Table 71. Grade 3-4 adverse events, NSABP P-1 trial

Adverse Event	Placebo	Tamoxifen	Total
WBC	3	1	4
Platelets	2	6	8
SGOT	1	0	1
SGPT	0	1	1
Bilirubin	0	3	3
Creatinine	9	7	16
Alkaline phosphatase	9	2	11
Cardiac dysrhythmia	15	10	25
Cardiac function	7	4	11
Hypertension	47	52	99
Hypotension	1	3	4
Neuro-sensory	6	5	11
Neuro-motor	8	9	17
Neuro-cortical	3	2	5
Neuro-cerebellar	1	2	3
Neuro-headache	19	29	48
Neuro-constipation	2	2	4
Neuro-hearing	1	3	4
Neuro-vision	9	17	26
Hemorrhage	1	1	2
Infection/sepsis	8	5	13
Nausea	8	6	14
Vomiting	6	3	9
Diarrhea	13	8	21
Stomatitis	0	2	2
Hematuria	2	1	3
Alopecia	3	4	7
Pulmonary	12	7	19
Skin	5	4	9
Allergy	9	9	18
Fever	2	1	3

Reviewer Comments:

1. The original submitted database did not allow an assessment of the number of women with elevated liver enzymes or hematologic abnormalities. Grade 3-4 ADR events and primary laboratory data were submitted 7/29, 7/31, and 8/3/98.

2. Hematologic parameters

There were no significant differences between WBC counts between the two groups. Few women had grade 3-4 platelet abnormalities. In the ERSMAC report, the number of women with platelet abnormalities of any grade was reported as 18 on placebo compared to 43 on tamoxifen. It is possible that tamoxifen is associated with thrombocytopenia; however, these data suggest that if there is an association, thrombocytopenia is a rare event.

No information on hemoglobin or hematocrit levels was collected.

2. Liver Function Test Abnormalities

In the database, as reported in Table 60, 1 participant on each arm had an elevation of either SGOT or SGPT. However, the ERSMAC report states that 9 participants on placebo and 7 on tamoxifen had grade 3-4 elevations of SGOT, and that 9 and 2 respectively had grade 3-4 elevations of SGPT. If an Access query is performed on the original toxicity database (BCPT2), 14 participants on placebo and 8 on tamoxifen experienced a grade 3-4 elevation of either SGOT or SGPT. Overall, there were more transaminase elevations on placebo than on tamoxifen.

Three patients on tamoxifen had grade 3-4 bilirubin elevations compared to none on placebo.

There were no differences in the incidence of rises in alkaline phosphatase; if anything, there were more elevations of any grade on placebo (97 v. 46).

3. The incidence of other adverse events was not significantly different between the two treatment arms. However, review of the case report forms indicated that events may have been missed. The database contains only information reported at scheduled visits. Adverse events are more likely to occur at timepoints that do not correspond to scheduled appointments.

10.7.2 Gynecologic Symptoms

There are no CTC grades for hot flashes or vaginal discharge, 2 well-recognized symptoms associated with tamoxifen. The following categories were therefore used:

Level 0:	No symptoms or symptoms were not bothersome
Level 1:	Slightly bothersome
Level 2:	Moderately bothersome
Level 3:	Bothered quite a bit
Level 4:	Extremely bothersome

Hot flashes and vaginal discharge were more common and more severe on the tamoxifen arm compared to control. These symptoms are shown in the following table:

Table 72. Gynecologic symptoms among NSABP P-1 participants (ERSMAC report, Table 1, page 32, volume 109.3)

Self-reported symptom/level	Placebo (n=6469)*		Tamoxifen (n=6441)*	
	No. Pts.	%	No. Pts	%
Hot flashes:				
Level 0	2053	31.7	1269	19.7
Level 1	1184	18.3	909	14.1
Level 2	1398	21.6	1352	21.0
Level 3	1189	18.4	1794	27.9
Level 4	645	10.0	1117	17.3
Vaginal discharge:				
Level 0	4230	65.4	2908	45.1
Level 1	1408	21.8	1686	26.2
Level 2	544	8.4	1058	16.4
Level 3	212	3.3	591	9.2
Level 4	75	1.2	198	3.1
Vaginal bleeding				
Level 0	5057	78.2	4974	77.2
Level 1	682	10.5	741	11.5
Level 2	399	6.2	387	6.0
Level 3	208	3.2	215	3.3
Level 4	123	1.9	124	1.9
Vaginal dryness				
Level 0	3140	48.5	3094	48.0
Level 1	1140	17.6	1135	17.6
Level 2	993	15.4	987	15.3
Level 3	694	10.7	769	11.9
Level 4	502	7.8	456	7.1

* Number of participants with follow-up QOL forms

Overall, hot flashes of any severity were reported in 68% of the placebo patients and in 80% of the tamoxifen patients. Level 3-4 hot flashes were reported in 28% of women on placebo and in 45% of women on tamoxifen. For vaginal discharge, the report of any severity occurred in 35% of women on placebo and in 55% of women on

tamoxifen. Level 3-4 events occurred in 4.5% and 12.3% of participants respectively. There was no difference in the incidence of vaginal dryness or vaginal bleeding between the two treatment arms.

11.0 Quality of Life analyses

The NSABP provided copies of the QOL analysis to the FDA on July 31, 1998. The introductory statement indicated that this analysis and slides had been prepared for a presentation at the Annual Meeting of the Society for Clinical Trials May 17-20, 1998.

In the analysis, data from 11,064 women randomized in the first 24 months of the study were used. The data was obtained at baseline and at follow-up at 3, 6, 12, 18, 24, and 36 months. This group of women was equally distributed by age group between tamoxifen and placebo, and the age distributions were similar to those of all 13, 388 participants in the BCPT trial. Three components of the Quality of Life questionnaire were analyzed: the Center for Epidemiological Studies--Depression Scale (CES-D), the Medical Outcomes Study (MOS), and the Sexual Activity Item from the MOS Sexual Functioning Scale.

The CES-D measures "non-specific psychological distress"--the items are related to affective distress but not to a particular psychiatric disorder. The data was presented as the mean score of all analyzed participants by treatment arm at each timepoint, and also stratified by the age groups 35-49, 50-59, and age 60 or older. A second series of tables showed the proportion of participants at each timepoint with a score $\geq$ 16, as 16 is considered the upper limit of normal, for the entire group and by age. All of these curves are superimposable and show no difference between treatment arms or differences between age groups. Each age group had approximately the same mean score, and no differences from baseline to the 36-month timepoint were observed.

The MOS analysis used the mean Physical Summary Scale and the mean Mental Health Summary Scale. These scales permit comparison with the general population of the United States as well as between-scale comparisons. The mean in the U.S. general population is defined as 50. The scales show no difference between treatment arms for the entire group or the subsets by age; there does not appear to be any difference in mean scores between the age groups. The scales appear to fall along a mean of 50, consistent with the mean in the general population. There was no difference between baseline and 36 month scores.

The Sexual Activity Item reports on the proportion of BCPT participants who were sexually active during the 6 months prior to each evaluation. Approximately 60-65% of the entire BCPT population was sexually active during the trial. The proportion of participants who were sexually active varied according to age:80% of women aged 35-49 were sexually active, compared to 65-70% of women aged 50-59 and 40% of women over the age of 60. There was no difference between treatment arms overall or by age; the proportions did not vary significantly over the 36-month time period that was studied.

Reviewer Comment:

1. There was a two-month delay in submitting the analyses, despite our agreement to accept the final analysis only rather than the raw data.

2. As described in the next comments, the CES-D scores were submitted. The statistical reviewer, Tony Koutsoukos, Ph.D., evaluated all randomized participants for extent of missing data. There is a gradual decline in the amount of missing data until approximately 36-42 months; after this timepoint, there is an acceleration in the amount of missing data, consistent with the unblinding of the trial. For the subset of women used in the QOL analyses, the pattern and amount of missing data appears to be the same between the two treatment arms.

2. Because of reports in the literature that associated tamoxifen with depression and because of the higher grades of depression reported during the trial, the Division chose to examine this aspect of the Quality of Life assessment in greater detail. At the request of the FDA medical reviewer, the QOL form with the CES-D and the depression scores for each participant were submitted.

The QOL questionnaire consisted of an 8-page form; the second page of the questionnaire contained the CES-D scale, which is comprised of 20 statements. The statements were a list of feelings, attitudes, and behaviors ("I was bothered by things that usually don't bother me"). Participants were asked to describe how often they had experienced these feelings in the past week: rarely or none of the time (less than one day), some of the time (1-2 days), moderately (3-4 days), or most of the time (5-7 days). These responses were scored from 0 to 3. A score of 15 or less is considered normal.

3. We first looked at the prior history of depression, nervous or emotional disorder, or psychiatric problems in the study population, and past or current use of antidepressants or tranquilizers.

Table 73. Past history of depressive/psychiatric illness and past/current use of antidepressant medication

Condition	Placebo (n=6707)	Tamoxifen (n=6681)	Total (n=13,388)
PHx depression, nervous/emotional disorder, psychiatric problem			
No	5535	5471	11,006
Yes	1172	1210	2382
Currently on antidepressants or tranquilizers:			
No	6020	5974	11,994
Yes	687	707	1394
Previously on antidepressants or tranquilizers			
No	4829	4792	9621
Yes	1878	1889	3767
Any past or present use of these drugs			
No	4794	4755	9549
Yes	1913	1926	3839

Only about 64% of the women on each arm with a past history of antidepressant or tranquilizer use stated that they had a prior history of depression, a nervous/emotional disorder, or psychiatric problems. Antidepressants are used for a variety of non-psychiatric conditions, but the medical reviewers noted, in reviewing the requested CRFs, that participants often added a note identifying a psychological reason for the use of the medication. Some participants may not have viewed short-term use of these medications in the past as consistent with a psychiatric diagnosis.

4. Review of the case report forms indicates that the baseline depression scores were assessed prior to randomization and were forwarded to the NSABP Biostatistical center. On the results of the entry/eligibility review, a "special note" was included if a woman's score exceeded the normal cut-off. However, there is no indication that P.I.s were required to discuss this finding with participants or perform further evaluation/treatment.

5. Information about depression was gathered in two ways: by a "Neuro-mood" toxicity grading obtained through the patient's self-reported symptoms and discussion with the study coordinator, and by the depression scores, calculated from the CES-D. The following table reports the distribution of neuro-mood toxicity in the trial:

Table 74. (Table 2, ERSMAC report, volume 3, page 12). Distribution of participants by depression grade.

	PLACEBO (N=6484)		TAMOXIFEN (N=6492)	
Depression Grade	No. pts	%	No. pts	%
None (0)	5785	89.2	5740	88.4
Mild (1)	317	4.9	313	4.8
Moderate (2)	328	5.1	375	5.8
Severe (3)	33	0.5	36	0.6
Suicidal (4)	20	0.3	26	0.4
Death (5)	1	<0.1	2	< 0.1

These figures were verified by the reviewer in an MS Access query of the database.

The tamoxifen arm had a greater number of participants with grade 3-5 depression: 64 compared to 54. In 2 patients treated with tamoxifen and 1 patient treated with placebo, the depression led to suicide. Overall, the NSABP felt there was a slight shift towards higher grades of depressive toxicity associated with tamoxifen, as stated in the ERSMAC report.

6. Based on this concern, the NSABP evaluated the depression scores for the women in the study:

Table 75. Distribution of highest depression score reported by NSABP P-1 participants (ERSMAC report, Table 3, page 13, volume 109.3)

	PLACEBO (N=6469)		TAMOXIFEN (N=6441)	
Depression Scores	No. pts	%	No. pts	%
0-15	4261	65.9	4242	65.9
16-22	1032	16.0	998	15.5
23-29	610	9.4	642	10.0
30-36	334	5.2	325	5.0
37+	232	3.6	234	3.6

Because there was no difference in the range of scores reported, the NSABP concluded there was no difference in the incidence of depression between the two arms.

7. We examined the range of depression scores in participants who had reported grade 3 through grade 5 neuro-mood toxicity.

Table 76. Range of depression scores for participants with grade 3, 4, or 5 neuro-mood toxicity

Grade, Neuro-mood toxicity	Placebo	Tamoxifen	Total
Grade 3:	33	36	69
Depression scores			
0-15	9	9	18
16-22	6	2	8
23-29	5	9	14
30-36	6	5	11
≥ 37	6	11	17
Unknown	1	0	1
Grade 4:	20	26	46
0-15	1	3	4
16-22	5	3	8
23-29	2	3	5
30-36	4	1	5
≥ 37	8	15	23
Unknown	0	1	1
Grade 5:	1	2	3
0-15	0	0	0
16-22	0	1	1
23-29	0	0	0
30-36	1	1	2
≥ 37	0	0	0

These data show that even within severe grades of reported neuro-mood toxicity, a wide range of depression scores was reported. The 2 measures did not always correlate.

8. In order to assess why these measures did not correlate, we examined data from participants identified in the requested CRFs. When the reviewers looked at the CRFs for other endpoints, we noted whenever a participant reported "feeling depressed" and also collected information on the initiation of antidepressant drugs. We did not include antidepressant drugs started for non-psychiatric reasons. [For example, some participants used these medications for fibromyalgia.] We then looked at the depression scores recorded around the time of the event.

Placebo:

Randomized 10/8/93; stopped drug therapy 4/18/94
Began Paxil 1/97
No scores obtained after 10/94

Randomized 7/23/92; stopped drug 7/25/94
Reported depression from the 3/31/94-7/25/94 follow-up
Depression score 7/24/94 14
Randomized 10/21/92; stopped drug therapy 11/7/94
12/9/96: "profound depression requiring medical therapy"
No scores obtained after 10/19/94
Randomized 10/26/92; off study 4/1/96
5/3/95: begun on antidepressant medication
Depression scores: 4/24/95 16
10/25/95 5
Randomized 10/15/92; off study approximately 10/24/97
Began Prozac while on study
Baseline depression score = 6
All others taken on study range from 25 to 43
Began study drug 3/15/94; stopped drug 8/12/94
6/9/94 Recorded with depression
Depression scores: 1/28/94 11
6/9/94 15
On study drug 9/23/92; remained on drug as of 9/17/97
Antidepressant medication started 3/95
Depression scores: 9/21/94 37
3/28/95 37
9/19/95 22
On study 1/26/94; remained on study drug as of 2/3/98
Zoloft prescribed 1997
Trazodone prescribed 8/87
Depression scores: 1/31/97 52
7/31/97 45

Tamoxifen:

Randomized 9/15/92; stopped drug therapy 12/26/92
Began Prozac 9/94
Last depression score obtained 5/16/94 = 7
Randomized 3/17/93; stopped drug 11/1/94
Begun on antidepressant drug therapy 4/94
Depression scores: 10/6/93 0
6/1/94 11
Randomized 1/4/93; drug stopped 7/7/96
11/30/94 Amitryptilene begun
Depression score: 7/1/94 23
1/4/95 11
Randomized 6/10/93; drug stopped 10/26/95
6/15/95 Depression requiring medical therapy
Depression scores: 12/12/94 16
6/15/95 5

Randomized 9/2/92; off study approximately 7/96
3/93 Reported as "severe depression"

Depression scores:	11/30/92	9
	3/3/93	7
	9/8/93	3

Randomized 12/3/92; off study 7/29/93
Reported mild depression in the follow-up period 6/13/93 to 12/29/93

Depression scores:	3/10/93	5
	6/11/93	5
	3/18/94	2

Began study drug 5/19/93; stopped drug 19/9/96
3/6/96 Elavil prescribed for depression

Depression scores:	6/28/95	8
	3/5/96	12

Began study drug 8/24/92; off drug 2/7/95
Described with depression in the follow-up period 8/26/93-2/28/94

Depression scores:	2/25/93	3
	8/24/93	8
	2/27/94	14

Began study drug 2/10/94; off study 3/31/97
Began Zoloft 7/11/94

Depression scores:	5/9/94	7
	9/9/94	1

Began study drug 8/9/92; stopped drug 1/10/97
Prozac started in 1995

Depression scores:	8/2/94	18
	2/15/95	29
	1/31/96	22

Began study drug 9/3/92; off therapy 7/2/97
Prozac begun 1994
Zoloft begun 1996

Depression scores:	3/2/94	12
	9/16/94	17
	3/6/95	15
	10/6/95	18
	3/14/96	11
	12/2/96	15

The women who committed suicide on study are described separately below:

Placebo:

Randomized 12/10/93; stopped study drug 1/14/95
Died from suicide 11/4/95

Depression scores:	11/23/93	3
	4/12/94	1
	7/3/94	41
	1/5/95	42
	7/14/95	18

Tamoxifen:

Randomized 11/18/92; off study 11/24/97 because of death from suicide
Suicide attempt 5/4/94

Depression scores:	11/3/93	6
	5/17/94	48
	11/17/94	1

Suicide attempt 9/5/97

Depression score	5/19/97	3

Suicide 11/24/97

Depression score	11/10/97	49

Began study drug 9/4/92; off study 4/8/96 due to suicide
1/7/94 Prozac prescribed

Depression scores:	9/1/93	5
	3/9/94	14

4/8/96 Suicide

Depression score	2/26/96	5

No additional values

The data from women who did not commit suicide show that out the 19 women discussed, 11 had normal depression scores or had not had depression scores obtained around the time of the event. Of the 8 remaining women, 3 had grade 1 scores, 2 had grade 2, and 3 had grade 4 scores. In the women who committed suicide, scores were high around the time of a reported suicide attempt, but were normal at other points.

An additional 4 participants were identified, all randomized to tamoxifen, who did not have depression scores reported in the database table. Three decided to go off study because of depression ; one went off study and reported depression requiring medication within 1 month of stopping study drug.

This information suggests several points:

- Depression scores are likely to be accurate only if taken at the time of the participant's acute distress
- A number of the participants began drug therapy when scores were grade 0-2. One possibility is that depression scores were not measured at the time of greatest distress; another is that clinical decisions are made on the basis of symptoms that score as

grade 1-2. Perhaps the usual reporting system of grade 3-4 adverse events underestimates clinically significant changes in mood.

- The analyses performed in this trial do not indicate that tamoxifen causes depression. However, given the limitations of the testing intervals and the correlation of the scores with clinical events, it is not possible to exclude this possibility conclusively.

12.0 NSABP P-1 Substudies

The NSABP sponsored several substudies in conjunction with P-1, which are listed below.

1. NSABP P-1B: The bone mineral density and biochemical marker study to determine the effect of tamoxifen on bone in premenopausal and postmenopausal women

The goal of this trial was to evaluate the effect of tamoxifen on osteoporosis. Bone mineral density measurements, in addition to blood/urine collections to study serum markers of bone turnover, were to be performed at years 0, 1, 2, and 5. The study opened in March 1995 at 20 sites. Eighteen sites accrued participants to this substudy. A total of 107 participants entered this study; however, 5 sites contributed 67% of the participants. Because of the closure of NSABP P-1 with subsequent unblinding, year 5 data cannot be obtained. Data will be collected through either the 1 or 2-year timepoint, depending on the participant's time of entry, and the data will subsequently be analyzed and published.

2. NSABP P-1U: A protocol to evaluate the effect of tamoxifen therapy on the endometrium in participants enrolled in the BCPT

This trial was never finalized, due to the temporary closure of all NSABP trials in 1994. When NSABP P-1 re-opened, endometrial sampling was added to the original protocol as an amendment, superceding the need for this substudy.

3. NSABP P-1E: A protocol to evaluate the prevalence and detection of ophthalmic abnormalities associated with long-term low dose tamoxifen administration

This trial, although submitted as a substudy under the P-1 IND, enrolled only patients from NSABP B-14. This study resulted in a publication, which is discussed with the eye findings from the current trial.

4. NSABP P-1G: A study of the association between inherited mutations and the effect of tamoxifen on breast cancer incidence

This study will examine:

- The effect of tamoxifen compared to placebo on the incidence of invasive breast cancer in women with inherited mutations of BRCA1 or BRCA2
- Whether the effect of tamoxifen versus placebo differs for women with inherited BRCA1 or BRCA2 mutations compared to women without mutations
- The proportion of participants with inherited mutations
- The risk of developing breast cancer among mutation carriers compared to non-carriers in this trial

Blood samples were collected at baseline *(in 96% of participants)* and at the time of a protocol-defined event for research purposes as part of the BCPT study. These samples will be used in this substudy, which will be initiated in the near future by Mary-Claire King, Ph.D.

In addition to the NSABP P-1 substudies, local institutions could submit proposals for single-site trials to be performed in conjunction with the parent study. Four single agent trials were conducted:

1. The effect of tamoxifen on the hemostasis system in women without breast cancer: Implications for cardiovascular disease prevention and assessment of thrombotic risk, David Krag, M.D., University of Vermont.

The objectives of this study were:

- To determine whether tamoxifen alters the hemostasis factors which are thought to be important as risk factors for coronary heart disease
- To determine if tamoxifen has a negative effect on measures of inflammation, since these are also implicated as risk factors for cardiovascular disease
- To determine if tamoxifen alters hemostatic factors in a manner that would predict increased risk of venous thrombosis

One hundred thirty-seven women were accrued to the study. Blood samples were collected and have been analyzed; correlation of the results with randomized therapy has not been completed.

2. Participant adherence for the NSABP breast cancer prevention trial, Richard Day, M.D., University of Pittsburgh

Four BCPT centers collaborated on this study: Rush-Presbyterian, UCLA, Fox Chase Cancer Center, and Georgetown University Medical Center. One hundred women were monitored using an electronic Medication Event Monitoring system (MEMS). A microchip was inserted in a bottle cap, which recorded the number, date, and time of pill bottle openings. The women were evaluated at the 3, 6, and 12 month time periods. These results were compared to the pill counts and the staff estimates which were performed as part of the NSABP P-1 protocol using a kappa statistic. There was poor correlation between MEMS and the other methods. According to the abstract that was submitted, MEMS indicated a higher degree of compliance than the other methods. However, this abstract does not provide detailed information about the results, and no definitive conclusions can be drawn.

3. Breast Cancer Prevention Trial: A multidisciplinary support and education group, George Peters, M.D., Baylor-Sammons Cancer Center, Dallas, TX

This trial was designed to test the effect of 3 group interventions on compliance with the NSABP P-1 trial. One group used educational approaches, one emphasized emotional and spiritual support, and the third combined educational tools and spiritual

support. Twenty women enrolled in the study; 9 agreed to attend the support group (support group assignment not given). However, the average attendance at the sessions was between 5 and 7 participants. After 1 year, the study was closed because of lack of interest on the part of the participants.

4. The BCPT: Nurses' observations, Tovia Freedman, University of Pennsylvania.

In this study, a letter of invitation was sent to all NSABP P-1 principal investigators, who were asked to forward the letter to the local nurse-coordinator of the trial. Fifty nurses agreed to participate in the study. Thirty participated in a telephone interview and 20 participated in 4 focus groups (one in the United States and 3 in Canada) given as part of two oncology nursing conferences. A publication resulted from this effort and reported the opinions of the nurses expressed during these interviews.

Reviewer's Comment:

1. The first 2 studies might yield information relevant to the risks of the trial and to actual compliance during the trial.

2. The second 2 studies lacked sufficient participation to draw any meaningful conclusions.

13.0 Sponsor's Summary of Safety and Efficacy

The ERSMAC report contained tables summarizing a global safety analysis: