleep Latency from MSLT in Study 3022

Method *#	Armodafinil 150 mg N Mean (SD) p-value	Placebo N Mean (SD)	Treatment difference 95% CI
Analysis 1	112 3.0 (4.38) <0.0001	104 0.4 (2.86)	1.57, 3.56
Analysis 2	112 **3.0 (4.35)** <0.0001	104 **–0.3 (2.89)**	1.69, 3.68
Analysis 3	112 2.9 (4.30) <0.0001	104 0.3 (2.87)	1.54, 3.51
Analysis 4	112 2.8 (4.18) <0.0001	104 0.3 (2.87)	1.52, 3.45
Analysis 5	107 2.7 (4.12) <0.0001	103 0.4 (2.87)	1.37, 3.30
Analysis 6	112 2.9 (4.35) <0.0001	104 0.3 (2.87)	1.58, 3.57
Original NDA	112 3.1 (4.46) <0.0001	104 0.4 (2.87)	1.67, 3.69

Source: Sponsor's 10-31-2005 submission.
* Please refer to the detailed analysis methods listed on page 8 in Section 3.1.
 # Due to the small percentage of "flawed" sessions in this study, no worst case analyses were performed

CGI-C rating was assessed using a CMH chi-square test adjusted for country and the result was statistically significant (p=0.0010).

Table 18: Proportion of Patients with at Least Minimal Improvement in the CGI-C in Study 3022

Variable, n (%)	Armodafinil 150 mg (N=112)	Placebo (N=104)	p-value
At least minimal improvement[a]	89 (79)	61 (59)	0.0010
No improvement	23 (21)	43 (41)	

Source: Sponsor's Clinical Study Report C10953/3022/CM/MN

3.1.3.6 **Sponsor's Secondary Efficacy Results**

The key secondary efficacy variable of the study is the mean quality of episodic secondary memory from the CDR system. The difference between armodafinil and placebo groups was statistically significant (p-value<0.0001).

3.1.4 STUDY 3025

3.1.4.1 Study Objectives

The primary objective of this study was to determine whether treatment with armodafinil is more effective than placebo treatment for patients with residual excessive sleepiness associated with obstructive sleep apnea/hypopnea syndrome (OSAHS).

3.1.4.2 Study Design

This was a 12-week, multicenter, randomized, double-blind, placebo-controlled, parallel-group study to evaluate the efficacy and safety of armodafinil at 150 mg/day as treatment for adults with residual excessive sleepiness associated with OSAHS. Patients were randomized to receive armodafinil or a placebo at a ratio of 1:1 within each country. 30-minute version MWT were performed same as study 3021.

3.1.4.3 Efficacy Measures

(1) Primary Efficacy Endpoint

The primary efficacy variables for this study were the mean change from the baseline assessment in MWT mean sleep latency (average of 4 naps at 0900, 1100, 1300, and 1500) and the proportion of patients with at least minimal improvement in the CGI-C ratings (as related to general condition) at week 12 (or the last postbaseline observation).

(2) Secondary Efficacy Endpoints

The key secondary variable was the mean from baseline in the mean power of attention from the tests of attention from the CDR system (average of 4 tests at 0930, 1130, 1330 and 1530). Other secondary efficacy variables included attention and memory assessed by CDR system, mean sleep latency from MWT at later time points, ESS score, BFI score, CGI-C by visit and mean sleep latency by visit.

3.1.4.4 Patient Disposition, Demographic and Baseline Characteristics

Patient Disposition

A total of 466 patients were screened and 263 patients were randomized in the study at 36 centers in 5 countries. Table 19 summarizes the disposition of patients.

Table 19: Disposition of Patients in Study 3025

Patient disposition	Number (%) of patients		
	Armodafinil 150 mg (N=131)	Placebo(N=132)	Total (N=263)
Screened	..	..	466
Randomized	131 (100)	132 (100)	263 (100)
Randomized, not treated	2 (2)	2 (2)	4 (2)
Safety analysis set	129 (98)	130 (98)	259 (98)
Full analysis set	116 (89)	120 (91)	236 (90)
Completed	111 (85)	118 (89)	229 (87)
Discontinued	20 (15)	14 (11)	34 (13)

Source: Sponsor's Clinical Study Report C10953/3025/AP/MN table 5 on page 57

Patients' Demographics Characteristics

The treatment groups were matched in regard to age, gender and race. 84% patients who took at least one dose of treatment were white. 73% patients were male and mean age of patients was 50.7. The information was summarized in Table 20.

Table 20: Patients' Demographics Characteristics in Study 3025

Demographic variable Statistic	Armodafinil 150 mg (N=129)	Placebo (N=130)	Total (N=259)
Age, years			
Mean	50.7	50.6	50.7
SD	9.17	8.85	8.99
Sex, n (%)			
Male	97 (75)	93 (72)	190 (73)
Female	32 (25)	37 (28)	69 (27)
Race group, n (%)			
White	107 (83)	111 (85)	218 (84)
Nonwhite	22 (17)	18 (14)	40 (15)
Missing	0	1 (<1)	1 (<1)

Patients' Baseline Characteristics

Baseline characteristics compared among patients in the 2 treatment groups included categorical variables of CGI-S rating and ESS total score. CGI-S ratings at baseline were similar among two treatment groups. Specifically, 72 (56%) patients were moderately ill and 57 (44%) patients were markedly, severely, or extremely ill in armodafinil group. 75 (58%) patients were moderately ill and 55 (42%) patients were markedly, severaly or extremely ill in placebo group.

ESS total score was also matched at the baseline between two groups. Armodafinil group patients had mean of 15.6 in ESS score while placebo group had mean of 16.0.

3.1.4.5 Sponsor's Primary Efficacy Results

The baseline*treatment interaction was significant in the primary ANCOVA model. Per the statistical analysis plan, the alternative ANOVA model without baseline covariate was applied and the test result was significant (p=0.0003). The original NDA analysis also included a Wilcoxon rank-sum test and the result was consistent.

Table 21: Changes from Baseline to Endpoint in Mean Sleep Latency from MWT in Study 3025

Method *	Armodafinil 150 mg N Mean (SD) p-value	Placebo Mean (SD)	treatment difference 95% CI
Analysis 1	116 **2.8 (7.67)** <0.0001	120 **–1.5 (6.39)**	2.50, 6.06
Analysis 2	116 **1.7 (7.87)** 0.0037	120 **–1.2 (7.11)**	0.93, 4.75
Analysis 3	116 **2.4 (7.84)** 0.0002	120 **–1.3 (7.06)**	1.74, 5.56
Analysis 4	116 **2.6 (7.61)** 0.0002	120 **–0.9 (6.81)**	1.65, 5.31
Analysis 5	78 **2.7 (7.71)** 0.0002	73 **–1.6 (6.20)**	2.08, 6.57
Analysis 6	111 **2.6 (8.16)** 0.0007	112 **–1.0 (7.29)**	1.49, 5.51
Analysis 7.1	116 1.7 (8.33) 0.1418	120 0.1 (7.18)	(-0.50, 3.47)
Analysis 7.2	116 2.1 (7.98) 0.0315	120 -0.1 (7.14)	(0.19, 4.00)
Analysis 8.1	116 1.0 (8.34) 0.9285	120 1.0 (7.71)	(-2.15, 1.97)
Analysis 8.2	116 1.0 (8.18) 0.9959	120 1.0 (7.72)	(-2.04, 2.03)
original NDA	116 2.3 (7.80) 0.0003	120 **–1.3 (7.08)**	1.67, 5.46

Source: Sponsor's 10-31-2005 and 12-16-2005 submissons.

* Please refer to the detailed analysis methods listed on page 8 in Section 3.1.

Most ad-hoc analyses and the original NDA analysis showed significant treatment difference for the primary comparison between the combined armodafinil and placebo treatment groups in favor of Armodafinil (Table 21). The worst case analyses, however, failed to differentiate among treatment groups.

The proportion of patients who had at least minimal improvement in CGI-C from baseline to endpoint was statistically significant (p=0.0069).

3.1.4.6 Sponsor's Secondary Efficacy Results

The key secondary efficacy variable is the power of attention from the CDR system. The result is not significant (p=0.8181).

3.2 Evaluation of Safety

The evaluation of safety is not performed in this report. Please refer the clinical review of this application for safety evaluation.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Age, Gender and Ethnic group

Subgroup analyses were only performed with "flawed" sessions replaced by the actual awakening time at local sites. Table 22, Table 23 and

Table 24 showed mean change from baseline in each treatment group stratified by gender, race and age, respectively. There appeared to have no numerically inconsistent treatment effect across the subgroups.

Table 22: Subgroup Analysis by Gender

Study	SEX	TGRP	n	Mean (minutes)	std
3020	Male	250mg	24	2.40	6.85
	Male	150mg	26	1.29	5.42
	Male	placebo	29	-2.05	5.33
	Female	250mg	35	2.72	4.59
	Female	150mg	32	1.23	4.94
	Female	placebo	29	-2.53	6.10
3021	Male	250mg	83	2.25	8.22
	Male	150mg	91	1.92	7.08
	Male	placebo	86	-1.18	7.43
	Female	250mg	38	2.35	7.89
	Female	150mg	29	2.83	5.40
	Female	placebo	37	-0.37	8.90
3022	Male	150mg	60	2.99	4.53
	Male	150mg	58	0.28	3.44
	Female	placebo	52	3.31	4.45
	Female	placebo	47	-1.32	3.26
3025	Male	1	85	2.93	8.45
	Male	2	85	-0.61	5.99
	Female	1	31	2.46	5.25
	Female	2	35	-3.63	7.12

Table 23: Subgroup Analysis by Race

Study	RACE	TGRP	n	Mean (minutes)	std
3020	Other	250mg	7	4.32	7.99
	Other	150mg	5	-0.48	3.64
	Other	placebo	1	4.00	
	White	250mg	43	2.55	5.33
	White	150mg	40	1.50	4.98
	White	placebo	45	-2.85	5.74
	Black	250mg	7	2.11	4.99
	Black	150mg	11	1.31	6.65
	Black	placebo	9	0.24	5.16
3021	Other	250mg	6	11.02	7.37
	Other	150mg	6	3.27	4.70
	Other	placebo	7	-0.20	4.81
	White	250mg	105	2.17	7.77
	White	150mg	100	2.09	7.02
	White	placebo	108	-0.95	8.08
	Black	250mg	10	-1.83	8.54
	Black	150mg	14	2.01	5.16
	Black	placebo	8	-1.48	7.76
3022	Other	150mg	11	5.32	4.72
	Other	placebo	9	-0.78	2.86
	White	150mg	68	3.15	4.64
	White	placebo	73	-0.41	3.30
	Black	150mg	33	2.38	3.92
	Black	placebo	23	-0.39	4.16
3025	Other	150mg	14	1.47	7.70
	Other	placebo	7	-4.25	3.65
	White	150mg	94	2.88	7.97
	White	placebo	102	-1.22	6.50
	Black	150mg	8	4.22	3.72
	Black	placebo	10	-2.39	7.76

Table 24: Subgroup analysis by Age

Study	Age	Treatment	N	Mean (minutes)	std
3020	<=30	250mg	28	1.55	5.92
	<=30	150mg	18	1.51	5.44
	<=30	placebo	16	-2.63	5.33
	31-45	250mg	21	3.11	4.65
	31-45	150mg	15	1.74	3.79
	31-45	placebo	21	-1.89	6.09
	>45	250mg	11	4.16	6.26
	>45	150mg	25	0.79	5.68
	>45	placebo	21	-2.44	5.77
3021	<=45	250mg	43	3.32	8.69
	<=45	150mg	38	3.31	7.92
	<=45	placebo	37	-0.99	8.99
	46-55	250mg	48	1.28	7.99
	46-55	150mg	44	1.57	6.15
	46-55	placebo	41	-0.91	6.58
	>55	250mg	30	2.35	7.36
	>55	150mg	38	1.63	5.97
	>55	placebo	46	-1.09	8.11
3022	<=35	150mg	45	3.05	4.82
	<=35	placebo	33	-0.64	3.35
	36-45	150mg	30	3.60	4.60
	36-45	placebo	32	0.50	4.02
	>45	150mg	37	2.87	4.01
	>45	placebo	40	-1.01	2.90
3025	<=45	150mg	34	0.75	7.80
	<=45	placebo	31	0.13	7.91
	46-55	150mg	47	4.31	8.94
	46-55	placebo	50	-1.33	6.11
	>55	150mg	35	2.72	5.10
	>55	placebo	39	-2.94	5.44

4.2 Other Subgroup Populations

No other subgroup analyses were performed.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Raw Data Validation

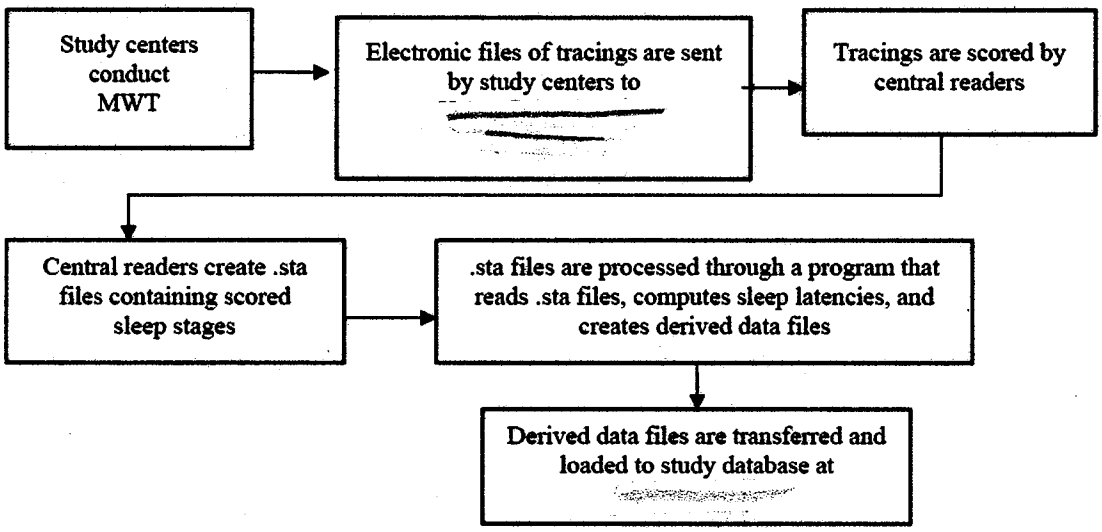
The submission has 4 efficacy studies targeting 3 indications, namely, narcolepsy, obstructive sleep apnea/hypopnea syndrome (OSAHS), and shift work sleep disorder (SWSD). There were some serious problems with the raw score data of those studies.

The discrepancy between local sleep technicians and the central readers artificially prolonged the primary endpoint measure. Specifically, in the situation where a patient was judged to be in sleep by the local sleep technicians, the technicians would disconnect patients from the equipment before the session terminated. The pruned tracing was then sent to the central reading lab. If the **central reader didn't find any protocol-defined sleep** pattern, the time to sleep of the patient in that session was set to the full length of the session (20 minutes or 30 minutes).

In addition to the data issue discovered by the reviewer, the sponsor also claimed in the NDA that a programming error was discovered in the computer program used to derive the primary efficacy variable after unblinding of the data. The data processing is shown in Figure 1. Specifically, the tracings from MWT were first sent to [redacted] and scored by a central reader. The raw scores were stored as .sta files and a data reduction program written at [redacted] then computed the sleep latencies from .sta files. The original algorithm incorrectly calculated the sleep latency from the MWT as the time to 6 consecutive epochs (3 minutes) of sleep stage 1, 2, 3, or 4, or REM sleep, whereas the definition as specified in the protocols is the time to onset of the first of 3 consecutive epochs of stage 1 sleep or to 1 epoch of stage 2, 3, or 4, or REM sleep. The sponsor discovered and corrected the programming error after unblinding study data. Studies 3020, 3021 and 3025 were affected. b(4)

The sponsor mentioned that they had no prior knowledge regarding the discrepancy between local sleep technicians and the central readers due to that the raw score was processed in the [redacted] and only the derived datasets were delivered to the sponsor for further **analyses. Interestingly, a variable named "NA" in the original efficacy datasets identifies naps in** which the patient did not reach 3 consecutive minutes of sleep of any stage but was terminated early. Even though the variable was created by the central reader according to the incorrect definition of sleep and does not represent the **"flawed" sessions if followed the corrected** algorithm, it still indicates that the problem of early session termination without predefined sleep pattern was noticed by the central reader and recorded in the derived datasets. This raw data issue was brought up by the reviewer after the reviewer looked into the raw score data from 30 randomly selected patients. According to the information collected by the DSI field inspector from [redacted], the programmer who wrote the derivation program to process the raw data was instructed by the sponsor to set sleep latency to 20 or 30 minutes if the pre-defined sleep patterns were not found. b(4)

Figure 1: Data processing flowchart



b(4)

Source: Sponsor's documentations in statistical method

The original NDA data are contaminated with 5-15**% of "flawed" sessions. The original analyses** are thus invalid. Several ad-hoc analyses were **conducted by the sponsor per the agency's** request, the results except the worst case analyses were consistent and showed treatment significance. However, it remains a question whether those ad-hoc analyses accurately reflect the treatment effect.

Data Distribution and Model Assumption

The sleep latency data are indeed truncated data. The sleep session only lasted 20 or 30 minutes but a patient may fall into sleep at a later time point. If the proportion of the truncated data is non-ignorable, the sponsor may not be able to simply rely on ANCOVA or ANOVA. The reviewer made graphs of the distribution of time to sleep and the change from baseline for studies 3020 and 3021 as examples (Figure 2 -Figure 5). It appears that the proportion of truncated observation is high. When computing the change from baseline, the truncated data create a high proportion of zeroes (Table 25). The distribution of the primary efficacy data cannot be considered as normally distributed. This also explains why the medians of the armodafinil group and the placebo group are close in the primary analyses.

Multiplicity in Primary Analysis

All four studies listed two primary efficacy variables and two studies included 2 armodafinil dose groups (150mg and 250mg). The multiplicity issue was raised during the IND phrase by the Division and remained unadjusted in the NDA submission. The sponsor claimed that a closed testing procedure was used to protect type I error. Specifically, individual armodafinil dose group comparisons to the placebo group were made only if the combined armodafinil group

comparison to the placebo group was statistically significant. However, it is not clear how the sponsor would take into consideration of the multiple primary efficacy variables. Specifically, does the sponsor claim efficacy of the drug if both primary efficacy variables win? If the sponsor intends to make a claim if either of the two primary efficacy variables wins, then the combined armodafinil group comparison should be conducted at 0.025 instead of 0.05.

Key secondary endpoint

Studies 3021 and 3025 were designed to show the efficacy in treating OSAHS. Each study specified a key secondary endpoint. The key secondary variables are the mean quality of episodic secondary memory and mean power of attention in study 3021 and study 3025, respectively. Neither is significant.

5.2 Conclusions and Recommendations

There are serious sleep latency derivation problem in these studies and the original primary analyses from the sponsor are likely to be invalid. Upon the agency's requests, the sponsor conducted a number of ad-hoc analyses. The ad hoc analyses except worst case analyses show consistently significant treatment effect. However, since the original data were not collected in the right way and the ad-hoc analyses can only based on the "flawed" data to assess part of the impact, it remains questionable that the analyses accurately assess the treatment effect.
The reviewer would feel more confident to support the efficacy claim if at least some of the worst case analyses are statistically significant.

Appendix

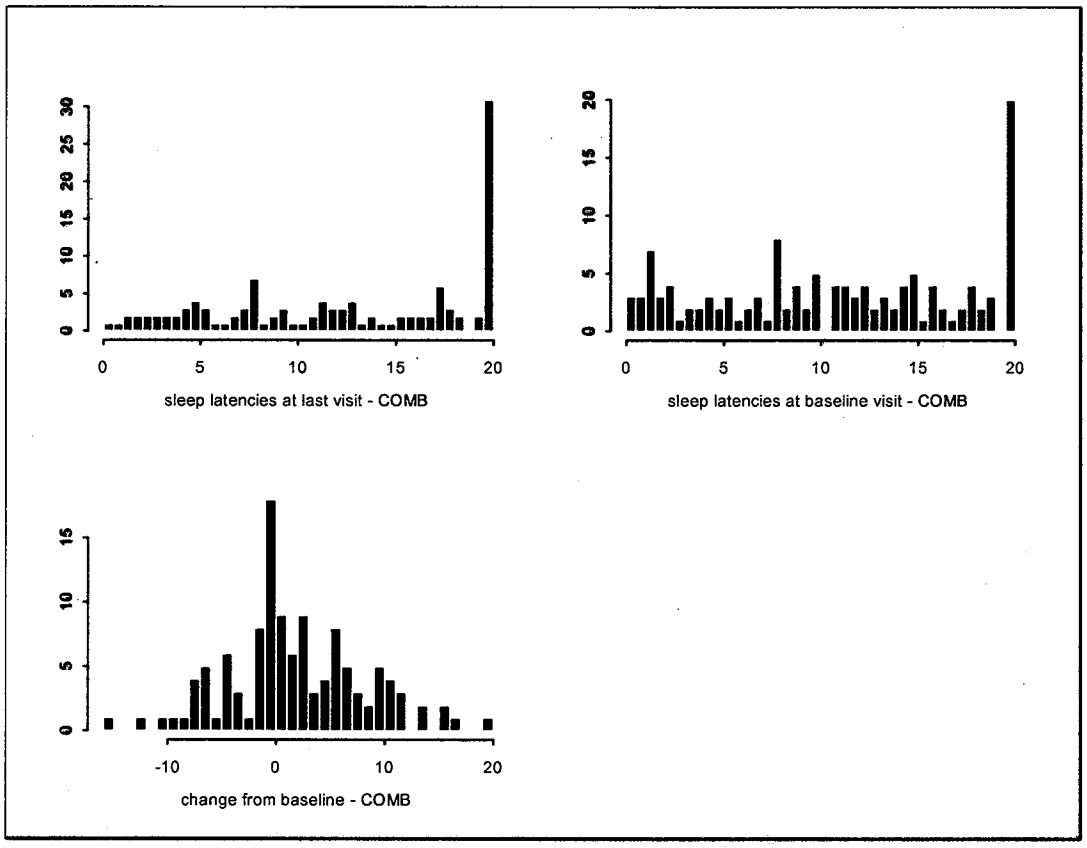


Figure 2: Distributions of sleep latencies and change from baseline in Study 3020 (150mg and 250 mg combined)

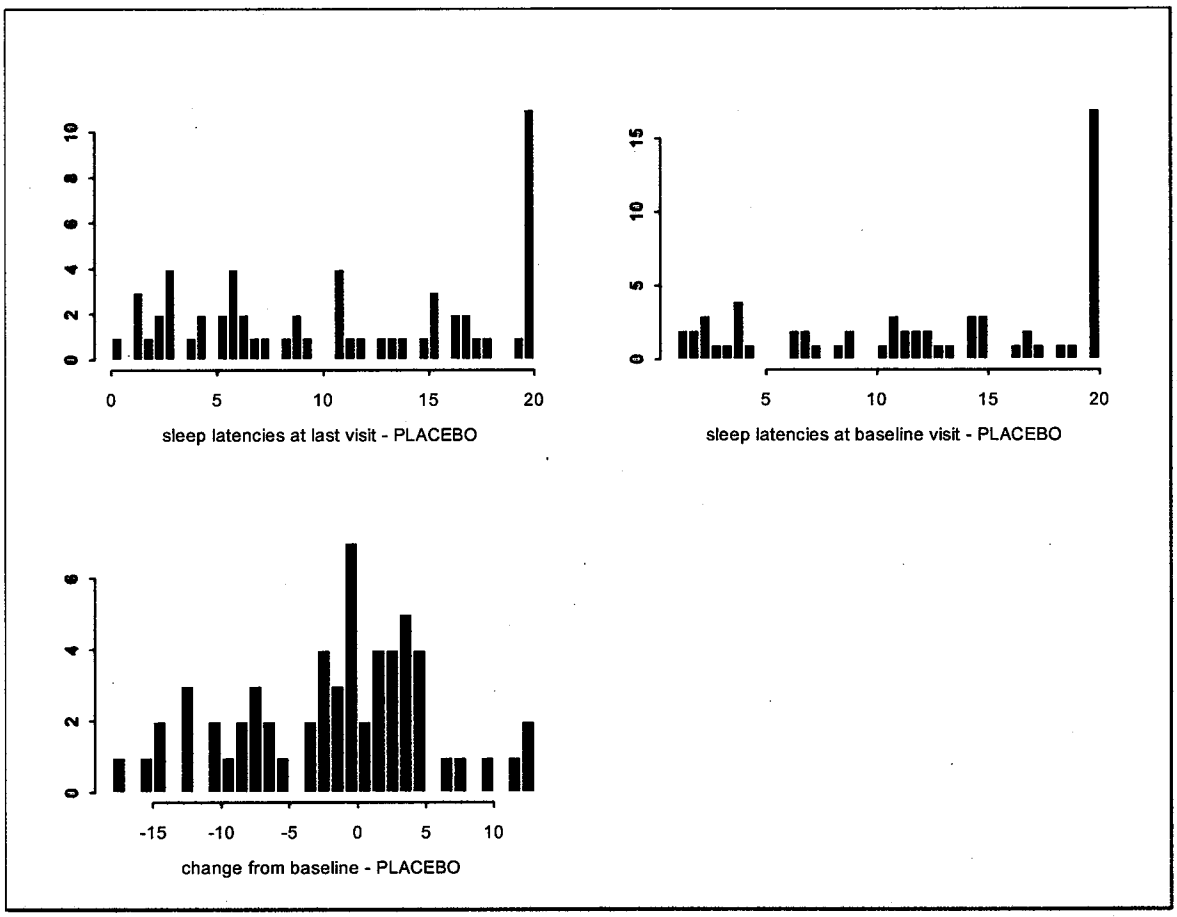


Figure 3: Distributions of sleep latencies and change from baseline in Study 3020 (placebo group)

Appears This Way
On Original

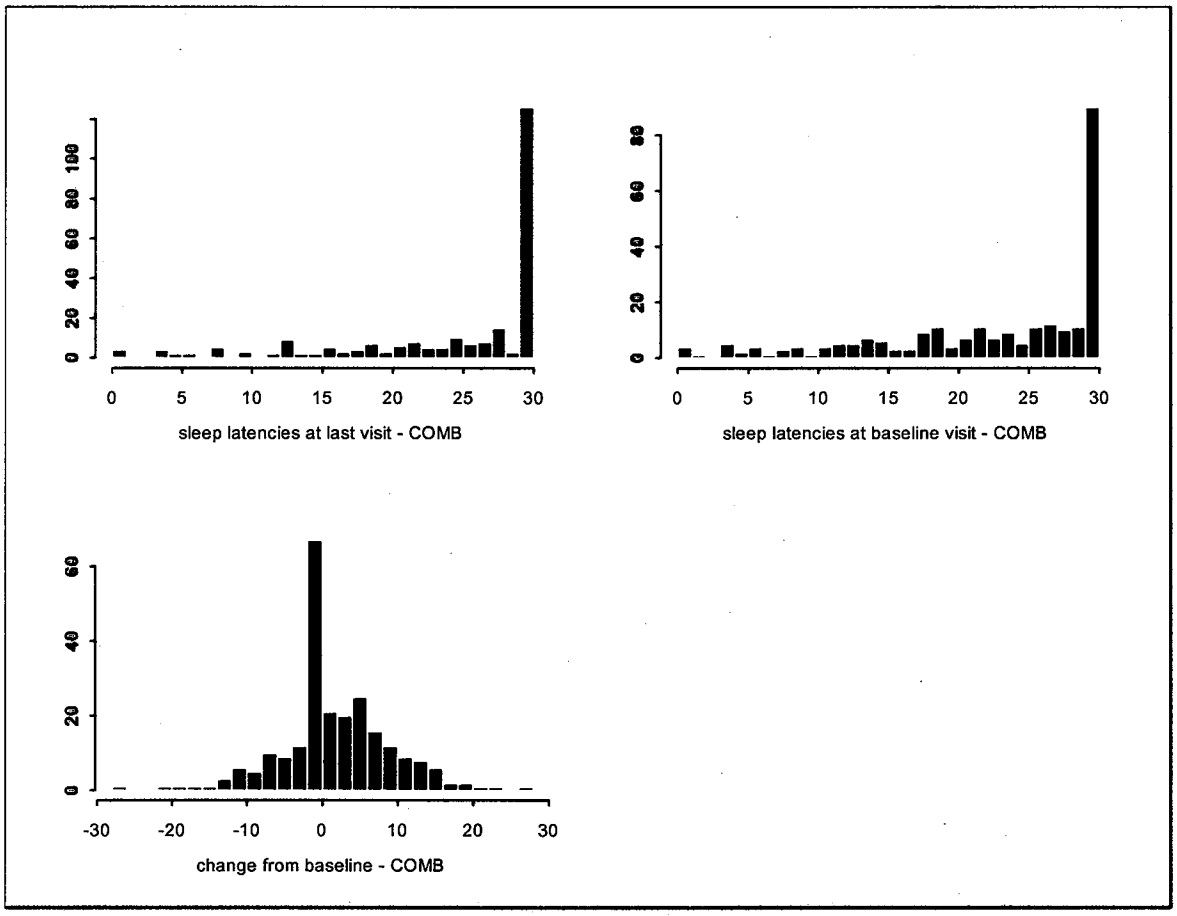


Figure 4: Distributions of sleep latencies and change from baseline in Study 3021 (150mg and 250 mg combined)

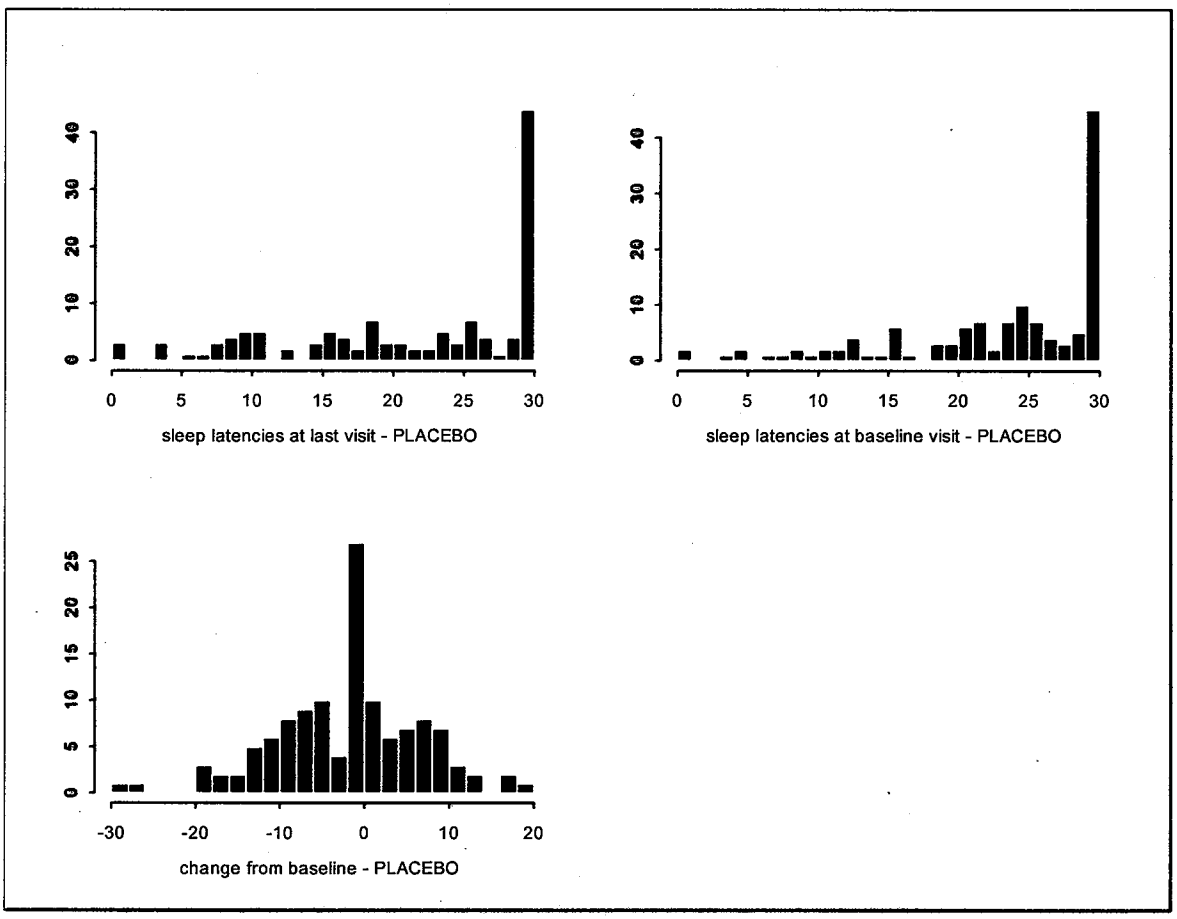


Figure 5: Distributions of sleep latencies and change from baseline in Study 3021 (placebo group)

Study	treatment	Freq	# of zeroes	Total Pt
3020	Placebo	10.17%	6	59
	150mg	17.24%	10	58
	250mg	3.33%	2	60
3021	placebo	17.74%	22	124
	150mg	20%	24	120
	250mg	23.14%	28	121
3025	placebo	28.33%	34	120
	150mg	35.34%	41	116

Table 25: Frequency table of zeroes in change from baseline in mean sleep latencies

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

/s/

Jialu Zhang
3/10/2006 02:54:43 PM
BIOMETRICS

Kun Jin
3/10/2006 02:57:42 PM
BIOMETRICS

Kooros Mahjoob
3/10/2006 07:53:04 PM
BIOMETRICS

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 STUDIES

IND/NDA Number:	NDA 21-875
Drug Name:	Armodafinil (Nuvigil™) Tablets
Indication(s):	26 Week Carcinogenicity in Transgenic Mice
Applicant:	Cephalon, Inc. 41 Moores Road, P.O. Box 4011 Frazer, PA 19355
Documents Reviewed:	Submission submitted electronically, Dated January 27, 2006 Data submitted electronically
Review Priority:	Standard
Biometrics Division:	Division of Biometrics VI
Statistical Reviewer:	Mohammad Atiar Rahman, Ph.D.
Concurring Reviewer:	Karl Lin, Ph.D.
Medical Division:	Division of Neurology Products
Reviewing Pharmacologist:	Ed Fisher, Ph.D.
Project Manager:	Jackie Ware
Keywords:	Carcinogenicity, Dose-Response

1. Background

In this submission the sponsor included a report of an animal carcinogenicity study in Tg.AC mice. This study was intended to assess the carcinogenic potential of armodafinil when applied topically for 26 weeks. The results of this review have been discussed with the reviewing pharmacologist Dr. Fisher.

2. Design

Two separate experiments, one in males and one in females were conducted. In each of these two experiments there were three treated groups along with a positive control group and a vehicle control group. The treated groups were 125, 250, and 500 mg/kg/day. The positive control group was 1.25 μg TPA applied 3 times per week and the vehicle control received methanol. The purpose of the positive control was to assess the sensitivity of the study. One hundred Tg.AC mice of each gender were randomly allocated to the control and treated groups of equal size of 20 animals.

The animals were observed regularly for mortality and appearance of skin papillomas. Tissue masses were diagnosed and scored as latent papillomas when they attained a 2 mm diameter and protruded from the surface of the skin. Upon remaining present for three consecutive weeks, the skin papillomas subsequently scored as actual papillomas. The weekly papilliomas were compiled for statistical analysis.

2.1 Sponsor's analyses

Figures 1A and 1B in the appendix show the Kaplan-Meier estimates of the cumulative proportion of mice with papillomas at site of application (SOA) in males and females, respectively. The data did not show any apparent dose response in proportion of animals with papillomas in males or females, respectively. Figures 2A and 2B the appendix show the survival adjusted average maximum papilloma burden in males and females, respectively. The data did not show any apparent dose response in papilloma burden in males or females. The sponsor**'s analysis** further showed a probability of less than 0.0065 for an animal getting a skin papilloma at SOA at the selected doses during the 26 weeks of treatment.

The sponsor analyzed the weekly papilloma count data of individual animal at SOA by gender, using the model proposed by Dunson et al. [David B. Dunson, Joseph K. Haseman, Angelique P. J. M. van Birgelen, Stanley Stasiewiez, and Raymond W. Tennant. Statistical Analysis of Skin Tumor Data from Tg.AC Mouse Bioassays. *Toxicological Sciences* **55**, 2000, 293-302]. The following model was fitted:

$$E(Yij) = \exp\{\beta_1 + (b_1+\gamma_1)t_j d_i\} \quad \text{if } M_{ij} = 0$$
$$= \exp\{\beta_2 + \gamma_2 d_i\} \quad \text{if } M_{ij} > 0$$

where d_i =is the dose level for i^{th} mouse, T is the duration of the study, $t_j = j/T$ is the fraction of time i^{th} mouse stayed in the study, M_{ij} is the maximum number of papilloma burden observed for i^{th} mouse prior to week j, and $Yij = (M_{ij} - M_{i\,j-1})$ is the increase in the maximum papilloma burden observed for i^{th} mouse between week j-1 and j. Yij is assumed to follow a Poisson sampling distribution with mean specified via above model. The interpretations of the parameters are as follows:

b_1 = Mouse-specific susceptibility variable
β_1 and β_2 = Intercepts related to the rate of appearance of spontaneous papillomas
γ_1 and γ_2= Slopes associated with the dose-response relationships in papilloma latency time and in papilloma

multiplicity after the appearance of the first papilloma.

In Tg.AC transgenic model, the carcinogenic effect of a drug is decomposed into two components (1) the dose-response relationship in papillomas latency time, and (2) the dose-response relationship in papillomas multiplicity after the appearance of the first papilloma. The hypotheses of interest are H_{01}: $\gamma_1 = \gamma_2 = 0$, H_{02}: $\gamma_1 = 0$ and H_{03}: $\gamma_2 = 0$, corresponding to incidence, latency, and multiplicity of papillomas. If H_{01} is not rejected, it can be concluded that there is no evidence of a dose-response in papilloma incidences. However, if H_{01} is rejected then it will be of interest to know whether the trend is due to a shortening of the latency time and/or an increment in papilloma multiplicity. If H_{02} is rejected then it can be concluded that there is a significant decrease in papilloma latency with increasing dose. If H_{03} is rejected then it can be concluded that there is a significant increase in papilloma multiplicity with increasing dose.

For the analysis of the data the Dunson et al. also provided a computer program which was available in their website (presently it has been withdrawn). However, sponsor wrote their own program using Proc NLMIXED in SAS. Tables 1A and 1B in the appendix show sponsor**'s analyses results for** male and female mice respectively. Results did not show any statistically significant dose responses in papilloma latency time and in papilloma multiplicity after the appearance of the first papilloma.

2.2 Reviewer's analysis

Since the sponsor wrote their own analysis program instead of using the one originally provided by Dunson et al., to verify their results this reviewer independently performed analysis on papilloma data at SOA using the analysis program developed by Dunson of NTP. Tables 2A and 2B in the appendix show **this reviewer's analysis results. This reviewer's** analysis results differed slightly from those of the sponsor, however the overall conclusion agreed i.e. the data did not show any statistically significant dose responses in papilloma latency time and in papilloma multiplicity after the appearance of the first the papilloma. The insignificant differences **between the sponsor's and this reviewer's an**alysis results might be due to the use of different sets of initial values in estimation of parameters in the non-linear mixed effect model.

3. Summary

In this submission the sponsor included a report of an animal carcinogenicity study in Tg.AC mice. This study was intended to assess the carcinogenic potential of armodafinil when applied topically for 26 weeks. Two separate experiments, one in male and one in female mouse were conducted. In each of these two experiments there were three treated groups along with one positive control group. The treated groups were 125, 250, and 500 mg/kg/day. The positive control group was 1.25 μg TPA applied 3 times per week and the vehicle control received methanol. One hundred mice of each gender were randomly allocated to the control and treated groups of equal size of 20 animals. Tissue masses were diagnosed and scored as latent papillomas when they attained a 2 mm diameter and protruded from the surface of the skin. Upon remaining present for three consecutive weeks, the skin papillomas subsequently scored as actual papillomas. The weekly papilliomas were compiled for statistical analysis. The analysis of the data did not show any statistically significant dose responses in papilloma latency time and in papilloma multiplicity after the appearance of the first papilloma.

Mohammad Atiar Rahman, Ph.D.
Mathematical Statistician, Biometrics VI

Concur: Karl K. Lin, Ph.D.
Team Leader, Biometrics VI

cc:
Archival NDA 21-875
Dr. Freed
Dr. Fisher
Dr. Bryan
Dr. Hershkowitz
Dr. Ware
Dr. Machado
Dr. Huque
Dr. Lin
Dr. Rahman
Dr. Jin
Dr. O'Neill
Ms. Patrician

Appendix

Table 1A

MLEs of the Parameters for Male Mice Using PROC NLMIXED

Parameter	Estimate	Standard Error	DF	t-Value	P-Value	Alpha	Lower CL	Upper CL
beta1	-6.8438	1.1787	79	-5.81	<.0001	0.05	-9.1900	-4.4976
beta2	-10.0021	105.06	79	-0.10	0.9244	0.05	-219.11	199.10
gamma1	-7.3770	15.6458	79	-0.47	0.6386	0.05	-38.5192	23.7651
gamma2	-0.00094	167.19	79	-0.00	1.0000	0.05	-332.78	332.78
sd	29.1658	79.8247	79	0.37	0.7158	0.05	-129.72	188.05

Table 1B

MLEs of the Parameters for Female Mice Using PROC NLMIXED

Parameter	Estimate	Standard Error	DF	t-Value	P-Value	Alpha	Lower CL	Upper CL
beta1	-9.3703	2.6856	79	-3.49	0.0008	0.05	-14.7159	-4.0247
beta2	-12.5027	294.91	79	-0.04	0.9663	0.05	-599.51	574.50
gamma1	-4.6566	3.8852	79	-1.20	0.2343	0.05	-12.3900	3.0768
gamma2	-2.0750	67.2181	79	-0.03	0.9755	0.05	-135.87	131.72
sd	68.3067	51.8606	79	1.32	0.1916	0.05	-34.9191	171.53

Table 2A

The Maximum Likelihood Estimate of the Parameters and P-values
Male Mouse
(Reviewer's Table)

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta1	-6.8555	1.1735	79	-5.84	<.0001	0.05	-9.1912	-4.5198	-0.00002
beta2	-12.4127	350.65	79	-0.04	0.9719	0.05	-710.36	685.53	8.379E-6
g1	-6.9487	14.2076	79	-0.49	0.6261	0.05	-35.2282	21.3308	-0.00002
g2	-3.9377	2304.16	79	-0.00	0.9986	0.05	-4590.26	4582.39	2.183E-7
sd	26.9805	69.9475	79	0.39	0.7007	0.05	-112.25	166.21	-3.64E-6

Table 2B

The Maximum Likelihood Estimate of the Parameters and P-values
Female Mouse
(Reviewer's Table)

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta1	-9.4785	2.3921	79	-3.96	0.0002	0.05	-14.2400	-4.7171	0.340972
beta2	-9.0761	205.88	79	-0.04	0.9649	0.05	-418.87	400.71	0.000076
g1	-5.2388	5.8431	79	-0.90	0.3727	0.05	-16.8693	6.3916	0.279538
g2	-4.4619	111.94	79	-0.04	0.9683	0.05	-227.27	218.34	0.000062
sd	85.6234	85.8901	79	1.00	0.3219	0.05	-85.3365	256.58	0.018959

Figure 1A
Proportion of Animals with Papillomas
In Male Mice

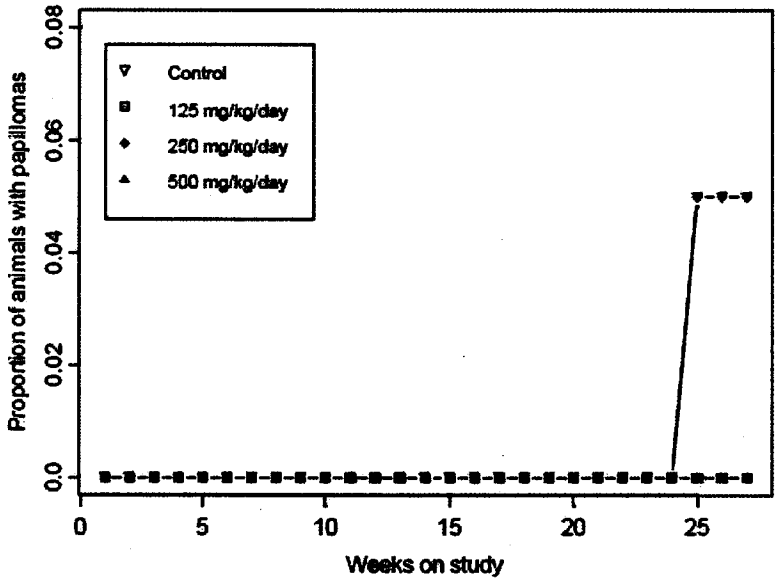


Figure 1B
Proportions of Animals with Papillomas
in Female Mice

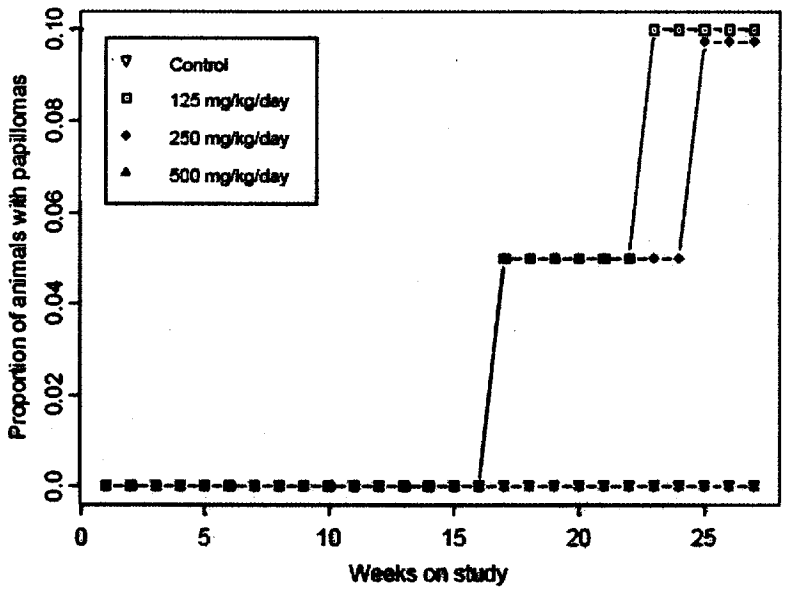


Figure 2A
Average Maximum Papilloma Burden
in Male mice

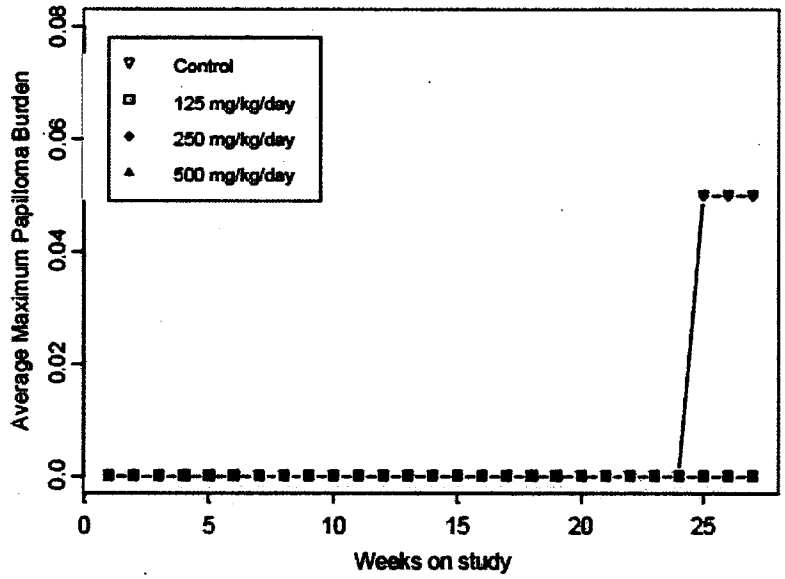
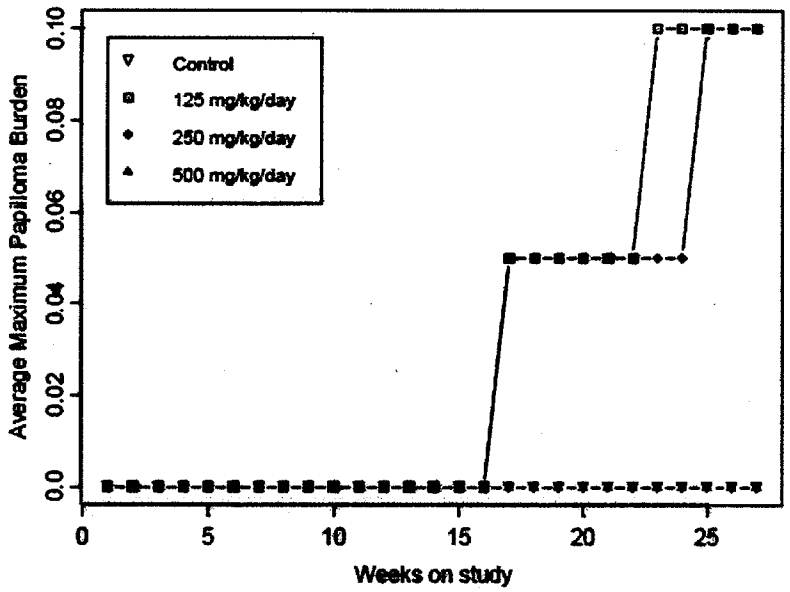


Figure 2B
Average Maximum Papilloma Burden
in Female mice



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

/s/

Atiar Rahman
2/28/2006 03:37:35 PM
BIOMETRICS

Karl Lin
2/28/2006 04:20:08 PM
BIOMETRICS
Concur with review