

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:
21-368

STATISTICAL REVIEW(S)

Memorandum of Statistical Review

NDA 21368

Name of Drug: Cialis (tadalafil)

Indication: Treatment for erectile dysfunction

Applicant: Lilly Icos LLC

Documents: \\Cdsub1\N21368\N_000\2003-05-27

Reviewer: Mike Welch, Ph.D., HFD-715

Medical: Mark Hirsch, M.D, HFD-580

Background

The sponsor has submitted a complete response to the Approvable Letter dated April 29, 2002. The sponsor's responses to the issues raised in the letter are based on guidance provided by the Division at teleconferences and meetings after the approvable action. Of the issues addressed by the sponsor in the resubmission, three require statistical comment and are discussed in this memorandum: Applicability of data to the U.S. population; Claims pertaining to _____ and Justification of starting dose in routine clinical practice.

Applicability of data to the U.S. population

The original NDA included six, phase 3, placebo controlled studies, conducted outside the U.S., to evaluate the efficacy and safety of different doses (5 mg, 10 mg, 20 mg) of tadalafil. All studies used the same three co-primary endpoints: the EF domain score and SEP questions 2 and 3. For the latter two, efficacy was based on the proportions of successful attempts. Statistical significance was required and was achieved on all three endpoints. Refer to the original statistical review for details.

To show that the effects of tadalafil are similar in the U.S. population, the sponsor completed two additional studies, LVCR and LVEF in a total of 402 subjects (305 assigned to tadalafil). Both studies compared 20 mg tadalafil to placebo and followed the design of the non-U.S. studies, including a 4-week run-in period to assess baseline ED followed by a 12 week treatment period. The inclusion and exclusion criteria and the primary endpoints and analysis methods were identical to those in the original studies.

The efficacy results for the U.S. studies are consistent with the results of the original studies. See Table 5.5 in the sponsor's response document *Applicability of Clinical Trial Data to the U.S. Population*. The sponsor also performed subgroup analyses of the pooled, U.S. data base, comparing Hispanic and Non-Hispanic subgroups and comparing African-descent and non-African-descent subgroups. Although the resulting samples sizes are too small for high statistical confidence, the efficacy endpoints show levels of response that are larger in the treated group as compared to placebo. No significant safety issues are noted for the U.S. studies; the reader is referred to the clinical review.

Claims pertaining to _____ of effectiveness

The sponsor's principle studies intending to support labeling language relating to _____ of effectiveness are studies LVCK, LVDG, and a new study, LVFD. Study LVCK, submitted with the original NDA, attempted to examine _____ for tadalafil 10 mg and 20 mg, compared to placebo, within 30 minutes after dosing. Study LVDG, also in the original NDA, examined the effectiveness of tadalafil (20 mg dose vs. placebo) by assessing the number of successful attempts occurring 24 or 36 hours post-dose to infer that _____ was at least 36 hours.

The statistical reviewer of the original submission concluded that study LVCK did not provide any valid, statistical characterization of _____ which was its primary objective, _____ and no such labeling would be warranted. For study LVDG, the reviewer noted that the study failed to measure _____ in a meaningful way and that the study only demonstrated a statistically significant difference between drug and placebo at 24 or at 36 hours with regard to percentage of successful attempts.

For study LVCK, labeling could include a descriptive summary of the percentages of patients, by treatment group, who had at least one successful attempt (out of four) within 30 minutes of dosing. For study LVDG, labeling discussion could include presentation of the primary efficacy measure, the percentage of successful attempts at 24 and 36 hours after dosing; any such table should also show the numbers of patients. A more meaningful presentation might be the percentages of patients who had at least one successful event. As noted in the original statistical review, labeling should not include any reference to _____

Study LVFD was a multicenter, randomized, double-blind, placebo controlled trial to evaluate the efficacy and safety of 10 mg and 20 mg tadalafil when sexual activity occurs at pre-specified time-points after dosing. The study randomized a total of 438 subjects to six parallel groups (10 mg, 20mg and placebo, each by assigned time: 24 or 36 hours). Subjects were asked to take four doses of study medication separated by at least seven days with each dose followed by a sexual attempt at the assigned time. The primary outcome was defined as the subject's percentage of "yes" responses to question 3 of the SEP diary taken over the four attempts. This reviewer has reviewed the study report, protocol and statistical methods and finds them satisfactory. The primary efficacy analyses show the mean percent "yes" responses are 42%, 56%, and 67% for the placebo, 10 mg, and 20 mg groups respectively for the 24-hour time point; and 33%, 56%, and 62% for the placebo, 10 mg, and 20 mg groups respectively for the 36-hour time point. (Refer to Table LVFD.11.6 in the clinical study report.) The results for both doses of tadalafil as evaluated by ANCOVA are statistically different from placebo for both time points. This reviewer has no objection to including these primary results in the proposed label.

Justification of starting dose in routine clinical practice

In an attempt to show an efficacy advantage of the 20 mg dose over the 10 mg dose, the sponsor presented in the original submission a pooled analysis of all studies with either dose. The statistical reviewer noted that when pooling (more properly) included only the three studies that tested both doses (LVCO, LVDJ, and LVBK) no evidence of statistical superiority is shown.

In the new analysis, the sponsor pools studies LVCO and LVDJ (study LVBK included diabetes mellitus patients) and argues that superiority of the 20 mg dose for the EF domain score is indicated ($p = .049$). The SEP endpoints, however, show no indication of any difference. Additional analysis are used to infer that more clinically and statistically important dose differences in EF domain score exist for patients with either moderate or severe baseline scores. The sponsor argues that as the safety and tolerability profiles of both regimes are similar, the gain in efficacy indicates a risk-benefit ratio in favor of the 20 mg dose. The sponsor's efficacy analyses, however, are ; the tolerability issue is a clinical review concern.

**APPEARS THIS WAY
ON ORIGINAL**

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

/s/

Mike Welch
11/6/03 11:41:02 AM
BIOMETRICS

Statistical Review and Evaluation Carcinogenicity

NDA No:	21-368
Applicant:	Eli Lilly and Company
Trade Name:	Cialis
Generic Name:	Tadalafil
Code name:	IC351 (LY450190)
Pharmacologist:	Yangmee Shin, Ph.D. (HFD-580)
Statistical Reviewer:	Moh-Jee Ng (HFD-715)

Summary

- In the 2-year mouse study, there was no statistically significant positive trend in survival and statistically significant difference in survival distributions among treatment groups in both males and females. There were statistically significant differences in tumor incidence in hemangiosarcoma in liver ($p=0.0381$) in females when comparing the control groups with the 400 mg/kg/day (high dose) group; in alveolar/bronchiolar adenoma and alveolar/bronch carcinoma combined in lung in females when comparing vehicle control 2 with the 60 mg/kg/day group ($p=0.0053$) and when comparing vehicle control 2 with the 400 mg/kg/day group ($p=0.0099$).
- In the 2-year rat study, there was no statistically significant positive trend in survival and statistically significant difference in survival distributions among treatment groups in both males and females. There were statistically significant positive-dose relationships in tumor incidence in fibroma in subcutaneous tissue ($p=0.0008$) in males, in hepatocellular adenoma and hepatocellular carcinoma combined in liver ($p=0.036$) in males. There was statistically significant difference in tumor incidence in adenocarcinoma in mammary gland ($p=0.0012$) in females when comparing vehicle control 1 with the 60 mg/kg/day (medium dose) treatment group.

1. Introduction

This reviewer evaluated the oncogenic potential of IC351 (LY450190), given to mice and rats by daily oral administration for 2 years. This report includes the results of the analyses of the survival and tumor data.

2. Studies Designs

The study designs of mice and rats are summarized in the following Table.

Table 1
Summary of Study Design

Species	Mice	Rat
Strain	Albino Mouse	Wister Rats
Route of Administration	Oral	Oral
Frequency of Drug Administration	Daily	Daily
Dose Unit	mg/kg/day	mg/kg/day
Dose Level (Control 1, Control 2, Low, Medium, High)	0, 0, 10, 60, 400	0, 0, 10, 60, 400
Number of Animals/sex/per treatment group	50 males/group 50 females/group	50 males/group 50 females/group
Length of Study	24 months	24 months

In each of these experiments there were two control groups and three treated groups known as low, medium, and high. The dose levels for the treatment groups were 0, 0, 10, 60 and 400 mg/kg/day, respectively, in the mouse and rat studies. There were 50 animals of each sex in each treatment group. All surviving males and females were necropsied following a minimum of 104 weeks of dosing. The terminal sacrifice started at and after weeks 103.

3. Sponsor's Tumor Analyses and Findings

The sponsor used the method of Tarone (1975) to perform dose-related trend for mortality data and used the SAS procedure, Proc Multtest (SAS release 6.12), to perform tumor analysis. The two control groups were pooled for the trend tests if there was no statistically difference ($p > 0.05$) in survival rates, otherwise, a separate analysis was conducted using each control group as dose level 0 for the trend tests. All tests used the two-sided trend test at the 0.05 significance level and were performed for each sex separately.

The sponsor listed the following findings in its reports.

In survival analysis:

- No significant differences in the mortality among the treated groups were detected when compared to the controls for both mice and rats.

In tumor analysis:

- No significant positive linear trends in incidence rate in tumor data for both mice and rats were detected.
- There was an increase in incidence of hepatocellular adenomas in male mice but no statistically significant when comparing the controls to the high dose group. The sponsor indicated that hepatocellular tumors were not associated with the development of foci of cellular lateration.
- Hemangiosarcoma was recorded as a primary tumor in multiple tissues in both male and female mice. However, there was no increase in the combined incidences of hemangiosarcoma from all tissues.

The sponsor concluded that the treatment with IC351 (LY450190, ———) at dose levels of 0, 0, 10, 60 or 400 mg/kg/day for 2 years showed no evidence of any treatment-related effects of the parameters examined and did not result in a statistically significant increased incidence of any tumor type for both mice and rats.

4. Reviewer's Evaluation

This reviewer performed independent analyses on the survival and tumor data submitted by the sponsor, using the programs written by Dr. Ted Guo of Division of Biostatistics II. The primary statistical methods used were described by Peto *et al.* (1980), and Lin and Ali (1994). These methods adjust differences in animal mortality and take the fatal or prevalence context of observation of the tumor into consideration. The intervals used for the adjustment of mortality were 0-52, 53-78, 79-91 and 92-103 weeks and terminal sacrifice for animals. The actual doses were used as weights in the analyses.

The statistical analyses of carcinogenicity study data consisted of two parts, namely, the survival data analysis and the tumor data analysis. The survival data analysis was: 1) to examine the differences in survival distributions among the treatment groups (homogeneity test); and 2) to determine if there is a positive trend in the proportion of deaths with respect to the dose levels (Trend test). Two statistical tests were used in the survival data analysis: the Cox test and the generalized Kruskal-Wallis test. The theoretical background of these tests was described by Lin and Ali (1994) and Thomas *et al.* (1977).

The tumor data analysis was to determine if there is a positive trend in the proportions of a selected tumor type in a selected organ/tissue with respect to the dose levels. The tumors were classified as either fatal (lethal) or non-fatal (non-lethal), according to Peto *et al.* (1980). The reviewer applied the death-rate method to fatal tumors and the prevalence method to non-fatal tumors. For tumors that caused death for some, but not for all, animals, a combined test was performed.

A rule for adjusting the effect of multiple testings proposed by Haseman (1983) can be used to adjust for the effect of multiple testings in pairwise comparisons. Haseman's rule says that rare tumors should be tested at 0.05 level of significance and common tumors should be tested at 0.01 level of significance. A similar rule proposed by the Office of Biostatistics, CDER/FDA for trend tests was used in this review for tests for positive trend. The rule states that in order to keep the overall false-positive rate at the nominal level of approximately 0.1, tumor types with spontaneous tumor rates of 1% or less (rare tumors) should be tested at 0.025 significance level, otherwise (common tumors) at 0.005 significance level (Lin and Rahman, 1998).

4.1 Evaluation of Carcinogenicity Study on Mice

This reviewer's evaluation comprises the following components:

- Survival data analysis
- Tumor data analysis

4.1.1 Survival Data Analysis of Mice

The survival data analysis determines whether the dose-mortality trend in mortality is statistically significant. A positive result indicates that mortality increases as the dose level increases.

- Tables 2 and 3 present the cumulate percentages of death by dose group for female and male, respectively. The time interval "Final Kill 104-106" presents the terminal-sacrifice interval.
- Figures 1 and 2 present plots of Kaplan-Meier estimates of the survival distributions of the treatment groups for female and male, respectively.
- Tables 4 and 5 present results of test for dose-mortality trend for female and male using the methods described in the paper "Trend and Homogeneity Analysis of Proportions and Life Table Data" version 2.1, by Donald G. Thomas, National Cancer Institute.

**APPEARS THIS WAY
ON ORIGINAL**

Table 2
Cumulative Percentages of Death in Female Mice

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
CTR1	0-52	50	1	49	98	2
	53-78	49	7	42	84	16
	79-91	42	8	34	68	32
	92-103	34	7	27	54	46
	FINALKILL104-106	27	27	0	54	46
CTR2	0-52	51	3	48	94.1	5.9
	53-78	48	8	40	78.4	21.6
	79-91	40	4	36	70.6	29.4
	92-103	36	11	25	49	51
	FINALKILL104-106	25	25	0	49	51
LOW	0-52	50	0	50	100	0
	53-78	50	4	46	92	8
	79-91	46	6	40	80	20
	92-103	40	11	29	58	42
	FINALKILL104-106	29	29	0	58	42
MED	0-52	51	2	49	96.1	3.9
	53-78	49	4	45	88.2	11.8
	79-91	45	9	36	70.6	29.4
	92-103	36	9	27	52.9	47.1
	FINALKILL104-106	27	27	0	52.9	47.1
HIGH	0-52	52	4	48	92.3	7.7
	53-78	48	8	40	76.9	23.1
	79-91	40	7	33	63.5	36.5
	92-103	33	7	26	50	50
	FINALKILL104-106	26	26	0	50	50

Figure 1
Kaplan-Meier Survival Functions for Female Mice

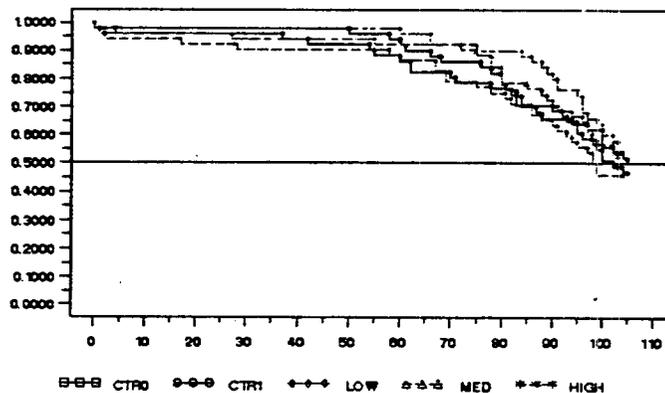
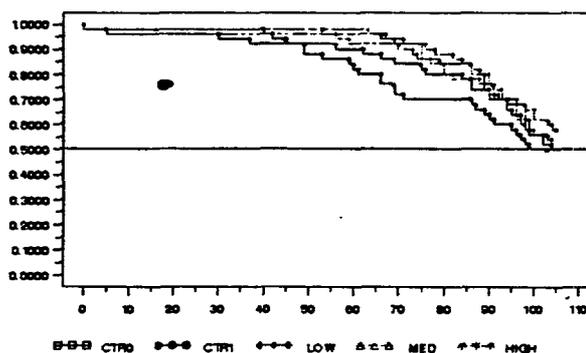


Table 3
Cumulative Percentages of Death in Male Mice

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
CTR1	0-52	50	4	46	92	8
	53-78	46	10	36	72	28
	79-91	36	5	31	62	38
	92-103	31	6	25	50	50
	FINALKILL104-106	25	25	0	50	50
CTR2	0-52	50	3	47	94	6
	53-78	47	6	41	82	18
	79-91	41	4	37	74	26
	92-103	37	9	28	56	44
	FINALKILL104-106	28	28	0	56	44
LOW	0-52	50	1	49	98	2
	53-78	49	6	43	86	14
	79-91	43	7	36	72	28
	92-103	36	8	28	56	44
	FINALKILL104-106	28	28	0	56	44
MED	0-52	50	0	50	100	0
	53-78	50	6	44	88	12
	79-91	44	6	38	76	24
	92-103	38	10	28	56	44
	FINALKILL104-106	28	28	0	56	44
HIGH	0-52	50	0	50	100	0
	53-78	50	5	45	90	10
	79-91	45	7	38	76	24
	92-103	38	7	31	62	38
	FINALKILL104-106	31	31	0	62	38

Figure 2
Kaplan-Meier Survival Functions for Male Mice



The dose-mortality trend tests for female mice (presented in Table 4) and male mice (presented in Table 5) are not statistically significant using the Cox test and the Kruskal-Wallis test.

Table 4
Results of Tests for Dose-Mortality trend for Female Mice

	Method			
	Cox		Kruskal-Wallis	
	Statistics	P-Value	Statistics	P-Value
Depart from Trend	1.3248	0.7224	1.9539	0.5820
Dose-Mortality Trend	1.3346	0.2480	1.6974	0.1926
Homogeneity	2.6630	0.6157	3.6513	0.4552

Table 5
Results of Tests for Dose-Mortality trend for Male Mice

	Method			
	Cox		Kruskal-Wallis	
	Statistics	P-Value	Statistics	P-Value
Depart from Trend	1.1457	0.766	2.0868	0.5546
Dose-Mortality Trend	1.0038	0.3164	1.4251	0.2326
Homogeneity	2.1495	0.7083	3.5119	0.4761

4.1.2 Tumor Data Analysis for Mice

The tumor data analysis determines whether the dose-tumor positive linear trend in tumor incidence is statistically significant. This reviewer tested this trend for every organ and tumor combination with the data provided by the sponsor. This reviewer analyzed the dose-tumor trend among the two control groups and 3 treated groups. The daily doses 0, 0, 10, 60, and 400 mg/kg/day were used as weights for those tests. The time intervals used for the adjustment of mortality were 0-52, 53-78, 79-91, 91-103 weeks, and terminal sacrifice. The resulting p-values are compared against the p-value cutoff point set by the FDA procedures.

This reviewer performed an additional statistical analysis combining hepatocellular adenoma and hepatocellular carcinoma in liver and combining hemangiosarcomas in all organs.

This reviewer also performed pairwise comparisons between the two control groups and the medium dose group as well as between the two control groups and the high dose group.

Tables 6a and 6b contain incidence rates of the combined tumor types and summaries of all the tumor types with statistically significant positive trends.

Table 6a
Results of Trend Tests of Combined Tumor Types for Male and Female Mice

Organ	Tumor	Tumor-Bearing Animal	P-Value
Female			
	Combined hepatocellular adenoma & hepatocellular carcinoma	3, 1, 2, 2, 2	0.4674 ³
	Combined hemangiosarcoma (18001)	1, 1, 5, 3, 3	0.2891 ²
Male			
	Combined hepatocellular adenoma & hepatocellular carcinoma	14,17,21,15,28	0.0066 ¹
	Combined hemangiosarcoma (18001)	0, 1, 3, 3, 2	0.4162 ²

Table 6b
Significant Trends or Differences in Tumor Incidence for Male and Females Mice

Organ	Tumor	Tumor-Bearing Animal	P-Value
Female			
Liver (1800)	Hemangiosarcoma (180017)	0, 1, 1, 0,3	0.0092 ¹
Pairwise comparison between control groups and high dose group			
Liver (18000)	Hemangiosarcoma (180017)	0, 1, 3	0.0381 *
Pairwise comparison between control group 2 and the medium dose group			
	Combined alveolar/bronchiolar adenomas & alveolar/bronchio carcinomas in Lung	6, 18	0.0053 **
Pairwise comparison between control group 2 and the high dose group			
	Combined alveolar/bronchiolar adenomas & alveolar/bronchio carcinomas in Lung	6, 17	0.0099 **
Male			
Liver (1800)	Hepatocellular adenoma (180011)	13,15,18,12,25	0.0090 ¹
Hemolym Tissue (4500)	Malignant lymphoma (450001)	4, 1, 6, 5,3	0.0307 ¹

* indicates statistically significant at level 0.05 (pairwise comparison in tumors with spontaneous tumor rates of 1% or less)

** indicates statistically significant at level 0.01 (pairwise comparison in tumors with spontaneous tumor rates of more than 1%)

¹: Using Asymptotic p-value, since the overall tumor type is both fatal and incidental with spontaneous tumor rates of more than 1%, it should be tested at 0.005 significant level

²: Using Asymptotic p-value, since the overall tumor type is both fatal and incidental with spontaneous tumor rates of 1% or less, it should be tested at 0.025 significant level.

³: Using Exact p-value, since the overall tumor type is incidental with spontaneous tumor rates of more than 1% , it should be tested at 0.005 significant level.

The results of tumor analysis are as follows:

- No significant positive linear trend in incidence rates in tumor data in both males and females.
- Statistically significant difference in tumor incidence (p=0.0381) in hemangiosarcoma in liver for females when comparing control group 1 with the high dose group.
- Statistically significant difference in tumor incidence (p=0.0053) in combined alveolar/bronchiola adenoma and alveolar/bronchiola carcinoma in lung for females

when comparing control group 2 with the medium dose group.

- Statistically significant difference in tumor incidence ($p=0.0099$) in combined alveolar/bronchiola adenoma and alveolar/bronchiola carcinoma in lung for females when comparing control group 2 with the high dose group.
- No statistically significant positive trend in tumor incidence in hemangiosarcoma in all organs combined in both males and females.
- No statistically significant positive trend in tumor incidence in hepatocellular adenoma and hepatocellular carcinoma combined in liver in both males and females.

4.1.3 Conclusion of Mouse Study

In the 2-year mouse study, there were no significant positive trend in survival and statistically significant difference in survival distributions among differences in survival between treatment groups in both females and males. There were statistically significant differences in tumor incidence in hemangiosarcoma in liver ($p=0.0381$) in females when comparing the control groups with the high dose group; in alveolar/bronchiolar adenoma and alveolar/bronchiola carcinoma combined in lung in females when comparing control group 2 with the medium dose group, and when comparing control group 2 with the high dose group.

4.2 Evaluation of Carcinogenicity Study on Rats

This reviewer's evaluation comprises the following components:

- Survival data analysis
- Tumor data analysis

4.2.1 Survival Data Analysis of Rats

The survival data analysis determines whether the dose-mortality trend in mortality is statistically significant. A positive result indicates that mortality increases as the dose level increases.

- Tables 7 and 8 present the cumulative percentages of death by dose group for female and male, respectively. The time interval "Final Kill 104-106" presents the terminal-sacrifice interval.
- Figures 3 and 4 present plots of Kaplan-Meier estimates of the survival distributions of the treatment groups for female and male, respectively.
- Tables 9 and 10 present results of test for dose-mortality trend for female and male using the methods described in the paper "Trend and Homogeneity Analysis of Proportions and Life Table Data" version 2.1, by Donald G. Thomas, National Cancer Institute.

Table 7
Cumulative Percentages of Death in Female Rats

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
CTR 1	0-52	50	2	48	96	4
	53-78	48	4	44	88	12
	79-91	44	5	39	78	22
	92-103	39	9	30	60	40
	FINALKILL104-106	30	30	0	60	40
CTR 2	0-52	50	1	49	98	2
	53-78	49	5	44	88	12
	79-91	44	3	41	82	18
	92-103	41	7	34	68	32
	FINALKILL104-106	34	34	0	68	32
LOW	0-52	50	4	46	92	8
	53-78	46	3	43	86	14
	79-91	43	5	38	76	24
	92-103	38	10	28	56	44
	FINALKILL104-106	28	28	0	56	44
MED	0-52	50	5	45	90	10
	53-78	45	5	40	80	20
	79-91	40	3	37	74	26
	92-103	37	10	27	54	46
	FINALKILL104-106	27	27	0	54	46
HIGH	0-52	50	4	46	92	8
	53-78	46	4	42	84	16
	79-91	42	6	36	72	28
	92-103	36	5	31	62	38
	FINALKILL104-106	31	31	0	62	38

Figure 3
Kaplan-Meier Survival Functions for Female Rats

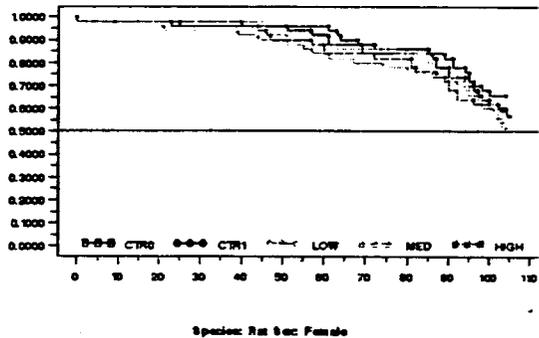
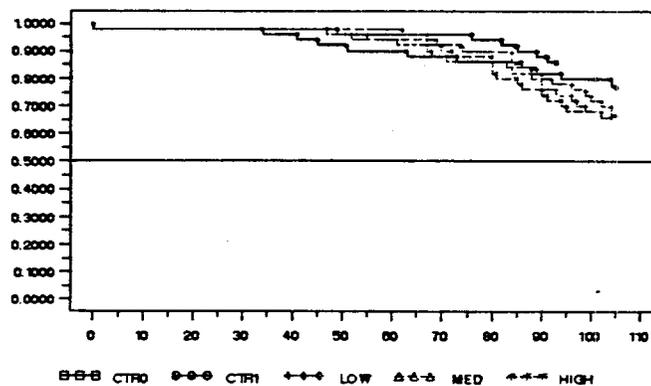


Table 8
Cumulative Percentages of Death in Male Rats

Analysis of Mortality		No. Risk	No. Died	No. Alive	Pct Survival	Pct Mortality
CTR0	0-52	50	1	49	98	2
	53-78	49	1	48	96	4
	79-91	48	4	44	88	12
	92-103	44	1	43	86	14
	FINALKILL104-106	43	43	0	86	14
CTR1	0-52	50	4	46	92	8
	53-78	46	2	44	88	12
	79-91	44	2	42	84	16
	92-103	42	2	40	80	20
	FINALKILL104-106	40	40	0	80	20
LOW	0-52	50	2	48	96	4
	53-78	48	4	44	88	12
	79-91	44	4	40	80	20
	92-103	40	6	34	68	32
	FINALKILL104-106	34	34	0	68	32
MED	0-52	50	4	46	92	8
	53-78	46	5	41	82	18
	79-91	41	5	36	72	28
	92-103	36	36	0	0	100
	FINALKILL104-106	0	0	0	0	100
HIGH	0-52	50	1	49	98	2
	53-78	49	4	45	90	10
	79-91	45	8	37	74	26
	92-103	37	3	34	68	32
	FINALKILL104-106	34	34	0	68	32

Figure 4
Kaplan-Meier Survival Functions for Male Rats



The dose-mortality trend for female rats (presented in Table 9) and male rats (presented in Table 10) are not statistically significant using the Cox test and the Kruskal-Wallis test .

Table 9
Results of Tests for Dose-Mortality trend for Female rats

	Method			
	Cox		Kruskal-Wallis	
	Statistics	P-Value	Statistics	P-Value
Depart from Trend	2.6324	0.4518	2.4708	0.4806
Dose-Mortality Trend	0.0633	0.8013	0.1750	0.6757
Homogeneity	2.6957	0.6100	2.6459	0.6187

Table 10
Results of Tests for Dose-Mortality trend for Female rats

	Method			
	Cox		Kruskal-Wallis	
	Statistics	P-Value	Statistics	P-Value
Depart from Trend	4.3806	0.2232	4.0242	0.2589
Dose-Mortality Trend	2.6802	0.1016	2.5564	0.1098
Homogeneity	7.0609	0.1327	6.5806	0.1598

4.2.2 Tumor Data Analysis for Rats

The tumor data analysis determines whether the dose-tumor positive linear trend in tumor incidence is statistically significant. This reviewer tested this trend for every organ and tumor combination with the data provided by the sponsor. This reviewer analyzed the dose-tumor trend among the two control groups and 3 treated groups. The daily doses 0, 0, 10, 60, and 400 mg/kg/day were used as weights for those tests. The time intervals used for the adjustment of mortality were 0-52, 53-78, 79-91, 91-103 weeks, and terminal sacrifice. The resulting p-values are compared against the p-value cutoff point set by the FDA procedures.

This reviewer performed an additional statistical analysis combining hepatocellular adenoma and hepatocellular carcinoma in liver and combining hemangiosarcomas in all organs.

This reviewer also performed pairwise comparisons between the two control groups and the medium dose group as well between the two control groups and the high dose group.

Table 11a and 11b contain incidence rates of the combined tumor types and summarizes of all the tumor types with statistically significant positive trends.

Table 11a
Results of Trend Tests of Combined Tumor Types for Male and Female Rats

Organ	Tumor	Tumor-Bearing Animal	P-Value
Female			
Combined hepatocellular adenoma & hepatocellular carcinoma		0, 1, 5, 5, 2	0.4760 ²
Combined hemangiosarcoma (18001)		0, 1, 1, 1, 0	0.7842 ²
Male			
Combined hepatocellular adenoma & hepatocellular carcinoma		2, 4, 3, 4, 7	0.0036 ¹ **
Combined hemangiosarcoma (18001)		1, 0, 2, 1, 0	0.8640 ²

Table 11b
Significant Trends or Differences in Tumor Incidence for Male and Females Rats

Organ	Tumor	Tumor-Bearing Animal	P-Value
Female			
Uterus (3400)	Adenocarcinoma (340016)	3, 0, 1, 1, 5	0.0101 ¹
Pairwise comparison between control 1 and medium dose group			
Mammary Gland	Adenocarcinoma	0, 10	0.0012 ¹ *
Male			
Subcutaneous Tissue (3100)	Malignant Schwannoma (310001)	0, 1, 0, 0, 1	0.0082 ¹
Subcutaneous Tissue (3100)	Fibroma (310009)	0, 1, 1, 0, 1	0.0008 ¹ **

* indicates statistically significant at level 0.01 (pairwise comparison in tumors with spontaneous tumor rates of more than 1%)

** indicates statistically significant at level 0.005

1: Using Asymptotic p-value, since the overall tumor type is both fatal and incidental with spontaneous tumor rates of more than 1%, it should be tested at 0.005 significant level.

2: Using Asymptotic p-value, since the overall tumor type is both fatal and incidental with spontaneous tumor rates of 1% or less, it should be tested at 0.025 significant level.

The results of tumor analysis are as follows:

- Statistically significant positive dose-response relationship (p=0.0012) in denocarcinoma in mammary gland for males when comparing control group 1 with medium dose group.
- Statistically positive dose-response relationship (p=0.0008) in incidence rate of fibroma (310009) in subcutaneous tissue (3100) in male rats.
- Statistically positive dose-response relationship (p=0.0036) in combined hepatocellular adenoma and hepatocellular carcinoma in liver for males.
- No significant positive linear trend in incidence rates in tumor data in females.
- No statistically significant positive dose-response relationship combined hemangiosarcoma in all organs for both males and females.
- No statistically significant positive dose-response relationship in combined hepatocellular adenoma and hepatocellular carcinoma in liver for females.

4.2.3 Conclusion of Rat Study

In the 2-year rat study, there was no significant positive trend in survival and statistically significant difference in survival distributions among treatment groups in both females and males. There were statistically significant positive-dose relationships in fibroma in subcutaneous tissue ($p=0.0008$) in males, in hepatocellular adenoma and hepatocellular carcinoma combined in liver in males. There was statistically significant difference in tumor incidence in adenocarcinoma in mammary gland ($p=0.0012$) in females when comparing control group 1 with the medium dose group.

5. References

- 1) Bart, Chi, and Tarone (1979). "Statistical Issues in Interpretation of Chronic Bioassay Tests for Carcinogenicity." Journal of the National Cancer Institute. Vol. 62, pp.957-974.
- 2) Chu, Cueto and Ward (1981). "Factors in the evaluation of 200 national cancer institute carcinogen bioassay." Journal of Toxicology and environmental Health. Vol. 8, pp.251-80.
- 3) Haseman, J. K. (1983). "A re-examination of false-positive rates for carcinogenesis studies." Fundamental and Applied Toxicology, 3, pp.334-9
- 4) Haseman, J. K. (1985). "Issues in carcinogenicity testing: Dose selection." Fundamental and Applied Toxicology. Vol. 5. Pp. 66-78.
- 5) Lang, P. L. (1992). "Spontaneous neoplastic lesions and selected non-neoplastic lesions in the Crl: — rat." Table 5b. neoplasms 24 Month Studies Female CD®Rats. pp.23
- 6) Lin, K. K. and M. Ali (1994), "Statistical Review and Evaluation of Animal Tumorigenicity Studies." Statistical in the Pharmaceutical Industry, Second Edition, Revised and Expanded, edited by C.R.Buncher and J.Y. Tsay, Marcel Dekker, Inc., New York. pp. 19-57.
- 7) Lin, K. K. and M. A. Rahman (1998). "Overall False Positive Rates in Tests for Linear Trend in Tumor Incidence in Animal Carcinogenicity Studies of New Drug." Journal of Biopharmaceutical Statistics, 8(1), 1-15 (1998)
- 8) Peto *et al* (1980). "Guidelines for Simple, Sensitive Significance Tests for Carcinogenic Effects in Long-Term Animal Experiments," In Long-Term and Short-Term Screening Assays for Carcinogens: An Critical Appraisal, International Agency for Research on Cancer, Lyon, France." IARC monographs supplement, 2, pp.311-426
- 9) Thomas *et al* (1977). "Trend and Homogeneity Analyses of Proportions and Life Table Data," Computer and Biomedical Research, 10, pp.373-381.

/S/

Moh-Jee Ng
Mathematical Statistician

/S/

Concur:

Karl Lin, Ph.D.
Expert Mathematical Statistician
(Applications in Pharmacology & Toxicology)

cc: Original NDA 21-318
HFD-510/Division File
HFD-510/YShin, AJordan
HFD-715/Division File, Chron
HFD-715/ENevius, MWelchmh, Canello, KLin, MNg

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

/s/

Moh-Jee Ng
12/20/01 07:33:12 AM
BIOMETRICS

Karl Lin
12/21/01 07:50:57 AM
BIOMETRICS
Concur with review

Executive CAC

Date of Meeting: 1/15/02

Mouse/Rat Carcinogenicity Study

Committee: Joseph Contrera, Ph.D., HFD-901, Acting Chair
Bob Osterberg, Ph.D., HFD-520, Alternate Member
John Leighton, Ph.D., HFD-150, Alternate Member
Alex Jordan, Team Leader
Yangmee Shin, Presenting Reviewer

Author of Draft: Yangmee Shin

The following information reflects a brief summary of the Committee discussion and its recommendations. Detailed study information can be found in the individual review.

NDA #21-368

Drug Name: Cialis

Sponsor: Lilly ICOS LLC

Background: Cialis is a PDE5 inhibitor being developed for treatment of erectile dysfunction. The Sponsor initiated the 2-year oral gavage carcinogenicity studies in rats and mice and selected the high doses based on saturation of absorption at doses >400 mg/kg/day prior to review by the Executive CAC. However, parent drug exposure increased slightly with doses up to 2000 mg/kg/day for rats (R20791) and up to 800-1200 mg/kg/day in mice (M21259) and more importantly, there were no data showing that drug metabolites did not accumulate with higher doses.

On June 16, 1999, the Executive CAC concurred on doses of 10, 60 and 400 mg/kg based on AUC multiples of free drug for both rats and mice. Subsequently, the Sponsor increased the clinical dose from 10 mg to 20 mg. This resulted in an AUC in humans of approximately 7700 ng.hr/ml (LVDK), nearly 4 times higher than that had been reported for the 10 mg dose. As a result, the AUC multiples between rodents and humans were reduced approximately 4 fold.

Mouse Carcinogenicity Study: There were no statistically significant increases in any tumors observed in the studies. The AUC's (males and females) for the unbound parent drug were approximately 10 times the human AUC at the proposed clinical dose of 20 mg.

Rat Carcinogenicity Study: There were no statistically significant increases in any tumors observed in the studies. The AUC's for the unbound parent drug were approximately 14 times in males and 26 times in females the human AUC at the proposed clinical dose of 20 mg.

Executive CAC Recommendations and Conclusions:

The Committee concluded that no evidence of carcinogenicity was seen in rats or mice.

The Committee noted that evidence of saturation of absorption such as measurement of metabolites or total radioactivity in plasma with increasing doses was not provided. Furthermore, there was no evidence of dose limiting toxicity at the high doses tested.

The Committee noted that the AUC ratio for the drug was below 25 in male rats. The doses produced acceptable drug blood levels in female rats.

For mice, The Committee found that adequate exposure was not achieved for males or females, as dose ratios were well below the 25-fold minimum required for a valid study based on pharmacokinetics.

The Committee noted that Cialis has been adequately investigated for carcinogenicity only in female rats. Studies in male rats and male and female mice were performed at doses below those recommended by the ICH guidelines. Therefore, the Committee concluded that the carcinogenic potential of Cialis has not been adequately investigated and recommended that an additional alternative mouse carcinogenicity assay be conducted. Alternatively, the Committee felt that if the Sponsor could provide evidence of saturation of absorption at the high doses tested, an additional study would not be necessary.

The Committee felt that the negative study results in adequately dosed female rats, taken together with the data from the inadequately dosed male rats and male and female mice, did provide some assurance of safety. Therefore the Committee would concur with the Division allowing the additional mouse carcinogenicity study to be completed after drug approval.

|S|

Joseph Contrera, Ph.D.
Acting Chair, Executive CAC

cc:\n
/DRUDP, Division File, HFD-580
/Alex Jordan, Team leader, HFD-580
/Yangmee Shin, Reviewer, HFD-580
/Dornette Spell-Lesane, Project manager, HFD-580
/Adele Seifried, HFD-024

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

/s/

Joe Contrera
2/22/02 01:57:18 PM

Statistical Review and Evaluation
Clinical Studies

NDA#: 21-368

Applicant: Lilly Icos LLC

Name of Drug: Cialis (tadalafil)

Documents Reviewed: Vols. 90-106 dated June 28, 2001

Medical Officer: Ashok Batra, M.D., HFD-580

Statistical Reviewer: David Hoberman, Ph.D., HFD-715

Background

The sponsor has submitted six placebo-controlled, randomized, double-blind, multi-center, international trials as evidence of tadalafil's efficacy and safety in the treatment of erectile dysfunction. There are other trials to assess _____ and comparisons to sildenafil. This review examines the six major trials and the two trials conducted to address _____ only.

Studies LVDJ, LVCO, LVCE, LVBN, LVBK, LVCQ

All of these 6 Phase 3 trials except LVCQ consisted of a 4 week run-in period and a 12 week double-blind period. Trial LVCQ had a 26 week double-blind period. All were placebo controlled. LVCQ studied 20 mg tadalafil. LVDJ, LVCO and LVBK studied 10 mg and 20 mg tadalafil. LVBN studied 5 mg and 10 mg tadalafil, and LVCE studied 2.5 mg, 5 mg and 10 mg tadalafil. LVCQ was conducted in Australia, LVDG was conducted in Canada, LVCO in the Republic of China, and LVBN in Argentina, Canada and Mexico, LVCE in Canada, and LVBK in Spain with **diabetic patients, only**. The primary endpoints were the same in each trial: the score on the Erectile Function (EF) domain of the IIEF (International Index of Erectile Function), and the answers to Sexual Encounter Profile (SEP) Questions 2 (Were you able to insert your penis into your partner's vagina?) and 3 (Did your erection last long enough for you to have successful intercourse?). Patients made random numbers of attempts at intercourse within the first, second and third month. Patients had to have attempted at least 4 times during the run-in period in order to establish baseline measurements. For SEP 2 and SEP 3, the unit of analysis was the proportion of successful attempts in a month. **Statistical significance compared to placebo at 12 weeks was required on all three endpoints to declare a 'positive' trial.** Statistical analysis centered on the 'change from baseline' for the three endpoints using the LOCF method of imputation of missing data.

In each study, sample sizes were determined using bootstrap techniques to provide approximately 90% power to detect differences from placebo and the least effective dose of tadalafil. The sponsor used data from another trial (LVAC) from this purpose.

Results

Selected baseline characteristics were as follows: the range of mean ages at entry over the 6 studies was 56 to 60. Virtually all patients were Caucasian except for LVBN (17% neither Caucasian nor Asian) and LVCO (100% Asian). The range of percentages of subjects who had a history of coronary artery disease was 5%-12%. Cardiovascular history for the study in diabetics (LVBK) was unavailable. The range of mean IIEF Erectile Domain Scores was 12 to 15 among the studies. In general, between 80%-90% of the patients completed the full double-blind period of the studies.

The sponsor's **Table 1** displays the raw and normalized number of attempts as assessed by SEP Question 3. Note that in some trials, there is evidence that, on average, placebo subjects made some 5 fewer attempts than tadalafil subjects.

The sponsor's **Table 2** displays the randomized sample sizes and results, including baseline means, for the three endpoints in each of the six trials. These are LOCF analyses 'carried forward' to the 3 month endpoint (6 months in the case of LVCQ).

Reviewer's Comments

It is evident from the sponsor's analyses of these six trials that statistical significance between tadalafil groups and placebo has been achieved for 10 mg and 20 mg. However, the analyses concentrate upon performance at time points and do not indicate an expectation of performance for an individual over the course of 3 months. As an alternative description of the data, this reviewer has done the following:

- 1) Define a "clinical response" based upon the proportion of "yes" responses to SEP Question 3 at each of the 3 months. For purposes of illustration, there are 2 different 'thresholds' of clinical response: at least 1 success (i.e. greater than 0%), and at least 50% success.
- 2) Since there are 3 months, there are $2 \times 2 \times 2 = 8$ patterns (permutations) of response (Y=Yes, N=No) over the 3 months of the trial: The ordering used for this analysis is NNN, YNN, NYN, NNY, YYN, YNY, NYY, YYY. Subjects are classified into one of these 8 response categories and histograms are produced. The ordering is somewhat arbitrary. In this case, later response has been deemed more desirable than losing an earlier response.

Figures 1 and 2 display the histograms for trial LVDJ, one for the 0% threshold of success and the other for 50%, respectively. The pattern of response over 3 months is indexed on the horizontal axis. The most compelling feature of the patterns for LVDJ and the other trials (except for LVCO) is that, regardless of the threshold of response, the distributions are strictly bimodal

at the two extremes. The heights of the 6 central treatment group clusters of bars (indicating inconsistent response over the 3 months) are small and apparently random among the groups. In other words, the majority of patients either *never* get a meaningful clinical response (however defined) over the 3 months or they do respond *every* month.

Further, inspection of this characteristic pattern in each study reveals (not shown) that baseline performance plays a significant role. For instance, in trial LVDJ, 50% of the patients had *no* success during the baseline run-in period. Thus, the 'mean' baseline response reported by the sponsor in Table 2 is misleading because of the skewed quality of the distribution of baseline percentages of success. *Further, when the 8 pattern bar graphs are stratified by whether or not patients had zero or greater than zero success at baseline, the result is that the bimodal feature appears only in those patients who had no success at baseline.* For those who did have at least one success at baseline, the overwhelming majority of patients responded for all 3 months (i.e. fell in stratum 8) with the same random performances of the treatment groups among the 6 middle clusters. *The upshot of these observations is that the patients in the YYY stratum provide nearly all the information about distinguishing active drug from placebo, regardless of baseline.*

Consequently, an approach to describing the data in a clinically relevant way is the following: We restrict the description to patients in the YYY stratum, i.e. responders for all 3 months, given thresholds of response which run from 0% to 100%. We then plot the percentage of patients in each treatment group who were responders (i.e. attained *at least* a given percentage of success: the "threshold") against the value of the threshold on the horizontal axis. The result is necessarily a non-increasing curve, since the greater percentage of success required to reach the threshold for all 3 months, the fewer the patients who will achieve this 'higher bar'. Figures 3-8 display the plots for each of the 6 major studies. For example, in trial LVDJ (Fig. 3), the plot indicates that, in the 20 mg group, about 48% of the patients achieved *at least* 60% success on the SEP 3 for all 3 months. Note that in patients with diabetes (LVBK) performance of the 20 mg dose of tadalafil appears to be below that of 20 mg groups in other trials with healthier patients.

Accounting for 6 Months in Trial LVCQ

In trial LVCQ which was 6 months long, the percentage of patients in the placebo group who *failed to respond in all 6 months* was 43% while that in the 20 mg group was 10%. The percentage of patients in the placebo group who *achieved at least 1 response in all 6 months* was 30%, while that in the 20 mg group was 62%. When the threshold for response is at least 50% success in a month, the percentage of patients in the placebo group who *failed to respond in all 6 months* was 53% while that in the 20 mg group was 15%. The percentage of patients in the placebo group who *achieved at least 1 response in all 6 months* was 13% while that in the 20 mg group was 54%.

It is also of interest to examine the association of **no successes during the run-in period to no successes over the 3 month trial**. The percentage of patients who did not have any successes during the run-in period was approximately 50% in each group in each trial. Restricting the denominator to those 50%, the table below displays the percentage of each run-in no-success subgroup which went on to have no successes for the double-blind portion of the trial.

	<u>LVDJ</u>	<u>LVCO</u>	<u>LVBN</u>	<u>LVCE</u>	<u>LVBK</u>	<u>LVCQ</u>
Placebo	73%	34%	62%	71%	65%	84%
tad. 2.5 mg	----	----	----	43%	----	----
tad. 5 mg	----	----	50%	35%	----	----
tad. 10 mg	38%	5%	33%	43%	38%	----
tad. 20 mg	31%	16%	----	----	27%	27%

Finally, there may be interest in the *conditional* probability that a patient “responds” for all 3 months *given that he has responded in the first month*. Pooling the 4 trials which included 20 mg, the result is that when response is defined as “at least 50% success”, then the estimated conditional probability for the 20 mg group is 83%.

Trial LVDG: _____

The purpose of this 8-week placebo controlled trial was to measure the _____ of tadalafil 20 mg at 24 and 36 hours after dosing. After a 4-week run-in period, subjects were stratified by numerical category of the Erectile Function Domain of the IIEF: mild: ED=17-30, moderate: ED=11-16, and severe: ED=1-10). Patients were then randomized into either of two sequences shown below.

Order Group	First dose Time of attempt	Second dose Time of attempt	Third dose Time of attempt	Fourth dose Time of attempt
Sequence A	24 hours	24 hours	36 hours	36 hours
Sequence B	36 hours	36 hours	24 hours	24 hours

Patients were to attempt intercourse either 24 or 36 hours after dosing depending upon the period of the study shown in the table above. This was *not* a treatment ‘crossover’ trial. Thus, there were 4 doses of either placebo or 20 mg tadalafil, 2 for each attempt time point (24 and 36 hours) after dosing. Subjects were instructed to take two doses over approximately 2 to 3 weeks with each dose separated by 8-10 days.

The sponsor’s rationale for the study was to compare the treatment groups at the two attempt time points after dosing. Statistical significance of a test comparing 20 mg tadalafil to placebo indicated “response” at that time and therefore _____ was interpreted to exist up to at least the time point of statistical significance. For this purpose, the response (yes/no) to the

SEP (Sexual Encounter Profile) Question 3 (“Did your erection last long enough for you to have successful intercourse ?”) was the primary clinical endpoint.

To compare groups, the sponsor used a repeated measures categorical model separately for the 24 and 36 hour analyses. Since there were two doses at each time point, the repeated measures were the two yes/no responses at 24 hours and the two yes/no responses at 36 hours. Although several attempts might be made after a dose, only those in the following time windows were assigned to the particular time point: for the 24 hour time point, responses between 20 and 30 hours were used. For the 36 hour time point, responses between 30 and 48 hours were used. The statistical test at 36 hours would be done only if the test at 24 hours was significant. The primary analysis used only the *first* attempt in the time window on each of the 2 doses.

The planned sample size was 134 subjects/group to provide 90% power to detect the difference of 30% ‘yes’ response on placebo and 50% on tadalafil at 24 hours and to provide at least 95% power to detect a difference of 30% and 60%, respectively.

Results of the Primary Analysis

A total of 348 patients were randomized within 34 centers in Europe (57%) and the United States (43%). The randomized sample sizes were 175 in the tadalafil group and 173 in the placebo group. According to the sponsor, 94% of the subjects “completed the protocol”. There were no important baseline factor imbalances between the two groups. The sponsor’s Table 3 displays the number of patients who were available for each of the time point analyses, and Table 4 displays the attempt time distributions for each group. For purposes of analysis of the primary endpoint, however, there are 163 subjects in the placebo group and 171 subjects in the tadalafil group. Thus 14 patients did not provide data for the primary analysis.

The sponsor’s Table 5 displays the results of the primary analysis. Note again that the table and the repeated measures analysis pool the results from the *first attempts* on both doses for each time point. Thus the results shown are in terms of attempts and not number of patients. It is clear that statistically significant differences between tadalafil and placebo occur at the 24 and 36 hour time points. The sponsor’s major conclusion of the study is that _____ of responsiveness to a single 20-mg dose of IC351 is a least 36 hours, as measured by Question 3 of the Sexual Encounter Profile.”

APPEARS THIS WAY
ON ORIGINAL

Reviewer's Comments

1. This reviewer has found that, even though the data was pooled over the two doses in the table above, the group percentages of response were nearly identical on both doses. However the sponsor does not report the actual numbers of patients in the analyses. The table below displays those numbers:

	<u>24 hours</u>		<u>36 hours</u>	
	<u>Dose1</u>	<u>Dose2</u>	<u>Dose1</u>	<u>Dose2</u>
Tadalafil	142	125	139	136
Placebo	152	134	136	128

The missing values indicated by these numbers being lower than the number of subjects randomized is due to the fact that the sponsor's analysis ignores the fact that some patients did not take both doses or make two attempts (let alone the fact that 14 patients were never included in the analysis evidently due to never providing data in the time windows). In both groups, 20% of the potential attempts are missing. **Thus the validity of the sponsor's tabled numbers depend upon assuming that the probabilities of success would be the same in the sample of missing attempts as those in the sample actually observed.**

2. Another aspect of the results not addressed by the sponsor is the *consistency of response over all four doses within a subject at each time point*. Consequently, this reviewer did the following: At 24 and 36 hours, each group's subjects was divided into 4 categories, each corresponding to the 4 possible patterns of two responses (yes/no) on two successive doses: (no,no), (no,yes), (yes,no), (yes,yes). **Figure 9 displays the bar graph examining this issue. Note that at both time points, 60% of the placebo patients did not respond to the first attempt for either dose, while the respective percentage was 35% in the Tadalafil group. Further, 35%-40% of tadalafil subjects successfully had intercourse on both doses at each time point, while the figures for the placebo were 15-20%.**

3. The major problem with the design of this study is that it ignores the usual notion of what it means to measure '_____'. For instance, if relief of headache is studied, a subject can use a stopwatch to measure the time between relief of the headache and its recurrence. By analogy, this study should have measured the beginning of a "response" to begin with, instead of inferring that subjects were "in response" for the full time before 24 or 36 hours. **What the study as currently constituted shows is simply that there was a percentage of patients in each group whose first attempt at intercourse after dosing was successful at 24 or 36 hours. The 'or' is important here because there was no provision that in order to be declared a responder at 36 hours, a subject had to respond at 24 hours. Otherwise, what is the meaning of _____? This reviewer has examined this issue of _____ at 24 and 36 hours for the purpose of inferring "continued response" at 36 hours.**

If one defines response at 36 hours occurring only if response occurred at 24 hours for a subject, then the results are as follows: **In the tadalafil group, 40% of the randomized subjects with post-baseline data had at least one success on either dose at both 24 and 36 hours. This contrasts sharply with the 60% “response” reported by the sponsor as the percentage of attempts that were successful at 36 hours.** Recall that the sponsor’s analysis ignored the fact that some patients did not take both doses. In the reviewer’s analysis, subjects who did not attempt are treated as failures for that hypothetical attempt. After all, a subject cannot succeed if there is no attempt. This is proper since the protocol is being violated and an intent to treat approach should not ignore the existence of missing data.

Unfortunately there is little data to examine for consistency of success before 24 hours, so that “at least 24 hours duration” could be examined. There were 23 tadalafil subjects who had at least one attempt before 24 hours. Of those 23, 10 had no successes (42%) within the first 24 hours. This is not an encouraging result when “duration” is being evaluated.

Discussion and Conclusion

The sponsor has demonstrated that the tadalafil and placebo groups can be statistically distinguished at 24 and 36 hours after dosing. Whether that result alone should be interpreted as demonstration of “_____” is very dubious. The trial was not designed to confirm that a response was taking place before the cross-sectional time points. At any rate, it would seem odd to declare a response at 36 hours when there was no evidence of effect (success on SEP Question 3) at 24 hours. In that case, the percentage of “responders” in the tadalafil group at 36 hours is more like 40% rather than 60%.

Trial LVCK: TIME TO RESPONSE

The purpose of this placebo-controlled trial was to determine the earliest _____ of 10 mg and 20 mg of tadalafil within the first 30 minutes of dosing. After a 4 week screening period, patients with IIEF scores between 6 and 25 were centrally randomized to the 3 treatment groups. The randomization strata used the following domains for mild, moderate or severe ED using the Erectile Function Domain of the IIEF: 22-25, 11-21, and 6-10, respectively. During the following 25-37 days each subject was to take 4 doses of the assigned medication, each dose separated by 8-10 days. After each dose, each patient used a stopwatch to measure the time until ‘response’: occurrence of erection *and* an answer of “yes” to SEP Question 3 (Did your erection last long enough for you to have successful intercourse?).

The proposed sample size of 66/group was based upon a response rate of 58% within 30 minutes in an active group and 31% in the placebo group providing at least 90% power.

The sponsor's plan was to statistically compare the proportion of successful attempts in the active groups to that in the placebo group sequentially backward at each minute starting at 30 minutes using repeated measures logistic regression (via Generalized Estimating Equations). Data was not recorded after 30 minutes, i.e. patients time to response was censored at 30 minutes. Sequential testing was done by descending minute until the statistical test was no longer significant at the .05 level. The minimum minute at which statistical significance occurred was then declared the "earliest time to onset of action".

Results

At total of 223 subjects were randomized among 10 centers in the United States (Placebo: N=74, 10 mg: N=74, and 20 mg: N=75). Virtually all subjects completed the trial. Figure 10 displays percentage of successful attempts in each treatment group. The sponsor indicates that the first step-down sequential test failure to reach a p-value of below .05 at 15 minutes. Thus, the sponsor's major conclusion is that "_____ of 20 mg IC351 was within _____ after dosing as analyzed for the primary objective"

Reviewer's Comments

1. Note that the conclusion cited above fails to mention that the purpose of the study was to find the _____. Thus, the sponsor's statement in the conclusion is incorrect.
2. The sponsor's method of relying on a placebo comparison is *ad hoc* since the result depends upon the sample sizes. In fact, the role of the placebo group is confusing in this setting. If a comparison to placebo indicates that the drug is efficacious in principle, there would be no need for the sequential comparison to placebo to determine the earliest response time, unless the intent was to somehow "adjust for a placebo response". This is what the sponsor seemed to have in mind. But this rationale confuses the issue. There are two questions that might be addressed: 1) What is the minimum time that the drug takes to exert a pharmacodynamic effect? And 2) What is the minimum time that one sees a defined clinical effect after taking tadalafil? These are two very different questions and trial LVCK can address only question 2. To answer question 1, one would have to correlate drug levels in the blood with clinical response. All in all, it seems more appropriate to work only with the distributions of time to response in the active groups whether or not one is convinced that the null hypothesis of equal efficacy of drug and placebo has been rejected. The sponsor ignores these explicit distributions. A look at the actual distributions is instructive:

This reviewer found that in the 20 mg group, 35% of the randomized patients never got a response in 30 minutes on any of the 4 doses taken during the trial. The respective percentages for the 10 mg and placebo groups were 43% and 51%. These results raise doubts about whether finding a _____ is meaningful in the first place, unless the question is restricted to subjects who responded within 30 minutes for at least 1 dose. In that subset, each subject's vector of minutes to successful erection (note that a "yes" answer to question 3 of the SEP was required) was examined to calculate its minimum over the 4 doses. Figure 11 displays the distribution of the minima of the 49 subjects in the 20 mg tadalafil group who had at

least one successful erection in 4 doses. The fact that the median of the minima is 15 minutes does not mean that it is a clinically meaningful number, for no other reason than that it occurs in the presence of so much variability illustrated in Figure 11. In fact, by censoring data at 30 minutes, the sponsor lost an opportunity to assess the median time to response, clearly a statistic with more potential clinical meaning. After all, information to the patient and physician *via* promotion or the package insert should communicate the “expected” or “typical” response time rather than the most *atypical*. In that regard, it makes much more sense to recognize that each of the 4 doses is a mini-trial producing a *mixture* of two distributions of subjects: 1) those who attain at least one erection at or before 30 minutes and 2) those who may or may not have after 30 minutes (since observation was curtailed after 30 minutes). It is important to consider both distributions because consumers should know that in this trial 35% of the subjects taking 20 mg Tadalafil never attained an erection in 4 doses.

Conclusion

The sponsor’s strategy for finding the “earliest” response time produces a number based on statistical comparisons to placebo. The result is simply an empirical statistical fact of the data set and lacks a scientific interpretation. *Inspection of the data over 4 doses in the 20 mg tadalafil group suggests that there is no single number characterizing _____ which would not be misleading when communicated to the public by the label or promotional material.* First, any measure of minimum time is probably clinically irrelevant. Second, the median time to response is not available due to curtailment of information after 30 minutes. Third, there is too much variability in any distribution derived from this data to assign any clinically meaningful measure of central tendency. The best (if not particularly clinically meaningful) statement seems to be that 65% of the subjects using 20 tadalafil attained at least one successful erection in 4 doses within 30 minutes of dosing.

The Sponsor’s Case for 20 mg Over 10 mg

In a FAX dated March 5, 2002, the sponsor presented an analysis of pooled studies which purported to demonstrate that 20 mg was statistically significantly more effective than 10 mg with regard to the IIEF Erectile Function Domain, SEP Question 2, and SEP Question 3. The problem with this approach is that the sponsor 1) combined all trials with either 10 mg or 20 mg, not both 10 mg and 20 mg and 2) excluded Trial LVBK in Diabetic patients. When the only trials with both doses (LVCO, LVDJ, and LVBK) are combined in a simple way, statistical evidence of 20 mg’s superiority disappears.

Specifically, when trials LVBK, LVCF, LVCO, LVCQ, and LVDJ are combined using ANOVA, the least square means (lsm) of the changes from baseline are tabled below:

	<u>IIEF EF Domain</u>	<u>SEP #2</u>	<u>SEP #3</u>
10 mg	6.4	24.7	34.7
20 mg	8.6	31.4	40.8

The sponsor then reports p-values below .05 for all three comparisons of 10 mg to 20 mg.

However, examination of the results in Table 2 for the individual trials LVCO, LVDJ, and LVBK (the 3 trials which included both 10 mg and 20 mg doses) indicates that there is no evidence of 20 mg' superiority to 10 mg. This reviewer has performed a simple meta-analysis using the sponsors results. That analysis fails to provide statistical evidence of superiority.

Overall Conclusions

The submitted trials provide evidence that tadalafil 20 mg and 10 mg are statistically different from placebo. Clinically relevant effects are essentially the same for 10 mg and 20 mg. Evidence is weaker for 5 mg. Review of the results in Table 2 and subsequent pooling of appropriate trials by this reviewer indicate no statistically persuasive evidence that 20 mg is more efficacious than 10 mg. Results for the two highest doses also suggest that tadalafil may be slightly less effective in patients with diabetes.

Finally, the data suggest that the difference between exposure to tadalafil and placebo lies in the level of continued favorable response over the three months of the trials. There is a high probability that success in the first month will lead to further success over at least the next two months.

For purposes of the package insert or marketing and promotion, _____ is too variable to assign a specific number characterizing this endpoint. Moreover, the study was flawed by censoring response after 30 minutes. _____ is not well-defined in any clinically useful sense to be meaningful. Consequently, use of the sponsor's numbers now in the proposed labeling for these endpoints will almost certainly be used in a misleading fashion and should therefore not be allowed.

/s/

David Hoberman, Ph.D.
Mathematical Statistician

Concur: Dr. Welch

Dr. Nevius

cc:

Arch NDA# 21-368

HFD-580

HFD-580/ABatra, MHirsch, DSpell-LeSane

HFD-715/DHoberman, MWelch, ENevius, CAnello

11 page(s) have been removed because it contains trade secret and/or confidential information that is not disclosable.

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

/s/

S. Edward Nevius

4/22/02 12:09:48 PM

BIOMETRICS

Concur with review. Submitted for David Hoberman due to
DFS difficulty.

Mike Welch

4/22/02 12:55:37 PM

BIOMETRICS

Concur with review

Statistical Review and Evaluation

NDA: 21368

Date: February 4, 2002

Applicant: Lilly ICOS LLC.

Name of Drug: Cialis (Tadalafil) Tablets.

Document Reviewed: Document dated 6/28/2001.

Statistical Reviewer: Wen-Jen Chen, Ph.D. (HFD-715).

I. Introduction

8 page(s) have been removed because it contains trade secret and/or confidential information that is not disclosable.

(b4)

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

/s/

Wen-Jen Chen
2/4/02 04:16:48 PM
BIOMETRICS

Karl Lin
2/4/02 04:46:54 PM
BIOMETRICS
Concur with review