Iowa City, Iowa 52242

The University of Iowa Hospitals and Clinics
Department of Psychiatry
500 Newton Road

To Call Writer Direct
Phone (319) 353- 3719

Signed copies

JEW AUG 19 1985

August 13, 1985

J. W. Wernicke, Ph.D., M.D.
Clinical Investigator
Lilly Research Laboratories
307 E. McCarty Street
Indianapolis, IN 46285

Dear Dr. Wernicke:

We have reviewed the letter which Bruce Dornseif sent to Skip Woolson.

We are in agreement between the three of us that the analyses described in the first two paragraphs of Dr. Dornseif's letter ought to be the ones reported as a summary of our consultation activities with Lilly on this project. We feel these comparisons take into account the total person-weeks of exposure in a proper manner and lead to a direct, unbiased test of the hypothesis in question.

Again, we want to thank you for the opportunity to be involved in this project. If there is something else which is necessary or desirable on your part, please do get in touch with us.

Sincerely,

George Woodworth, M.D.
The Paul W. Cunningham Professor
and Head, Department of Psychiatry

Robert F. Woolson, Ph.D.
Professor & Head
Division of Biostatistics
Department of Preventive Medicine

William Coryell, M.D.
Associate Professor of Psychiatry

JFW JUL 24 1985

Lilly Research Laboratories
A Division of Eli Lilly and Company

307 East McCarty Street
Indianapolis, Indiana 46285
(317) 261-2000

July 22, 1985

Robert Woolson, Ph.D.
Professor of Biostatistics
Department of Preventive Medicine
and Head, Division of Biostatistics
The University of Iowa
Iowa City, Iowa 52242

Dear Dr. Woolson:

The following tables summarize our current statistical analyses to determine if there was increased suicide attempts associated with fluoxetine, in comparison to the comparator group (imipramine, desipramine, amitriptyline, placebo), for the pooled double-blind and open-label studies. All computations are based on our current knowledge of the number of weeks at risk for the various groups, and on the revised[1] number of suicide attempts for the combined studies (6 in fluoxetine and 1 in the comparator groups). These values are reflected in the tables. Following Dr. Woolson's suggestion of using a binomial distribution, 2 p-values are initially calculated. The first considers fluoxetine and comparator exposure to week 54, while the second considers comparator exposure to week 54 and fluoxetine exposure to week 174. From our computations, there appear to be no significant differences between fluoxetine and the comparator group ($p=0.165$ for exposure to week 54, $p=0.208$ for exposure to week 174) for suicide attempts, based on patient exposure.

If we perform a binomial analysis considering fluoxetine and comparator exposures only to week 5, thereby producing a "balanced" exposure comparison between fluoxetine and all 4 comparators, we reduce the total number of suicide attempts to 5 (4 in fluoxetine, 1 in the comparator group). The p-value obtained for this computation is equal to 0.190. Similarly, comparing fluoxetine exposure to week 54 with only imipramine and doxapin in the comparator group, thereby producing a "balanced" exposure comparison and reducing the number of suicides to 6 (6 in fluoxetine and 0 in the comparator group), we obtain a p-value from the binomial expansion equal to 0.079.

[1] As per Dr. Winokur, Department of Psychiatry, The University of Iowa.

based on revised figures

Fisher's Exact Test was also suggested as an alternative methodology to assess differences in suicide attempts between fluoxetine and the comparator group. Although we question its applicability for this situation in which patient weeks are considered as experimental units (lack of independence between units, mixing of suicidal which are in patient units with units of patient exposure), we are providing the results from our analysis:

Exposure Units	Duration	P-value (1-tail test)
Patient weeks	Week 34	0.16[illegible]
Patient weeks	Week 175	0.[illegible]

The above results agree almost exactly with those obtained from the binomial expansion for exposure to weeks 34 and 175, using all 4 comparators and 7 suicide attempts (6 in fluoxetine and 1 in the comparator group).

Using Fisher's Exact test (1-tail) and carrying patients as experimental units, and comparing fluoxetine to all comparators until week 5, we have 5 suicide attempts (4 in fluoxetine, 1 in the comparator group). Fisher's Exact Test for these data provide a p-value equal to 0.166. Similarly, a comparison of fluoxetine to only imipramine and doxepin in the comparator group, to week 34, and using patients as experimental units and a total of 5 suicide attempts (4 in fluoxetine and 0 in the comparator group), provides a p-value equal to 0.058 by Fisher's Exact Test.

An alternative approach, which also utilizes exposure time, is to compare the survival curves for the fluoxetine group with the comparator group for the pooled double-blind and open-label studies. The log-rank test (2-tails) provides a chi-square value of 2.60 (df=1) with a corresponding p-value equal to 0.107. Any conclusions drawn from this analysis must be considered in light of the fact that the number of events (suicide attempts) is extremely small for each group. This forces me to question the suitability of any survival analysis for these data.

In summary, we are analyzing a pooled group of Phase II & III clinical trials, as though it is a traditional "observational" study for which such variables as suicide attempts were not part of the a priori considerations but were "observed" as part of the adverse events recording mechanism. Hence a wide range of p-values are possible, depending on the analysis. Since we have analyzed a variety of ad hoc groupings we came up with a widely varying set of p-values. This should not be surprising.

Sincerely,

LILLY RESEARCH LABORATORIES
A Division of Eli Lilly and Company

Bruce E. Dornseif, Ph.D.
Senior Statistician

BED

Attachments

cc: W. Beyrault, M.D.
O. Wilbanks, M.D.

bcc: Dr. K. K. Houck
Dr. W. C. Lister
Dr. W. W. Offen
Dr. C. B. Sampson
Dr. J. F. Warnicke

The evidence in the form of the p-values are summarized in the following table:

Test	Experimental Units	Exposure Duration	Total Suicide Attempts	p-value
B	Patient weeks	Week 54	7	0.165
B	Patient weeks	Week 174	7	0.208
FE	Patient weeks	Week 54	7	0.165
FE	Patients weeks	Week 174	7	0.208
B	Patient weeks	Week 5	5	0.190
B*	Patient weeks	Week 54	6	0.079
FE	Patients	Week 5	5	0.166
FE*	Patients	Week 54	6	0.058

*Comparator group consists only of imipramine and doxepin.

B=Binomial
FE=Fisher's Exact

Comparator Exposure Rates

Week	Imipramine Pt Weeks	Doxepin Pt Weeks	Amitriptyline Pt Weeks	Fluoxetine Pt Weeks
1	326	134	70	[illegible]
2	286	118	61	260
3	247	[illegible]	56	196
4	213	[illegible]	53	165
5	189	81	[illegible]	137
6	156	61		96
10	412	176		
14	336	156		
22	328	132		
30	472	208		
38	492	176		
46	318	128		
56	240	104		
Total	4109	1587	289	1137

Exposure Rates

	Flovoxetine			Comparator		
Week	Weeks at Risk	Patients at Risk	Patients Weeks	Weeks at Risk	Patients at Risk	Patients Weeks
1	1	758	758	1	813	813
2	1	699	699	1	720	720
3	1	627	627	1	598	598
4	1	553	553	1	520	520
5	1	498	498	1	456	456
6	1	422	422	1	323	323
10	4	321	1284	4	147	588
14	4	275	1100	4	123	492
22	8	195	1560	8	100	800
30	8	152	1216	8	85	680
38	8	124	992	8	71	568
46	8	103	824	8	55	440
54	8	88	704	Zoloft 8	43	344
62	8	30	240	14238 8	43	344
70	8	33	264		Total	7342
78	8	24	192			
86	8	22	176			
94	8	17	136			
102	8	14	112			
110	8	12	96			
118	8	10	80			
126	8	7	56			
134	8	7	56			
142	8	5	40			
150	8	4	32			
158	8	3	24			
166	8	2	16			
174	8	1	8			

Total 12726

It became apparent to me that one of the things which was necessary was a systematic evaluation of the patients who were said to have attempted suicide in order to determine whether we would agree that these were suicide attempts. There was, of course, one completed suicide in the group.

As a consequence, I asked you for the charts on the patients and after you sent them to me I went over them. Out of the total group of cases I identified eight cases in which there might be considerable controversy as to whether these were indeed suicide attempts at all. These are:

1) [redacted] who was a polydrug abuser. He took cocaine, heroin, fluoxetine, and alcohol at one time and was eliminated from the study. He denied this was a suicide attempt. He had a history of having abused drugs in the past. I did not believe that should be considered a suicide attempt. Dr. Coryell was presented this data in Indianapolis, and he agreed.

2) [redacted] This person used cocaine, was never very suicidal and engaged in self-mutilation with broken glass. There was no evidence that this was a suicide attempt. The data were not too good, but both Dr. Coryell and I agreed on going over the material in Indianapolis that this was not a suicide attempt.

3) [redacted] At the start of the study, this person had no suicide ideation according to the Hamilton Rating Scale. The patient said that the overdose was taken because "I needed help", and the psychiatrist who evaluated the patient said, "This denies a specific suicide attempt and she

currently has no suicide ideation." I rated this person as not making a suicide attempt, but Dr. Coryell looking at another evaluation was enough in doubt that he thought it might be considered a suicide attempt.

4) [redacted]. This patient took an overdose of chloralhydrate, 8 capsules four weeks after stopping medication. At the start of the study, this person had considerable suicide ideation on the Hamilton Rating score but at visit 10 suicide ideation was rated a 0. Visit 11 did not occur but was scheduled for April 12, 1993. The drug was stopped by the patient herself on March 3 as she felt well. She stopped herself on the drug rather than being stopped by a physician. Thus, she was not on any drug for four weeks before this suicide attempt. Both Dr. Coryell and I agreed this could not be related to any treatment.

5) [redacted]. The patient had considerable suicide ideation prior to the onset of the study. The suicide attempt was by wrist slashing. The patient stopped medication March 13, and the suicide attempt was March 15. As the half-life of the drug is three-four days, we agreed that this in fact could be related to the drug.

6) [redacted]. This person had no suicide ideation at onset of study. The patient drank a bottle of rum and then took 10 fluoxetine capsules in divided doses of 2-hour intervals. Dr. Coryell and I agreed this was not a suicide attempt.

7) [redacted]. This person had mild suicidal ideation at the beginning of the study. There was some question as to whether this was a suicide attempt.

in my mind; but when Dr. Coryell and I evaluated it together in Indianapolis, we agreed on a positive.

8) [redacted] The patient had suicidal ideation at the beginning of the study and made a self-inflicted laceration of the skin with a razor blade. Dr. Coryell and I agreed that this was not a suicide attempt.

Thus, of the eight cases there was agreement that there was no suicide attempt in five, a positive suicide attempt in two and a divided vote in one [redacted]

To be evaluated as a suicide attempt, the attending doctor on seeing the patient had to comment that the behavior in fact was related to suicide intention. In other words, if a patient said "I cut my wrist because I was mad at my boyfriend and wanted to gain his attention" we did not consider it a suicide attempt. If, on the other hand, the patient said "I intended to take my life" or the doctor felt that the person had any intent at all to take his/her life, it was counted as a positive. Thus, at least five cases did not meet these criteria.

Prior to coming to Indianapolis, Dr. Coryell, Dr. Woolson and I met for an hour. We determined that there was a series of things which could be of some interest and might clarify the issue. First, it was possible that the patients who were on one or the other treatments were more suicidal than patients in other groups. We had no information on this and needed to get it in Indianapolis. Second, if the physicians could somehow break the blind because of the side effects and had more confidence that the fluoxetine patients were

taking a drug which would not be lethal, they might keep suicidal patients on fluoxetine longer than on comparator drugs. Was it possible that in the course of the studies the more suicide inclined patients were terminated from comparator drugs earlier than those from fluoxetine; thus accounting for an eventually higher rate of attempts in the fluoxetine group? This is still something which needs to be looked at, and the way to do this is determine if there is a differential change over time in baseline suicide ratings between fluoxetine and comparator groups. Another way to look at this is to look at the five point Hamilton Scale and determine the dropouts over time for each point, associating the drop outs with particular treatments. Finally we questioned whether clinicians, in fact, could guess which drugs the patients were on. We do not know the answer to this. Perhaps there was some evaluation.

Upon arrival in Indianapolis, we looked at the baseline suicide scores between the various groups. There were no differences. Thus, there was no reason to believe that there was misalignment. One of the most interesting things, however, is the fact that in all groups there was an enormous amount of suicide ideation. Thus, if one looks at the Hamilton Rating Scale and the 3 score only between 3% and 39% of the groups were rated as having no suicide ideation of any kind. This, of course, is the one item on the Hamilton Rating. What this means is that the patients had considerable suicide inclinations according to that measure. It is interesting that there is such a low suicide attempt rate under those circumstances. Of course, the physicians were supposed to evaluate the suicide potential as well. They were supposed to include patients only if they did not believe they were suicide risks. Nevertheless, in a systematic evaluation, there were suicide ideas and

thoughts. In evaluating over the various groups, there were 3% to 18% of the patients who had a rating of 3 on the Hamilton Rating Scale. This indicates considerable suicide ideation. So, by my criterion some of these patients were suicidal when they entered the study and I continue to be surprised at the low number of suicide attempts in the study.

In evaluating the end points, it was clear that there were decreases in suicide ideation on the Hamilton Rating Scale in the fluoxetine groups and also in the groups which were treated with comparator drugs. On the other hand, such a decrease did not occur in the placebo groups. There was no significant difference between the fluoxetine groups and the imipramine group as an example in terms of decreases of suicide ideation. The placebo group, however, did not show this salutatory effect.

The conclusion of this set of findings is that the cognitive aspects of suicide are not increased in the fluoxetine group as opposed to the comparator drug groups. This is a very important point.

Next, there is the question of whether there are increased suicide attempts in the fluoxetine group as opposed to the comparator groups. We decided that the appropriate way to look at this was to look at the amount of weeks at risk for the various groups. The reason for this is that the fluoxetine groups were associated with more weeks at risk than the other groups and this had to be taken into account as the more weeks at risk would give the fluoxetine group a better chance of having a suicide attempt. Dr. Coryell made an effort to do this in Iowa City, and we made another effort to do this in Indianapolis. However, there were such divergent viewpoints as to how many weeks at risk

should be attributed to each group that we have to go over this again in a more systematic fashion. There were marked discrepancies with each new evaluation of weeks at risk and this turns out to be something which must be clarified. One statistical workup which was suggested was Fisher's Exact Probability between the fluoxetine group and the comparative drugs plus placebo. Dr. Woolson suggested using a binomal expansion. Both of these should be done but the fact is that both are dependant on knowing an exact number of weeks at risk for the various groups. The original evaluations, both in Iowa City and in Indianapolis, were questionable because of the possibility that we were counting some weeks at risk twice. In any event, the significance favoring the fact that fluoxetine was associated with more attempts varied with P numbers of .031, .056, and .18. Thus, there is considerable variability and we must assume that we do not have a final answer to this question. It is of major importance that this be clarified.

There are some qualitative points that should be noted. 1) In any group the presence of suicide attempts is a very rare event, probably less common than you would see in other studies. 2) If it turned out that the fluoxetine group in fact had significantly more suicide attempts per weeks at risk, we would have an interesting finding. It would indicate that the cognitive suicide data does not separate the groups, but the behavioral suicide data does. The meaning of that is something which should be pursued vigorously. A possibility which comes to mind is that fluoxetine might be somewhat more stimulating as a drug and that individuals might be slightly more impulsive although their thinking was not changed. In any event, this is only meaningful if we have a significance difference between groups in terms of suicide attempts per unit of observation.