AD_____

Award Number: DAMD17-97-1-7058

TITLE: Development of an Integrated Program of Health-Related Quality-of-Life Research for the National Surgical Adjuvant Breast and Bowel Project

PRINCIPAL INVESTIGATOR: Richard D. Day, Ph.D.

CONTRACTING ORGANIZATION: University of Pittsburgh
Pittsburgh, Pennsylvania 15260

REPORT DATE: September 1999

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1999	Annual (1 Sep 98 - 31 Aug 99)

4. TITLE AND SUBTITLE
Development of an Integrated Program of Health-Related Quality-of-Life Research for the National Surgical Adjuvant Breast and Bowel Project

5. FUNDING NUMBERS
DAMD17-97-1-7058

6. AUTHOR(S)
Richard D. Day, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

E-MAIL:
rdfac@vms.cis.pitt.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

This is the 2nd Year (months 13-24) Report for a Career Development Award for the development of a Health-Related Quality of Life (HRQL) Program for the National Surgical Adjuvant Breast and Bowel Project (NSABP). Specific aims proposed for the award included: (a) Design and implementation of new HRQL components for planned NSABP treatment and prevention trials; (b) testing and implementation of data collection methods to be used in treatment and prevention trials; (c) analysis of HRQL data collected in the NSABP prevention and treatment trials; (d) refinement and extension of HRQL methods to analyze the data from new treatment and prevention studies; (e) enhancement of minority participation in NSABP trials. Primary achievements for the months 13-24 described in this report include: (i) Implementation of two new NSABP protocols with HRQL components (P-2, B-30); (ii) completion of patient recruitment in the first two NSABP treatment protocols with a HRQL component (B-23, C-06); (iii) conducting successful NSABP HRQL Workshop; (iv) development of real-time monitoring methods to reduce missing data in NSABP HRQL studies; (v) publication of initial quality of life communication from the NSABP P-1 protocol; (vi) development of a minority recruitment program for the P-2 (STAR) HRQL component.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	76
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

___ Where copyrighted material is quoted, permission has been obtained to use such material.

___ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

___ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

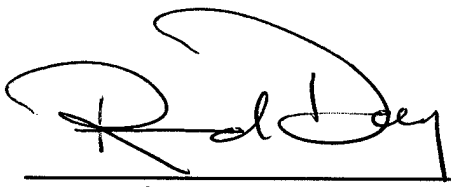
N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.



PI - Signature Date

Table of Contents

Career Development Award:

Development of an Integrated Program of Health-Related Quality of Life Research for the National surgical Adjuvant Breast and Bowel Project

Richard Day, Ph.D.
Department of Biostatistics
University of Pittsburgh

Second Annual Progress Report
September 1 1998 to September 30 1999

1. Introduction

This Career Development Award (CDA) was specifically intended to support Dr. Day in the development of a Health-Related Quality of Life Program (HRQL) for the National Surgical Adjuvant Breast and Bowel Project (NSABP). Specific aims proposed for the CDA included: (a) Design and implementation of new HRQL components for planned NSABP treatment and prevention trials; (b) testing and implementation of data collection methods to be used in treatment and prevention trials; (c) analysis of HRQL data collected in the NSABP prevention and treatment trials; (d) refinement and extension of HRQL methods to analyze the data from new treatment and prevention studies; (e) enhancement of minority participation in NSABP trials and the implementation of measures focusing on HRQL-related issues in women of color. Work completed during the second 12 months of Dr. Day's CDA will be summarized in terms of these specified aims and the schedule of technical objectives specified in the original proposal.

2. Body

2.1 Design and implementation of new HRQL components for planned NSABP treatment and prevention trials

a. **Protocol no. B-30** – *A Three Arm Randomized Trial to Compare Adjuvant Adriamycin and Cyclophosphamide Followed by Taxotere (AC-T); Adriamycin and Taxotere (AT); and Adriamycin, Taxotere and Cyclophosphamide (ATC) in Breast Cancer Patients with Positive Axillary Lymph Notes.* A complete HRQL component was developed for this study as an integral part of the trial protocol. Quality of life measures included in B-30 questionnaire are the Functional Assessment of Cancer Therapy-Breast (FACT-B), a treatment specific Symptom Checklist (SCL), the SF-36 Vitality Scale, and an overall HRQL rating scale (Appendix 2). Two additional questionnaires (Baseline and Follow-Up Menstrual History Questionnaires (Appendix 2) were developed in collaboration with Dr. Patricia Ganz, UCLA, to measure ovarian damage occurring as a long-term sequela of adjuvant chemotherapy. Spanish and French versions of these instruments are available.

b. **Protocol no. P-2** – *Study of Tamoxifen and Raloxifene (STAR).* This is the new NSABP prevention study following on the positive results of the P-1 (Breast Cancer Prevention Trial) Protocol. A new HRQL component was developed and approved by the National Cancer Institute and integrated into the study protocol. The P-2 HRQL questionnaire (Appendix 3) will be given to a sub-sample of the complete STAR cohort (22,000 women); the NSABP application to the Cancer Prevention and Control Protocol Review Committee was approved to give cancer control credits to CCOPS participating in this research.

c. **Protocols in development** – Two specific protocols are being considered for HRQL components and the present time:

Protocol no. B-33 – *A Randomized, Placebo Controlled, Double-Blinded Trial Evaluating the Effect of Exemestane in Stage I and II Post-Menopausal Breast Cancer Patients Completing at least Five Years of Tamoxifen Therapy.* The proposed HRQL component of this protocol will use the Menopause Specific Quality of Life Questionnaire (1, Appendix 4). Use of this instrument will permit a comparison of the B-33 findings to a study of letrozole being carried out by the NCI of Canada.

Protocol no. B-32 – *A Randomized, Phase III Clinical Trial to Compare Sentinel Node Resection to Conventional Axillary Dissection in Clinically Node Negative Breast Cancer Patients.* The development of a short

HRQL questionnaire for range of arm movement an pain following surgical resection is being cosidered.

d. **NSABP Quality of Life Workshop**, March 6-7, 1999, Pittsburgh, PA – Dr. Day served as the Co-Chair of this workshop which was designed to review the NSABP experience to date with HRQL studies and provide recommendations for future directions in HRQL research (Appendix 5). Workshop attendees included representatives from NSABP Headquarters and Biostatistical Centers, the NSABP Treatment and Prevention Quality of Life Committees, the NSABP Breast and Colon Cancer Committees, the National Cancer Institute, and other clinical trials groups (e.g., SWOG, ECOG). The primary goals of this workshop were to raise the visibility of HRQL research among the NSABP membership and to establish operating procedures, disease priorities and criteria for selection of future studies to have a HRQL components.

2.2 Testing and implementation of data collection methods to be used in treatment and prevention trials

a. **Operational methods for the reduction of missing data** - The initial NSABP HRQL components were implemented in protocols B-23 and C-06. Patient recruitment was completed in both of these trials over the past 10 months – recruitment was closed for B-23 on 31 December 1998 and for C-06 on 31 March 1999. Analysis of the initial data from these two protocols indicated marginal follow-up rates approximating 70% of the patients remaining on trial for at least 12 months. A review of operational procedures for data collection indicated that special measures need to be implemented in order to insure more complete follow-up treatment trial patients. Two specific measures have been proposed which are currently being designed for implementation in the B-30 and succeeding trials.

Real-time monitoring of data collection – This procedure will involve the monthly review of HRQL follow-up data for each operative protocol by a NSABP Biostatistical Center staff member. Clinical centers missing scheduled HRQL follow-up interviews will be contacted by a member of the NSABP Headquarters staff.

Missing data forms – HRQL follow-up interviews will become part of the routine delinquency assessment procedures. If a clinical center fails to complete a scheduled follow-up interview, they will be required to submit a Missing Data Form (Appendix 6) which provides information regarding the reasons for the failed interview.

The goals of these procedures are to increase the percentage of expected follow-up interviews and to provide some estimate of potential

biases in the trial data due to missing HRQL interviews. To out knowledge, these procedures have never been implemented before in large scale, multi-center clinical trials.

b. **Scannable forms** – Experimentation over the past 18 months, particularly in the C-06 protocol, have shown scannable HRQL forms to be less helpful than originally believed. In particular, they tend to exclude HRQL forms from the routine NSABP data processing procedures. As a result, it becomes difficult to trace the HRQL forms and to integrate them with routine delinquency assessments. As a result, all of the new NSABP treatment and prevention protocols have implemented traditional double entry data processing for HRQL forms.

2.3 Analysis of HRQL data collected in the NSABP prevention and treatment trials

Papers published:

Day R., Ganz P., Costantino J, Cronin W., Wickerham L and Fisher B., Health-related quality of life and tamoxifen in breast cancer prevention: A report from the National Surgical Adjuvant Breast and Bowel Project. J Clin Oncol 17, 1999, 2659-2669 (Appendix 7).

Papers in progress:

a. With Dr. Lisa A. Weissfeld (mentor): A comparison of mutivariate methods for the analysis of complex longitudinal study data with missing observations. This paper compares the practical (inferential) and theoretical implications of different methods (SAS Proc Mixed, GEE, and copula models) for the longitudinal analysis of P-1 data sets.

b. With Dr. Patricia Ganz (mentor): A detailed analysis of the P-1 data on rates of depressive symptoms in different at risk groups in the tamoxifen and control arms. This paper stratifies the P-1 cohort on baseline factors that predict a vulnerability to depression and poor psycho-social functioning (eg., a prior diagnosed history of depression or emotional illness, current and past psychiatric drug prescription, a history of undiagnosed periods of depression lasting months or years), then compares the rates of depressive symptoms among the stratified groups in each arm at follow-up. Patients that quit their study medication or quit the trial are assessed for evidence of depression or prescriptions of psychiatric medication.

c. With Dr. Patricia Ganz and Dr. David Cella (mentors): Factor analysis of the P-1 43-item Symptom Checklist data. Initial analyses suggest that most of the variance in baseline SCL scores can be explained by a

small number (7 or 8) independent latent variables. The goal of this paper is to simplify the SCL for future prevention studies and assess the stability of these initial latent factors on follow-up in the tamoxifen and placebo arms.

2.4 Refinement and extension of HRQL methods to analyze the data from new treatment and prevention studies

The work outlined above in Section 2.3 with Dr. Lisa A. Weissfeld, Dr. Day's statistical mentor for this award, applies to this specific aim. To date, the work under this specific aim has been limited to the P-1 prevention trial data. However, over the next 12 months, data will become available for analysis from the initial treatment trials designed by Dr. Day and his colleagues (B-23, C-06). Currently, follow-up is being completed in these treatment trials and data should be available for analysis in early 2000.

2.5 Enhancement of minority participation in NSABP trials and the implementation of measures focusing on HRQL-related issues in women of color

While planning recruitment for the HRQL component of the P-2 (STAR) trial, it was noted that the planned sample size (n=2000) of CCOP participants was unlikely to yield sufficient minority participation to permit statistical analysis. The following recommendations were put forward in order to increase the number of minority participants providing HRQL data in the P-2 study:

a. Minority completion of the HRQL questionnaire in the CCOP centers would not be cut-off after 2000 participants. All minority participants recruited in the CCOPs would be included in the HRQL component for the full period of the P-2 protocol.

b. Non-CCOP centers participating in the NSABP five city Community Outreach Program would collect HRQL information on all minority participants for the full period of the P-2 study.

2.6 Summary and conclusion

a. Summary of important achievements (Months 13-24)

- Implementation of two new NSABP protocols with HRQL components (P-2, B-30);
- Completion of patient recruitment in the first two NSABP treatment protocols with a HRQL component (B-23, C-06);
- Conducting successful NSABP HRQL Workshop;

- Development of real-time monitoring methods to reduce missing data in NSABP HRQL studies;
- Publication of initial quality of life communication from the NSABP P-1 protocol;
- Development of a minority recruitment program for the P-2 (STAR) HRQL component.

b. Summary of important delays (Months 13-24)

- The use of scannable forms in NSABP HRQL studies has been indefinitely delayed for operational reasons;
- Long-term breast cancer survivor study delayed due to NSABP emphasis on new treatment trials;
- Selection of minority members to NSABP HRQL Committees postponed due to selection of new committee chairs.

c. Planning priorities for months 24-36

The primary emphasis of work over the past 24 months has, by necessity, involved a focus on the operational tasks required for the development of an NSABP HRQL program. This has required that a significant portion of Dr. day's time be committed to the design of new treatment and prevention protocols, the selection and translation of HRQL instruments, the monitoring of data, and the development of operational procedures to facilitate data collection and processing and to cope with issues emerging from initial studies (e.g., missing data). As a result of this work, the data from the initial treatment trials containing a HRQL component will become available in the near future. Over the next twelve months, the primary goal will be to shift the emphasis of Dr. Day's work to the areas of methodological approaches to longitudinal data analysis and the publication of findings from completed treatment and prevention studies.

References in the Text

1. Hilditch J, Lewis J, Peter A van Maris B, Ross A, Franssen E, Guyatt G, Norton P, Dunn E, A menopause-specific quality of life questionnaire: development and psychometric properities. Maturitas 24 (1996) pp. 161-175.

Appendix 1

Key Research Accomplishments (Months 13-24)

Summary of Important Achievements
(Months 13-24)

- Implementation of two new NSABP protocols with HRQL components (P-2, B-30)
- Completion of patient recruitment in the first two NSABP treatment protocols with a HRQL component (B-23, C-06)
- Conducting successful NSABP HRQL Workshop
- Development of real-time monitoring methods to reduce missing data in NSABP HRQL studies
- Publication of initial quality of life communication from the NSABP P-1 protocol
- Development of a minority recruitment program for the P-2 (STAR) HRQL component

Appendix 2

B-30 HRQL Questionnaires

NSABP Protocol B-30
Quality of Life Questionnaire

INSTRUCTIONS TO INSTITUTION

This form applies to participants in the quality of life study.
The first page is to be completed by NSABP institution personnel.

Baseline Questionnaire: The baseline questionnaire must be administered before randomization. After the patient has signed the B-30 consent form, fill in the items listed below (leaving study number temporarily blank). Print the first 3 letters of the patient's last name at the top of page 3, and give the questionnaire to the patient for completion. After the patient has completed the questionnaire, verify that the date has been recorded on page 2. When the patient has been randomized, add the assigned study number to page 1 and submit the completed questionnaire to the NSABP Biostatistical Center.

Follow-up Questionnaires: Fill in the items listed below, print the first 3 letters of the patient's last name at the top of page 3, and give the questionnaire to the patient for completion. After the patient has completed the questionnaire, verify that the date has been recorded on page 2 and submit the completed questionnaire to the NSABP Biostatistical Center.

Patient Name _____

Study Number | 7 | 0 | | | | | | | | (1-9)

First 3 Letters of Patient's Last Name | | | | (10-12)

Time Point for this Questionnaire | | (13)

Evaluations should be carried out according to schedule, even if the patient discontinues protocol therapy or experiences breast cancer recurrence or second primary cancer.

0 - Baseline
[prior to randomization after consent form has been signed]

1 - Day 1 of Cycle #4
[at week 9. by mail or telephone, if the patient has been taken off protocol therapy]

2 - 6-Month Follow-up
[If AC →T patient will not have completed Taxotere by 6 months, delay questionnaire until 3 weeks after the last dose of Taxotere has been administered]

3 - 12-Month Follow-up

4 - 18-Month Follow-up

5 - 24-Month Follow-up

INSTRUCTIONS TO PATIENT

Please complete the following questionnaire by circling the number that corresponds to your response to each question. If you have any questions about how to answer the items in this questionnaire, please ask a staff member for help. Please use a pencil (rather than a pen) so that you will be able to erase a circle if you decide to change your response.

All information collected in this questionnaire will be kept confidential and will be used only for research purposes. If you feel uncomfortable about answering any question(s), you may leave the item blank. Your answers will not affect your continued participation in the B-30 trial.

Please write the date in the boxes provided below.

(14-21)

Date this questionnaire is completed:
(For example, if you were completing the questionnaire on September 8, 1998, you would write 09 08 1998 in the boxes.)

Mo	Day	Year	

As of today:

	excellent	very good	good	fair	poor	
1. In general, would you say that your health is	0	1	2	3	4	(22)

First 3 Letters of Patient's Last Name | | | |

Please indicate how true each statement has been for you <u>in the past 7 days</u>.

	not at all	a little bit	somewhat	quite a bit	very much	
2. I have a lack of energy.	0	1	2	3	4	(23)
3. I have nausea.	0	1	2	3	4	(24)
4. Because of my physical condition, I have trouble meeting the needs of my family.	0	1	2	3	4	(25)
5. I have pain.	0	1	2	3	4	(26)
6. I am bothered by side effects of treatment.	0	1	2	3	4	(27)
7. I feel ill.	0	1	2	3	4	(28)
8. I am forced to spend time in bed.	0	1	2	3	4	(29)
9. I feel distant from my friends.	0	1	2	3	4	(30)
10. I get emotional support from my family.	0	1	2	3	4	(31)
11. I get support from my friends and neighbors.	0	1	2	3	4	(32)
12. My family has accepted my illness.	0	1	2	3	4	(33)
13. Family communication about my illness is poor.	0	1	2	3	4	(34)
14. I feel close to my partner (or the person who is my main support).	0	1	2	3	4	(35)
15. Have you been sexually active during the past year? 1 = No 2 = Yes						(36)
If Yes: I am satisfied with my sex life.	0	1	2	3	4	(37)

Please indicate how true each statement has been for you in the past 7 days.

	not at all	a little bit	somewhat	quite a bit	very much	
16. I have confidence in my doctor(s).	0	1	2	3	4	(38)
17. My doctor is available to answer my questions.	0	1	2	3	4	(39)
18. I feel sad.	0	1	2	3	4	(40)
19. I am proud of how I'm coping with my illness.	0	1	2	3	4	(41)
20. I am losing hope in the fight against my illness.	0	1	2	3	4	(42)
21. I feel nervous.	0	1	2	3	4	(43)
22. I worry about dying.	0	1	2	3	4	(44)
23. I worry that my condition will get worse.	0	1	2	3	4	(45)
24. I am able to work (include work in home).	0	1	2	3	4	(46)
25. My work (include work in home) is fulfilling.	0	1	2	3	4	(47)
26. I am able to enjoy life.	0	1	2	3	4	(48)
27. I have accepted my illness.	0	1	2	3	4	(49)
28. I am sleeping well.	0	1	2	3	4	(50)
29. I am enjoying the things I usually do for fun.	0	1	2	3	4	(51)
30. I am content with the quality of my life right now.	0	1	2	3	4	(52)

Please indicate how true each statement has been for you <u>in the past 7 days</u>.

	not at all	a little bit	somewhat	quite a bit	very much	
31. I have been short of breath.	0	1	2	3	4	(53)
32. I am self-conscious about the way I dress.	0	1	2	3	4	(54)
33. One or both of my arms are swollen or tender.	0	1	2	3	4	(55)
34. I feel sexually attractive.	0	1	2	3	4	(56)
35. I am bothered by hair loss.	0	1	2	3	4	(57)
36. I worry about the risk of cancer in other family members.	0	1	2	3	4	(58)
37. I worry about the effect of stress on my illness.	0	1	2	3	4	(59)
38. I am bothered by a change in weight.	0	1	2	3	4	(60)
39. I am able to feel like a woman.	0	1	2	3	4	(61)

Please indicate how much you have been bothered by each of the following problems <u>in the past 7 days</u>.

	not at all	a little bit	somewhat	quite a bit	very much	
40. headaches	0	1	2	3	4	(62)
41. vomiting	0	1	2	3	4	(63)
42. mouth sores	0	1	2	3	4	(64)

Please indicate how much you have been bothered by each of the following problems in the past 7 days.

	not at all	a little bit	somewhat	quite a bit	very much	
43. diarrhea	0	1	2	3	4	(65)
44. skin problems (including rash, dry skin, irritation or redness)	0	1	2	3	4	(66)
45. numbness or tingling in hands or feet	0	1	2	3	4	(67)
46. fever or shivering (shaking, chills)	0	1	2	3	4	(68)
47. difficulty with bladder control	0	1	2	3	4	(69)
48. constipation	0	1	2	3	4	(70)
49. hot flashes	0	1	2	3	4	(71)
50. genital itching or irritation	0	1	2	3	4	(72)
51. mood swings	0	1	2	3	4	(73)
52. vaginal discharge	0	1	2	3	4	(74)
53. vaginal bleeding or spotting	0	1	2	3	4	(75)
54. vaginal dryness	0	1	2	3	4	(76)
55. pain with intercourse	0	1	2	3	4	(77)
56. cramps	0	1	2	3	4	(78)
57. general aches and pains	0	1	2	3	4	(79)

Please indicate how much you have been bothered by each of the following problems in the past 7 days.

	not at all	a little bit	somewhat	quite a bit	very much	
58. joint pains	0	1	2	3	4	(8
59. swelling of hands or feet	0	1	2	3	4	(8
60. muscle stiffness	0	1	2	3	4	(8
61. weight gain	0	1	2	3	4	(8
62. weight loss	0	1	2	3	4	(8
63. unhappy with appearance of my body	0	1	2	3	4	(8
64. forgetfulness	0	1	2	3	4	(8
65. night sweats	0	1	2	3	4	(8
66. cold sweats	0	1	2	3	4	(8
67. difficulty concentrating	0	1	2	3	4	(8

68. Please score your overall quality of life in the past 7 days on an 11-point scale between death and perfect health.

(90-91)

death	0	1	2	3	4	5	6	7	8	9	10	perfect health

NSABP Protocol B-30
Baseline Menstrual History Questionnaire

(For all B-30 patients)

INSTRUCTIONS TO INSTITUTION

After the patient has signed the B-30 consent form, fill in the items listed below (except study number), print the patient's name at the top of page 2, and give the questionnaire to the patient for completion.

The completed form must be faxed to the NSABP Biostatistical Center prior to randomization along with the other required pre-entry materials. After the patient has been randomized, record the assigned study number in the shaded boxes provided and retain the completed questionnaire for your records.

Patient Name _____

Study Number | 7 | 0 | | | | | | | | (1-9)

First 3 Letters of Patient's Last Name | | | | (10-12) (13) = 0

Date Questionnaire Completed | | | Month | | | Day | | | | | Year (14-21)

Patient's Name _____

INSTRUCTIONS TO PATIENT

Please answer all questions that apply by filling in the boxes provided.

SECTION A

- **Have you had your uterus removed?** ☐ (22)

  1 - Yes
  2 - No
  3 - Unsure

- **Have you had <u>both</u> ovaries removed?** ☐ (23)

  1 - Yes
  2 - No
  3 - Unsure

- **Have you had a menstrual period in the last 12 months?** ☐ (24)

  1 - Yes *(if yes, continue to Section B)*
  2 - No *(if no, remaining questions do not apply)*
  3 - Unsure *(if unsure, remaining questions do not apply)*

SECTION B

- **When was your last menstrual period (or vaginal bleeding)?** ☐☐ Month ☐☐☐☐ Year (25-30)

- **During the past 12 months, did the time between your menstrual periods change in any way?** ☐ (31)

  1 - No changes in periods, cycles usually regular
  2 - No changes in periods, cycles usually irregular
  3 - Yes, periods went from regular to irregular
  4 - Yes, periods stopped and have not resumed
  5 - Unsure

- **Have you taken hormone replacement therapy (estrogen and/or progesterone) at any time in the last 12 months?** ☐ (32)

  1 - Yes
  2 - No
  3 - Unsure

- **Have you taken hormonal contraceptives (birth control pills, Norplant, Depo-Provera) at any time in the last 12 months?** ☐ (33)

  1 - Yes
  2 - No
  3 - Unsure

NSABP Protocol B-30
Follow-Up Menstrual History Questionnaire

(For participants in the follow-up phase of the Menstrual History Study)

INSTRUCTIONS TO INSTITUTION

Fill in the items listed below, print the patient's name at the top of page 2, and give the questionnaire to the patient for completion. After the patient has completed the questionnaire, submit the completed questionnaire to the NSABP Biostatistical Center.

Patient Name _____

Study Number | 7 | 0 | | | | | | | | (1-9)

First 3 Letters of Patient's Last Name | | | | (10-12)

Time Point for this Questionnaire | | (13)

Evaluations should be carried out according to schedule, even if the patient discontinues protocol therapy or experiences breast cancer recurrence or second primary cancer.

1 - Day 1 of Cycle #4
[at week 9, by mail or telephone, if the patient has been taken off protocol therapy]

2 - 6-Month Follow-up
[If AC → T patient will not have completed Taxotere by 6 months, delay questionnaire until 3 weeks after the last dose of Taxotere has been administered]

3 - 12-Month Follow-up

4 - 18-Month Follow-up

5 - 24-Month Follow-up

Date Questionnaire Completed | | | Month | | | Day | | | | Year | (14-21)

Patient's Name _____

INSTRUCTIONS TO PATIENT

Please answer all questions by filling in the boxes provided.

- **Have you had your uterus removed?** ☐ (22)

  1 - Yes
  2 - No
  3 - Unsure

- **Have you had <u>both</u> ovaries removed?** ☐ (23)

  1 - Yes
  2 - No
  3 - Unsure

- **Have you had a menstrual period in the last 12 months?** ☐ (24)

  1 - Yes
  2 - No
  3 - Unsure

- **When was your last menstrual period (or vaginal bleeding)?** ☐☐ Month ☐☐☐☐ Year (25-30)

- **During the past 12 months, did the time between your menstrual periods change in any way?** ☐ (31)

  1 - No changes in periods, cycles usually regular
  2 - No changes in periods, cycles usually irregular
  3 - Yes, periods went from regular to irregular
  4 - Yes, periods stopped and have not resumed
  5 - Unsure

- **Have you taken hormone replacement therapy (estrogen and/or progesterone) at any time in the last 12 months?** ☐ (32)

  1 - Yes
  2 - No
  3 - Unsure

- **Have you taken hormonal contraceptives (birth control pills, Norplant, Depo-Provera) at any time in the last 12 months?** ☐ (33)

  1 - Yes
  2 - No
  3 - Unsure

Appendix 3

P-2 HRQL Questionnaire

QLQ

6/7/99

NSABP STUDY OF TAMOXIFEN AND RALOXIFENE

Quality of Life Questionnaire

(affix participant barcode label here)

Instructions to STAR Staff

Baseline Questionnaire: The baseline questionnaire must be administered before randomization. Print the first 3 letters of the participant's last name at the top of pages 2 through 5, and give the questionnaire to the participant for completion. When the participant has been randomized, add the assigned study number to page 1 and submit the completed questionnaire to the NSABP Biostatistical Center.

Follow-up Questionnaires: Fill in the items listed below, print the first 3 letters of the participant's last name on pages 2 through 5, and give the questionnaire to the participant for completion. Submit the completed questionnaire to the NSABP Biostatistical Center.

Participant Last Name _____

Study Number [1-9] | S | | | | | | | | |

Clinical Center Code [10-12] | | | | Subcenter Code [13-14] | | |

Date Form Completed [15-22] | | | | | | | | | |
Month Day Year

Contact/Clinic Visit [23] | | |

0 - Baseline
1 - 6 Month
2 - 12 Month
3 - 18 Month
4 - 24 Month
5 - 30 Month
6 - 36 Month
7 - 42 Month
8 - 48 Month
9 - 54 Month
10 - 60 Month
11 - 66 Month
12 - 72 Month
13 - 78 Month
14 - 84 Month

First 3 Letters of Participant's Last Name | | | |

Instructions to Participant

Please fill out this questionnaire and return it to a STAR staff member. If you have any questions about how to answer the items on this form, please ask for help. All of the information collected in this questionnaire will be kept strictly confidential and will be used only for research purposes. If you feel uncomfortable about answering any of these questions, please leave the item blank. Your answers to these questions will not affect your continued participation in this study.

YOUR FEELINGS DURING THE PAST WEEK

Below is a list of feelings, attitudes, and behaviors that you may have experienced during the PAST WEEK.
Please use the following scale and circle the one response that best describes how often you have had these experiences.

0 = Rarely or none of the time (less than one day)
1 = Some or a little of the time (1 - 2 days)
2 = Moderately (3 - 4 days)
3 = Most of the time (5 - 7 days)

	Rarely	Some of the Time	Moder-ately	Most of the Time	
I was bothered by things that usually don't bother me.	0	1	2	3	[24]
I did not feel like eating: my appetite was poor.	0	1	2	3	[25]
I felt that I could not shake off the blues even with help from family and friends.	0	1	2	3	[26]
I felt that I was just as good as other people.	0	1	2	3	[27]
I had trouble keeping my mind on what I was doing.	0	1	2	3	[28]
I felt depressed (blue or down).	0	1	2	3	[29]
I felt that everything I did was an effort.	0	1	2	3	[30]
I felt hopeful about the future.	0	1	2	3	[31]
I thought my life had been a failure.	0	1	2	3	[32]
I felt fearful.	0	1	2	3	[33]
My sleep was restless.	0	1	2	3	[34]
I was happy.	0	1	2	3	[35]
I talked less than usual.	0	1	2	3	[36]
I felt lonely.	0	1	2	3	[37]
People were unfriendly.	0	1	2	3	[38]
I enjoyed life.	0	1	2	3	[39]
I had crying spells.	0	1	2	3	[40]
I felt sad.	0	1	2	3	[41]
I felt that people disliked me.	0	1	2	3	[42]
I could not get "going".	0	1	2	3	[43]

First 3 Letters of Participant's Last Name ☐☐☐

PERSONAL HEALTH

This survey asks for your views about your health. Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In *general*, would you say your health is:

☐ [44]

0 = Excellent
1 = Very Good
2 = Good
3 = Fair
4 = Poor

2. Compared to *six months ago*, how would you rate your health in general now?

☐ [45]

0 = Much better now than 6 months ago
1 = Somewhat better now than 6 months ago
2 = About the same as 6 months ago
3 = Somewhat worse now than 6 months ago
4 = Much worse now than 6 months ago

3. The following items are about activities you might do during a typical day. Does *your health now limit you* in these activities? If so, how much? Please circle the one response that best describes your answer.

Activities	Yes, Limited A Lot	Yes, Limited A Little	No, Not Limited At all	
a. ***Vigorous activities,*** such as running, lifting heavy objects, participating in strenuous sports	0	1	2	[46]
b. ***Moderate activities,*** such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	0	1	2	[47]
c. Lifting or carrying groceries	0	1	2	[48]
d. Climbing ***several*** flights of stairs	0	1	2	[49]
e. Climbing ***one*** flight of stairs	0	1	2	[50]
f. Bending, kneeling, or stooping	0	1	2	[51]
g. Walking ***more than a mile***	0	1	2	[52]
h. Walking ***several blocks***	0	1	2	[53]
i. Walking ***one block***	0	1	2	[54]
j. Bathing or dressing yourself	0	1	2	[55]

4. During the *past 4 weeks*, have you had any of the following problems with your work or other regular daily activities *as a result of your physical health*? Circle your response.

	YES	NO	
a. Cut down the ***amount of time*** you spent on work or other activities	0	1	[56]
b. ***Accomplished less*** than you would like	0	1	[57]
c. Were limited in the ***kind*** of work or other activities	0	1	[58]
d. Had ***difficulty*** performing the work or other activities (for example, it took extra effort)	0	1	[59]

5. During the *past 4 weeks*, have you had any of the following problems with your work or other regular daily activities *as a result of any emotional problems* (such as feeling depressed or anxious)? Circle your response.

	YES	NO	
a. Cut down the ***amount of time*** you spent on work or other activities	0	1	[60]
b. ***Accomplished less*** than you would like	0	1	[61]
c. Didn't do work or other activities as ***carefully*** as usual	0	1	[62]

6. During the *past 4 weeks*, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups? [63]

0 = Not at all
1 = Slightly
2 = Moderately
3 = Quite a Bit
4 = Extremely

7. How much *bodily* pain have you had during the *past 4 weeks*? [64]

0 = None
1 = Very Mild
2 = Mild
3 = Moderate
4 = Severe
5 = Very Severe

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and house work)? [65]

0 = Not at all
1 = A Little Bit
2 = Moderately
3 = Quite a Bit
4 = Extremely

9. These questions are about how you feel and how things have been with you during the *past 4 weeks*. Circle the one answer that comes closest to the way you have been feeling. How much of the time during the *past 4 weeks*:

	All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time	
a. Did you feel full of pep?	0	1	2	3	4	5	[66]
b. Have you been a very nervous person?	0	1	2	3	4	5	[67]
c. Have you felt so down in the dumps that nothing could cheer you up?	0	1	2	3	4	5	[68]
d. Have you felt calm and peaceful?	0	1	2	3	4	5	[69]
e. Did you have a lot of energy?	0	1	2	3	4	5	[70]
f. Have you felt downhearted and blue?	0	1	2	3	4	5	[71]
g. Did you feel worn out?	0	1	2	3	4	5	[72]
h. Have you been a happy person?	0	1	2	3	4	5	[73]
i. Did you feel tired?	0	1	2	3	4	5	[74]

First 3 Letters of Participant's Last Name ☐☐☐

10. During the *past 4 weeks*, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)? ☐ [75]

0 = All of the Time
1 = Most of the Time
2 = Some of the Time
3 = A Little of the Time
4 = None of the Time

11. How TRUE or FALSE is each of the following statements for you? Circle your response.

	Definitely True	Mostly True	Don't Know	Mostly False	Definitely False	
a. I seem to get sick a little easier than other people.	0	1	2	3	4	[76]
b. I am as healthy as anybody I know.	0	1	2	3	4	[77]
c. I expect my health to get worse.	0	1	2	3	4	[78]
d. My health is excellent.	0	1	2	3	4	[79]

12. Have you been sexually active during the past SIX MONTHS? ☐ [80]

0 = No
1 = Yes

13. If you have been sexually active in the *past four weeks,* please answer the following questions. If you have not been sexually active, leave the items blank.

How much of a problem was each of the following over the past FOUR WEEKS?

a. Lack of sexual interest

☐ [81]

0 = Not a Problem
1 = A Little Problem
2 = A Definite Problem
3 = A Serious Problem

b. Difficulty in becoming sexually aroused

☐ [82]

0 = Not a Problem
1 = A Little Problem
2 = A Definite Problem
3 = A Serious Problem

c. Unable to relax and enjoy sex

☐ [83]

0 = Not a Problem
1 = A Little Problem
2 = A Definite Problem
3 = A Serious Problem

d. Difficulty in having orgasm

☐ [84]

0 = Not a Problem
1 = A Little Problem
2 = A Definite Problem
3 = A Serious Problem

Signature of Person Completing This Form _____

Thank You for Your Time In Completing This Questionnaire

Appendix 4

Hilditch J, Lewis J, Peter A van Maris B, Ross A, Franssen E, Guyatt G, Norton P, Dunn E, A menopause-specific quality of life questionnaire: development and psychometric properities. Maturitas 24 (1996) pp. 161-175.

ELSEVIER

Maturitas 24 (1996) 161–175

MATURITAS
JOURNAL OF THE CLIMACTERIC & POSTMENOPAUSE

A menopause-specific quality of life questionnaire: development and psychometric properties

John R. Hilditch[a,*], Jacqueline Lewis[a], Alice Peter[b], Barbara van Maris[a], Alan Ross[c], Edmée Franssen[d], Gordon H. Guyatt[e], Peter G. Norton[a], Earl Dunn[a]

[a]*Primary Care Research Unit, Room E349, Department of Family and Community Medicine, Sunnybrook Health Science Centre, University of Toronto, 2075 Bayview Ave., North York, Ontario, M4N 3M5, Canada*
[b]*HIV Project Centre, Sunnybrook Health Science Centre, 2075 Bayview Ave., North York, Ontario, M4N 3M5, Canada*
[c]*Mature Women Unit, Clinical Associates in Reproductive Endocrinology and Medicine, Mississauga, Ontario, Canada*
[d]*Clinical Trials and Epidemiology, Toronto Sunnybrook Regional Cancer Centre, Toronto, Ontario, Canada*
[e]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada*

Received 13 November 1995; revised 25 January 1996; accepted 4 April 1996

Abstract

Objective: to develop a condition-specific quality of life questionnaire for the menopause with documented psychometric properties, based on women's experience. *Methods*: Subjects: Women 2–7 years post-menopause with a uterus and not currently on hormone replacement therapy. Questionnaire development: A list of 106 menopause symptoms was reduced using the importance score method. Replies to the item-reduction questionnaire from 88 women resulted in a 30-item questionnaire with four domains, vasomotor, physical, psychosocial and sexual, and a global quality of life question. Psychometric properties: A separate sample of 20 women was used to determine face validity, and a panel of experts was used to confirm content validity. Reliability, responsiveness and construct validity were determined within the context of a randomized controlled trial.

Construct validation involved comparison with the Neugarten and Kraines' Somatic, Psychosomatic and Psychologic subscales, the reported intensity of hot flushes, the General Well-Being Schedule, Channon and Ballinger's Vaginal Symptoms Score and Libido Index, and the Life Satisfaction Index.

Results: The face validity score was 4.7 out of a possible 5. Content validity was confirmed. Test-retest reliability measures, using intraclass correlation coefficients were 0.81, 0.79, 0.70 and 0.55 for the physical, psychosocial, sexual domains and the quality of life question. The intraclass correlation coefficient for the vasomotor domain was 0.37 but there is evidence of systematic change. Discriminative construct validity showed correlation coefficients of 0.69 for the physical domain, 0.66 and 0.40 for the vasomotor domain, 0.65 and −0.71 for the psychosocial domain, 0.48 and 0.38 for the sexual domain, and 0.57 for the quality of life question. Evaluative construct validity showed correlation coefficients of 0.60 for the physical domain, 0.28 for the vasomotor domain, 0.55 and −0.54 for the psychosocial domain, 0.54 and 0.32 for the sexual domain, and 0.12 for the quality of life question. Responsiveness scores ranged from 0.78 to 1.34.

Conclusions: The MENQOL (Menopause-Specific Quality of Life) questionnaire is a self-administered instrument which functions well in differentiating between women according to their quality of life and in measuring changes in their quality of life.

Keywords: Menopause; Post-menopausal; Questionnaire; Quality of life; Methodology

* Corresponding author.

PII S0378-5122(96)01038-3

1. Introduction

Condition-specific quality of life can be defined as the extent that the physical, emotional and social aspects of an individual's life are intact and not adversely affected by that condition or treatment [1]. Only recently have studies considered the effect of a therapeutic intervention on the individual's evaluation of quality of life. Treatment goals may be achieved, but may produce unwanted side effects which can impact negatively on quality of life. It is increasingly recognized that the measurement of quality of life should be an integral part of any attempt to assess disease impact or to assess the effects of a medical intervention [2–4].

The purpose of this study was to develop a valid, reliable, responsive quality of life questionnaire specific to the post-menopausal period, based on women's reported experience.

2. Material and methods

2.1. Study population

Included were women 47–62 years old who had ceased menstruation for 2–7 years, who had not had a hysterectomy, and who had not used hormone therapy during the preceding 6 months. Women were excluded with contraindications to estrogen use or who had a current unstable medical or social problem. Participants were included in the questionnaire development regardless of their intention to take hormones or to take part in the subsequent randomized controlled trial of hormone therapy.

Recruitment utilized multiple sources: a mail-out to appropriately aged women who attended the Family Practice Department of a University of Toronto teaching hospital, newspaper articles and ads, strategically-placed posters, and word of mouth. Data from the first 88 completed item-reduction questionnaires were used to develop the instrument.

2.2. Item generation

A list of symptoms/problems which could be experienced by post-menopausal women was compiled from several sources: the menopause and quality of life literature, existing menopause and quality of life questionnaires [5–10] and the clinical experience of the investigators and consultants. In-depth semi-structured interviews were audiotaped with eight post-menopausal women and independently reviewed by two researchers to determine additional problems or symptoms. A panel of 8 medical and 2 quality of life experts reviewed the list for completeness. The final item-reduction questionnaire contained 106 items.

2.3. Question format

Respondents were asked to indicate if they had experienced each symptom or problem within the past month and, if so, to rate how bothersome it was on a seven-point Likert Scale [11]. The questionnaire allowed write-in of additional symptoms or problems.

To pretest the item-reduction questionnaire, seven women completed the questionnaire in the presence of the research coordinator.

2.4. Item reduction

For each question, an importance score was calculated as the product of the frequency that the problem was experienced, by the mean extent that the symptom was bothersome for women in whom the problem was experienced. The importance score has a larger value if a greater number of women experience the problem and/or if women find the problem more bothersome. The final item-reduction questionnaire consisted of 106 questions divided logically into physical, emotional, social, working life and sexual areas, plus additional demographic and attitudinal sections.

2.5. The Menopause-Specific Quality of Life (MENQOL) questionnaire

Before reduction, the researchers attributed each question to only one of five domains: physi-

cal, vasomotor, psychosocial, sexual and working life. Participants working outside the home had few complaints associated with working life so this section was dropped, leaving four domains. Each domain was required to have at least three questions, with a total of 30 questions in the final questionnaire [12].

The decision to include items in the final questionnaire was based on the importance score, with a few exceptions. In the areas of sleep disturbances, muscle and joint aches and tiredness, apparently redundant questions were eliminated from ranking by choosing the question with the higher importance score. The questions were ranked by domain from high to low importance score, and the top three questions in each domain were selected. Then, the remaining questions were ranked altogether from high to low and added to the appropriate domains in order of the value of their importance scores, until 29 questions were included. The final, overall quality of life question brought the total to 30 questions. See Appendix A.

2.6. Psychometric properties of the MENQOL questionnaire

Reliability, responsiveness and construct validity were determined in the course of a randomized parallel design trial comparing the effects of two forms of estrogen on quality of life [13]. Potential trial participants were seen twice, at monthly intervals, before randomization at the third visit to a 14-week course of therapy. Inclusion of participants' data for analysis required complete data by domain.

2.7. Reliability

Since no intervention was administered between the first two baseline visits for the randomized controlled trial, test-retest reliability was determined, by domain, using intraclass correlation coefficients [14,15]. Pearson's correlation coefficients were also calculated since this statistic has been more commonly reported. Internal consistency reliability (Cronbach's Alpha), measuring the extent that the questions in each domain tap a particular concept [16], was determined at the first baseline visit.

2.8. Validity

2.8.1. Face validity

A separate sample of post-menopausal women reviewed the questionnaire and answered the question: 'How well do you think the questionnaire measures the quality of a woman's life after she has gone through menopause?'. The women responded using a 5-point Likert scale from 1 (not at all) to 5 (very well).

2.8.2. Content validity

Ten experts in the area of quality of life or menopause examined the instrument for important omissions or inappropriate choice of items.

2.8.3. Construct validity

The domains of the instrument required validation against other instruments purported to measure the same construct. Validation questionnaires were chosen for their appropriateness to measure menopausal symptoms in a comparable population. The choice of questionnaires with known psychometric properties against which to validate the MENQOL questionnaire was limited. Validated instruments of functional ability and health status were not used because they were designed for studies of chronically ill people.

Table 4 displays the instruments chosen to compare with the MENQOL questionnaire. The Neugarten and Kraines' menopause questionnaire [9], developed from the Blatt and Kupperman Menopausal Index [8], is a self-administered questionnaire commonly used in published studies of the frequency of menopausal symptoms. Although test-retest reliability is reported as satisfactory, it is of unknown validity. The Somatic, Psychosomatic and Psychosocial subscales were used to validate the MENQOL Vasomotor, Physical and Psychosocial Domains respectively.

At baseline, participants were asked to report the intensity of the flushes they had experienced in the past 24 h (or past week if flushes were less frequent), measured on a visual analogue scale. This measure was also used to validate the Vasomotor domain.

The Psychosocial Domain was also validated against the General Well-being questionnaire (GWB) [5] which, unlike other measures of psychological well-being, has few items that can be interpreted as somatic in nature. The GWB has known and acceptable reliability and validity based on a time period of one month, matching the testing interval of the MENQOL questionnaire.

The Sexual Domain was validated against the 'Vaginal Symptoms score' and 'Libido Index' from the Channon and Ballinger questionnaire [6] used to investigate the relationship of anxiety and depression to sexual and vaginal symptoms in normal post-menopausal women. The instrument is not validated and has unknown reliability. Available sexual satisfaction scales were measures of sexual dysfunction and were therefore inappropriate for normal women.

The global quality of life question was validated against the 20 item Life Satisfaction Index A(LSI) [7] intended for use in an aging healthy population which is slightly older than the women in our study. However, it measures life satisfaction in physically healthy persons as opposed to other quality of life instruments developed for the chronically ill. There are no reliability data for this instrument but validity is reported.

The MENQOL questionnaire was completed by trial participants at each visit. The construct validation instruments were administered at either of the two baseline visits, and all were administered at the final visit, with the exception of the intensity of hot flushes. The methods used to score these questionnaires is described in Appendix B.

To determine discriminative construct validity, Pearson correlation coefficients were calculated between the scores of the validating instruments and the scores on the domains of our questionnaire at baseline. To determine evaluative construct validity, Pearson correlation coefficients were calculated between the change scores for the validating questionnaires and the change scores for the MENQOL questionnaire between the baseline and last visit.

2.9. Responsiveness to change

We examined responsiveness to change using the method described by Guyatt which relates the minimal clinically important difference to the variability in stable participants [17,18]. A clinically important difference was chosen to be '1' in a score that ranges from 0 to 6, based on the clinical opinion of the investigators. Responsiveness of each domain was determined by dividing this clinically important difference, 1, by the standard deviation of the change scores during the stable period between the two baseline visits. This responsiveness score also gives another measure of the adequacy of the sample size.

3. Results

3.1. Demographic characteristics of study participants

Table 1 displays the characteristics of the participants whose responses on the item-reduction questionnaire were used to develop the MENQOL questionnaire. In addition, about three-quarters were born in Canada, and English was the first language for 89%. Forty-three percent had one or more children.

Table 1
Demographic characteristics of item reduction participants

Characteristic	Number $n = 88$	Percent
Mean Age (years)	58	
Married	67	76.1
Divorced, separated	16	17.1
Completed secondary school	81	89.8
Family income > $40 000	56	63.6
Employed	55	62.5
Cigarette smoker	13	14.8
One or more drinks/day	23	26.1
Regular exercise	47	53.4
Age at menopause (years)	50.7	
Time since last period (months)	52.4	
Menopause symptoms	45	51.1
Past PMS	34	38.7
Past painful periods	33	37.5

Table 2
Test-retest reliability

Domain	Correlation coefficients		n
	Intra-class	Pearson	
Physical	0.81	0.82	101
Vasomotor	0.37	0.85	111
Psychosocial	0.79	0.81	104
Sexual	0.70	0.83	94
QOL	0.55	0.55	108

The characteristics of the participants in the RCT used for determining psychometric properties were, in most instances, very similar. More participants (80%) were working, a greater proportion (64%) reported menopausal symptoms in the past month, and they reported a shorter average time since last period (46 months).

3.2. Psychometric properties

3.2.1. Test-Retest reliability

The intraclass correlation coefficients and the Pearson correlation coefficients are shown in Table 2. The number of women for whom we had complete data varied by domain between 111 and 94. Cronbach's Alpha scores, measuring the internal consistency of questions in each domain are shown in Table 3. The number of women responding varied by domain between 144 and 107.

3.3. Validity

3.3.1. Face validity

All of the women rated the question 4 or higher and most gave it a top rating of 5, producing an overall mean of 4.7.

Table 3
Internal consistency

Domain	Cronbach's Alpha	n
Physical	0.87	137
Vasomotor	0.82	144
Psychosocial	0.81	139
Sexual	0.89	107

3.3.2. Content validity

The experts gave no advice for additions which could be incorporated into the 'frequency bothersome' question format. One expert would have preferred a quality of life questionnaire which incorporated both positive and negative items.

3.3.3. Construct validity

Table 4 displays the correlation coefficients measuring discriminative validity, between the domain scores and the validating instrument scores at baseline, and the correlation coefficients measuring evaluative validity, between the change scores from the pre- trial baseline to the last RCT visit. The number of participants completing the validating questionnaires varied by domain from 60 to 103 for discriminative validity, and from 24 to 54 for evaluative validity.

3.4. Responsiveness

At baseline, under stable conditions, the mean change score over 1 month ranged from -0.06 (QOL) to 0.28 (Psychosocial Domain). The calculated responsiveness scores ranged from 0.78 to 1.34 (Table 5). Calculations indicate that in order to detect a change of 1.0 in a scale of 7, with $\alpha = 0.05$ and $\beta = 0.1$, the sample size required per group would be 9-26, depending on the domain.

4. Discussion

Menopause is a unique experience. It is not a disease but rather a stage in development. Because of the complexity of changes during this time of life, assessment of quality of life must include not only consideration of physical changes, but also psychological and social changes.

Our intention to examine the effect of hormones on quality of life led to the examination of currently available measurement instruments. We required a scale of known psychometric properties and wanted the scale to be developed on the basis of women's own qualitative and quantitative experience, not simply on the basis of clinical or expert judgment. In our opinion, currently available questionnaires did not meet these criteria.

Table 4
Construct validity

MENQOL domain	Validation instrument	Discriminative (baseline)	Evaluative (change score)
Physical	Neugarten and Kraines' Menopause Symptom Checklist:		
	Somatic and Psychosomatic	0.69	0.6
	Subscales	(86)	(50)
		$P < 0.001$	$P < 0.001$
Vasomotor	Somatic Subscale	0.40	0.28
		(86)	(48)
		$P < 0.001$	$P = 0.57$
	Intensity of hot flushes	0.66	
		(102)	
		$P < 0.0001$	
Psychosocial	Psychologic Subscale	0.65	0.55
		(84)	(54)
		$P < 0.001$	$P < 0.001$
	The General Well-Being Schedule[a]	−0.70	−0.54
		(70)	(54)
		$P < 0.001$	$P < 0.001$
Sexual	Channon and Ballinger's Vaginal Symptoms Score	0.38	0.54
		(63)	(26)
		$P < 0.001$	$P = 0.004$
	Libido Index	0.48	0.32
		(60)	(24)
		$P < 0.001$	$P = 0.124$
Global Quality	Life Satisfaction Index	0.57	0.12
		(103)	(52)
		$P < 0.001$	$P < 0.391$

The numbers in parentheses represent the sample size.
[a]GWB and MENQOL scores move in opposite directions.

A limited number of assessment instruments exist which pertain specifically to the menopause. Early attempts at quantifying the menopausal experience consisted mainly of symptom indexes, the best known being an index developed by Blatt and Kupperman [8]. This index was further refined by Neugarten and Kraines [9], who added 17 more items to the scale and grouped them as somatic, psychosomatic and psychologic items. While these indices are useful in examining the frequency and severity of some symptoms commonly associated with certain areas of the menopause experience, they omit some areas such as sexual function. With regard to psychometric properties, only test-retest reliability was reported. The use of the Blatt-Kupperman or Neugarten-Kraines scales offers the possible advantage of making comparisons with older studies, but they were not designed to capture the effect of menopause on the health-related quality of life, nor were they tested for responsiveness.

Table 5
Responsiveness

Domain	Mean change at base-line	S.D.	Required sample size per group
Physical	0.21	0.75	9
Vasomotor	0.18	1.10	19
Psycho-social	0.28	0.91	13
Sexual	0.24	1.28	26
QOL	−0.06	1.05	17

Greene developed a symptom scale [10] largely based on Neugarten and Kraine's questionnaire and used factor analysis to identify three symptom groupings. The scale, which was developed using a sample of women referred by general practitioners to a 'Hormone Replacement Therapy Clinic' [19], has satisfactory psychometric proper

ties, but responsiveness was not measured, and it was not designed to measure quality of life.

Another approach to menopausal quality of life, used by Wiklund et al. [20] and Limouzin-Lamothe et al. [21] is to combine several existing instruments to measure the components of quality of life. This approach offers some confidence in using instruments which may have known psychometric properties. The results can be compared with other studies which used some or all of the same instruments on a variety of populations and conditions. In addition, by using existing instruments one can avoid the cost and time required to develop a new instrument.

However, there are disadvantages to this approach. Generally, the available instruments are not designed to be responsive to change and responsiveness has rarely been tested. In addition, questionnaires designed to assess emotional and psychological function, i.e. the General Well-being Scale, have generally been designed to measure community mental health or individual response to a wide range of problems. Although useful for the purposes for which they were designed, these instruments may not be specific enough to pick up subtle but important features in any one condition, such as the menopause. Also, to cover all relevant aspects of quality of life, many questionnaires must be administered, and for purposes of a clinical trial, the test battery must be administered repeatedly. The time and burden on the individual respondent may be too great for efficient and meaningful response. For example, the approach by Wiklund required response to more than 100 items and by Limouzin-Lamothe to 112 items.

Our condition-specific quality of life questionnaire considers the quality of life of those experiencing menopause. Focus on the experiences of this population and their participation in content development was felt to contribute to good face and content validity. The method chosen for questionnaire reduction, developed by Guyatt [12], has been used for instrument development for quality of life measurement for many conditions, including lung disease, inflammatory bowel disease, arthritis and congestive heart failure [22–25]. Unlike most questionnaires which are solely based on the frequency that symptoms are experienced, this method incorporates both frequency and the extent that the symptom is bothersome into a single score. Although most evaluations of effectiveness of treatments for menopausal symptoms emphasise vasomotor symptoms, only three vasomotor symptoms ranked high in importance. We believe this method offers some advantages over factor analysis. Sometimes in factor analysis, important items can be omitted when they do not conveniently fit into the factor structure. The sample size required for factor analysis is generally 3–5 participants per question [26]. Because the item generation method contributed 106 questions, the use of factor analysis was precluded as it would have demanded an impractically large sample size. The reduction method was intuitively easy to understand and contributed to the face and content validity.

The variation in number of respondents in the measurement of the psychometric properties deserves some comment. In general, there are fewer participants in the Sexual Domain measurements, primarily because responses were incomplete for women without sexual partners. Cronbach's Alpha and test-retest reliability were undertaken before the randomization visit, during which time some potential participants declined further participation or were excluded from the RCT. Therefore, more respondents were available for the first visit at which Cronbach's Alpha was measured. Fewer completed both first and second visit, required for test-retest reliability and discriminative validity. Subsequently, even fewer entered the RCT and completed it, a requirement to determine evaluative validity. Inclusion in the analysis required a complete data set by domain per participant.

Test-retest reliability, as measured by the intraclass correlation coefficients, was good in all domains except the vasomotor domain, at 0.37. On the other hand, the Pearson correlation coefficient for this domain was 0.85. This discrepancy indicates a systematic change in the Vasomotor domain responses between measurements [26]. Across domains, the measures of domain consistency were satisfactory. Face and content validity appear good.

Health-related quality of life questionnaires may have one or more purposes. If the purpose is discrimination the instrument must distinguish between people at a single point in time. Discriminative instruments require high reliability and high

cross-sectional construct validity [15,27]. If the purpose is evaluation the instrument must measure the magnitude of longitudinal change in an individual or group. Evaluative instruments must be responsive and demonstrate longitudinal construct validity, i.e. appropriate correlations between changes in the new questionnaire and changes in other measures [28]. We required an instrument that was both discriminative and evaluative. The instrument had to determine, at a single point in time, if the groups were equivalent and had to be able to show change over time in quality of life with treatment. We determined discriminative validity at baseline and evaluative validity by comparing the change scores over the course of the trial.

For most domains of the MENQOL questionnaire, baseline discriminative construct validity has been satisfactorily established. The Vasomotor domain correlated well with reported intensity of hot flushes, 0.66, but less well with the Somatic subscale of the Menopausal Index, 0.40. This low correlation could be explained by the fact that the comparison questionnaire, the Somatic subscale, contains 12 items, many of which are clearly not vasomotor in nature. We tested this assumption by determining the internal consistency of the Somatic subscale; the Cronbach's Alpha was 0.57, much lower than the 0.82 found in the Vasomotor domain of the MENQOL questionnaire. The problem could also originate from the low one-month test-retest reliability of the Vasomotor domain when measured with an intraclass correlation coefficient, which requires further clarification.

In testing for evaluative construct validity, the correlation coefficients between the change scores of the MENQOL domains and the comparison instruments are generally lower than the correlation coefficients for discriminative construct validity. This is especially evident for the vasomotor domain. One could conclude that either the vasomotor domain or the Somatic subscale used for validation is not valid in its evaluative role. In theory, this could also occur if there was not enough change in the participants' responses to determine evaluative validity, however, examination of the data allows the last hypothesis to be rejected. The scores for the vasomotor domain and the Somatic subscale improve over time and show moderate variance. We are unable to determine from the existing data whether the Vasomotor domain or the somatic subscale has unsatisfactory evaluative validity.

There are serious doubts about the discriminative and evaluative roles of the global rating question and we have excluded it from the instrument.

With regard to the responsiveness characteristics of this evaluative instrument, the more responsive an instrument, the greater its ability to lower the sample size needed to detect treatment effects in clinical trials. We estimated that the smallest change in the questionnaire that would be important to clinicians was a difference of 1 point in each of the domains, approximately a 15% change. Given this difference, and taking into account the variability in questionnaire score in stable women, sample sizes of less than 30 per group would be required to detect the smallest clinically important mean difference. If one uses a difference of 0.5 as the estimate of the minimal important difference, an estimate supported by other data [17,18,29], maximal sample size requirements would be just over 100 per group.

5. Conclusions

The MENQOL questionnaire is a self-administered instrument which shows potential both for determining differences in quality of life between menopausal women and measuring changes in their quality of life over time. Further research is required to determine test-retest reliability of the Vasomotor domain, and there is uncertainty about the evaluative construct validity of the Vasomotor domain. The single quality of life question has been removed because of inadequate discriminative and evaluative properties. The remaining domains show adequate test-retest reliability, good responsiveness and good to excellent discriminative and evaluative construct validity.

Acknowledgements

Funding for this project was provided by Ciba-Geigy Canada Ltd., Mississauga, Canada. Dr. Ross was an employee of Ciba-Geigy Canada Ltd. during the conduct of the study.

Appendix I

THE MENOPAUSE-SPECIFIC QUALITY OF LIFE QUESTIONNAIRE

Primary Care Research Unit
Department of Family and Community Medicine
Sunnybrook Health Science Centre
University of Toronto

The development of this questionnaire was funded by CIBA-Geigy Canada Ltd., Mississauga, Canada

INSTRUCTIONS

Each of the items in the questionnaire is in the form of the examples below:

		Not at all bothered								Extremely bothered
			0	1	2	3	4	5	6	
NIGHT SWEATS	☐ No	☐ Yes	→	0	1	2	3	4	5	6

Indicate whether or not you have experienced this problem in the ***last month.***

IF YOU ***HAVE NOT*** EXPERIENCED THE PROBLEM:

Mark "No"

NIGHT SWEATS	☐ No	☐ Yes	→	0	1	2	3	4	5	6

Go to the next item.

IF YOU ***HAVE*** EXPERIENCED THE PROBLEM:

Mark "Yes", then circle how ***bothered*** you were by the problem

NIGHT SWEATS	☐ No	☐ Yes	→	0	1	2	3	4	5	6

Go to the next item.

This questionnaire is completely confidential. Your name will not be associated with your responses. However, if for any reason you do not wish to complete an item, please leave it and go on to the next one.

Primary Care Research Unit
Department of Family and Community Medicine
Sunnybrook Health Science Centre
University of Toronto

The Menopause-Specific Quality of Life Questionnaire

For each of the following items, indicate whether you have experienced the problem in the **PAST MONTH.** If you have, rate how much you have been ***bothered*** by the problem.

		Not at all bothered		0	1	2	3	4	5	6	Extremely bothered
1.	HOT FLUSHES OR FLASHES	☐ No	☐ Yes	→	0	1	2	3	4	5	6
2.	NIGHT SWEATS	☐ No	☐ Yes	→	0	1	2	3	4	5	6
3.	SWEATING	☐ No	☐ Yes	→	0	1	2	3	4	5	6
4.	BEING DISSATISFIED WITH MY PERSONAL LIFE	☐ No	☐ Yes	→	0	1	2	3	4	5	6
5.	FEELING ANXIOUS OR NERVOUS	☐ No	☐ Yes	→	0	1	2	3	4	5	6
6.	EXPERIENCING POOR MEMORY	☐ No	☐ Yes	→	0	1	2	3	4	5	6
7.	ACCOMPLISHING LESS THAN I USED TO	☐ No	☐ Yes	→	0	1	2	3	4	5	6
8.	FEELING DEPRESSED, DOWN OR BLUE	☐ No	☐ Yes	→	0	1	2	3	4	5	6
9.	BEING IMPATIENT WITH OTHER PEOPLE	☐ No	☐ Yes	→	0	1	2	3	4	5	6
10.	FEELINGS OF WANTING TO BE ALONE	☐ No	☐ Yes	→	0	1	2	3	4	5	6
11.	FLATULENCE (WIND) OR GAS PAINS	☐ No	☐ Yes	→	0	1	2	3	4	5	6

Primary Care Research Unit
Department of Family and Community Medicine
Sunnybrook Health Science Centre
University of Toronto

The Menopause-Specific Quality of Life Questionnaire

	Not at all bothered								Extremely bothered	
		0	1	2	3	4	5	6		
12. ACHING IN MUSCLES AND JOINTS	☐ No	☐ Yes	–	0	1	2	3	4	5	6
13. FEELING TIRED OR WORN OUT	☐ No	☐ Yes	→	0	1	2	3	4	5	6
14. DIFFICULTY SLEEPING	☐ No	☐ Yes	→	0	1	2	3	4	5	6
15. ACHES IN BACK OF NECK OR HEAD	☐ No	☐ Yes	→	0	1	2	3	4	5	6
16. DECREASE IN PHYSICAL STRENGTH	☐ No	☐ Yes	→	0	1	2	3	4	5	6
17. DECREASE IN STAMINA	☐ No	☐ Yes	→	0	1	2	3	4	5	6
18. FEELING A LACK OF ENERGY	☐ No	☐ Yes	→	0	1	2	3	4	5	6
19. DRYING SKIN	☐ No	☐ Yes	→	0	1	2	3	4	5	6
20. WEIGHT GAIN	☐ No	☐ Yes	→	0	1	2	3	4	5	6
21. INCREASED FACIAL HAIR	☐ No	☐ Yes	→	0	1	2	3	4	5	6
22. CHANGES IN APPEARANCE, TEXTURE OR TONE OF YOUR SKIN	☐ Yes	☐ No	→	0	1	2	3	4	5	6
23. FEELING BLOATED	☐ No	☐ Yes	→	0	1	2	3	4	5	6

Primary Care Research Un
Department of Family and Community Medicine
Sunnybrook Health Science Centre
University of Toronto

The Menopause-Specific Quality of Life Questionnaire

	Not at all bothered		0	1	2	3	4	5	6	Extremely bothered
24. LOW BACKACHE	☐ No	☐ Yes	→	0	1	2	3	4	5	6
25. FREQUENT URINATION	☐ No	☐ Yes	→	0	1	2	3	4	5	6
26. INVOLUNTARY URINATION WHEN LAUGHING OR COUGHING	☐ No	☐ Yes	→	0	1	2	3	4	5	6
27. CHANGE IN YOUR SEXUAL DESIRE	☐ No	☐ Yes	→	0	1	2	3	4	5	6
28. VAGINAL DRYNESS DURING INTERCOURSE	☐ Yes	☐ No	→	0	1	2	3	4	5	6
29. AVOIDING INTIMACY	☐ No	☐ Yes	→	0	1	2	3	4	5	6

Primary Care Research Unit
Department of Family and Community Medicine
Sunnybrook Health Science Centre
University of Toronto

Appendix A: The Menopause-Specific Quality of Life questionnaire, and instructions for use and storing

Instructions for use and scoring of the menopause-specific quality of life questionnaire

Use:

(1) This questionnaire is designed to be self-administered, either in person or by mail.
(2) Specific instructions for the subject are part of the instrument.
(3) The questionnaire requires, on average, 7 min to complete, with a range of 5–15 min.
(4) The questionnaire is appropriate for English-speaking subjects.
(5) The psychometric properties are based on data collection periods one month apart.
(6) The questionnaire was developed using data from women who: (a) are between the ages of 47 and 62 years, (b) are 2–7 years post-menopause, (c) have an intact uterus, (d) have not been on hormone replacement therapy in the past 6 months.

Scoring:

(1) (a) Each domain is scored separately. (b) The scale contains four domains: (i) Vasomotor - Items 1, 2 & 3 (ii) Psychosocial - Items 4–10 (iii) Physical - Items 11–26 (iv) Sexual - Items 27–29. (c) There is no overall score that can be obtained from this questionnaire since the relative contribution of each domain to an overall score is unknown.
(2) For analyses, convert the item scores to a score ranging from 1 to 8 in the following manner: (1) The subject responded 'NO', she did not experience the problem; (2) The subject experienced the problem and rated it as '0' on the bothered scale; (3) The subject experienced the problem and rated it as '1' on the bothered scale; (4) Rated as '2'; (5) Rated as '3'; (6) Rated as '4'; (7) Rated as '5'; (8) Rated as '6'.
(3) Since the domain subscales are not comprised of an equivalent number of items, the mean of the subscale is used as the overall subscale score. Each domain score ranges from 1 to 8.

Appendix B: Scoring method for the validation questionnaires

B.1. Neugarten and Kraines menopause questionnaire (1986)

B.1.1. Scoring

If a symptom is present, the item is given a score of '1'.

If a symptom is not present, the item is given a score of '0'.

If missing, the item is given a score of '0'.

B.1.2. Domains

The Somatic subscale is the sum of items 1–12

The Psychosomatic subscale is the sum of items 13–17.

The Psychologic subscale is the sum of items 18–28.

B.2. General Well Being Schedule (Dupuy, 1977)

B.2.1. Scoring

Higher scores are more positive. The data were entered as seen on questionnaire. Items 1, 3, 6, 7, 9, 11, 13, 15 and 16 are reversed so that higher number is more positive. Missing values are set to missing and not included in the analysis.

B.2.2. Domain

GWB = Sum of items 1–18 minus 14 (range 0–110).

B.3. Life Satisfaction Index (Neugarten and Havighurst, 1961)

B.3.1. Scoring

For items 1, 2, 4, 6, 8, 9, 11, 12, 13, 15, 16 and 19:

Agreement with the statements is given a score of '2'.

Disagreement with the statement is given a score of '0'.

Not sure and missing values are given a score of '1'.

For items 3, 5, 7, 10, 14, 17, 18 and 20:

Agreement with the statement is given a score of '0'.

Disagreement with the statement is given a score of '2'.

Not sure and missing values are given a score of '1'.

B.3.2. Dimensions

LSI = sum of items 1–20

B.4. Channon and Ballinger (1986) Sexuality questionnaire

B.4.1. Scoring

Items 1–6:

If symptom is present score '1'.

If symptom is not present, or not applicable, score '0'.

Missing values are set to missing.

Items 8, 10, 11 and 12 are scored from 1 to 3, where 1 is 'important' or 'increased'. Missing values are set to missing.

B.4.2. Dimensions

Vaginal symptom score: the sum of items 1–6.

Libido score: the sum of items 8, 10, 11 and 12, minus 3.

References

[1] Fletcher AE, Bulpitt CJ. Measurement of quality of life in clinical trials therapy. Cardiology 1987; 75 Suppl 1: 41–52.

[2] Schipper H. Why measure quality of life? Can Med Assoc J 1983; 128: 1367–1370.

[3] Arora R, Sager J, Butler RN. Therapeutic goals: quality and quantity of life. Cardiol Clin 1986; 4: 305–312.

[4] Editorial. Quality of life and clinical trials. Lancet 1995; 346: 1–2.

[5] Dupuy H. The general well-being schedule, as described in McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. 1987, Oxford University Press, New York.

[6] Channon LD, Ballinger SE. Some aspects of sexuality and vaginal symptoms during menopause and their relation to anxiety and depression. Br J Med Psychol 1986; 59: 173–180.

[7] Neugarten BL, Havighurst RJ and Tobin SS. The measurement of life satisfaction. J Gerontol 1961; 16: 134–143.

[8] Kupperman HS, Blatt MHG, Wiesbaden H, Filler W. Comparative clinical evaluation of estrogenic preparations by menopausal and amennorheal indices. J Clin Endocrin 1953; 13: 688–703.

[9] Neugarten BL, Kraines RJ. "Menopausal symptoms" in women of various ages. Psychosom Med 1965; 28: 266–273.

[10] Greene JG. A factor analytic study of climacteric symptoms. J Psychosom Res 1976; 20: 425–430.

[11] Likert RA. A technique for the development of attitude scales. Educ Psycholog Measurement 1952; 12: 313–315.

[12] Guyatt G, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. Can Med Assoc J 1986; 134: 889–895.

[13] Hilditch JR, Lewis JE, Ross AH, et al. A comparison of the effects of oral conjugated equine estrogen and transdermal estradiol-17β combined with an oral progestin on the quality of life in postmenopausal women. Maturitas 1996; 24: 177–184.

[14] Kramer MS, Feinstein AR. Clinical biostatistics LIV. the biostatistics of concordance. Clin Pharmacol 1981; 26: 111–123.

[15] Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Controlled Clin Trials 1991; 12: 142S–158S.

[16] Aday LA. Designing and conducting health surveys. San Francisco, Jossey-Bass, 1989, pp. 45–47.

[17] Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epid 1994; 47: 81–87.

[18] Jaeschke R, Guyatt G, Keller J, Singer J. Measurement of health status. Ascertaining the meaning of a change in quality-of-life questionnaire score. Controlled Clin Trials 1989; 10: 407–415.

[19] Greene JG. Guide to the Greene Climacteric Scale, June, 1991. University of Glasgow.

[20] Wiklund I, Holst J, Karlberg J, et al. A new methodological approach to the evaluation of quality of life in postmenopausal women. Maturitas 1992; 14: 211–224.

[21] Limouzin-Lamothe M, Mairon N, Joyce CRB, Le Gal M. Quality of life after the menopause: Influence of hormonal replacement therapy. Am J Obstet Gynecol 1994; 170: 618–624.

[22] Bombardier C, Tugwell P. Methodological considerations in functional assessment. J Rheum 1987; 14 Suppl 15: 6–14.

[23] Guyatt G, Mitchell A, Irvine EJ, et al. A new measure of health status in clinical trials in inflammatory bowel disease. Gastroenterology 1989; 96: 804–810.

[24] Guyatt GH, Norgradi S, Halcrow S, Singer J, Sullivan MJ. Development and testing of a new measure of health status for clinical trials in heart failure. J Gen Int Med 1989; 4: 101–107.

[25] Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. Thorax 1987; 42: 773–778.

[26] Kline P. (1986) A handbook of test construction. Introduction to psychometric design. London: Methuen, 1986.

[27] Kirshner B, Guyatt GH. A methodologic framework for assessing health indices. J Chron Dis 1985; 38: 27–36.

[28] Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: What are the necessary measurement properties? J Clin Epidemiol 1992; 45: 1341–1345.

[29] Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Basic Sciences Review. Annals Int Med 1993; 70: 225–230.

Appendix 5

Report of the NSABP Quality of Life Workshop

National Surgical Adjuvant Breast and Bowel Project
Quality of Life Workshop (QOL) Report
March 6-7, 1999
Pittsburgh, PA

Participants

Jennifer Aikin, RN, MSN, AOCN/NSABP Operations Center
John Bryant, Ph.D./NSABP Biostatistical Center
Richard Day, Ph.D./NSABP Biostatistical Center
Patricia Ganz, M.D./UCLA
Charles Geyer, M.D./Joe Arrington Cancer Research and Treatment Center
J. Wendall Goodwin, M.D./CCOP Ozarks Regional
Jacek, Kopec, M.D., Ph.D./University of Toronto
Carol Moinpour, Ph.D./SWOG
Joyce Mull, MPM/NSABP Operations Center
Judy Negley, BS/NSABP Operations Center
Edward Romond, M.D./University of Kentucky
Sandra Swain, M.D./Bethesda, MD
Claudette Varicchio, DSN, RN/National Cancer Institute
Samuel Wieand, Ph.D./NSABP Biostatistical Center

I. Introduction (Richard Day, Ph.D.)

A. *Primary Goals of the Workshop*:

1. Review QOL experience from prevention (P-1, P-2) and treatment trials (B-23, C-06, B-30);

2. Identify successful and less successful aspects of this experience and key problem areas that need to be addressed in future research;

3. Discuss the feasibility of using companion studies and other strategies to generate new QOL research ideas;

4. Concisely state the essential role of QOL studies in accomplishing NSABP's mission;

5. Provide recommendations for future directions of QOL research by NSABP.

B. *Products of Workshop*:

1. Report on the QOL Workshop that can be distributed to the membership in order to raise the visibility of QOL and summarize current goals and activities for the NSABP Treatment and Prevention QOL Committees; and,

2. Plan of action for inclusion in the new Treatment Grant Application.

II. **Progress Report: Prevention QOL Committee** (Patricia Ganz, MD)

A. *P-1 Study*

1. Data from the QOL component of the P-1 study is summarized in two papers, one which reports baseline data (Ganz et al., *JNCI*,1995, 87 1372-82) and a second which summarizes the follow-up data (Day et al., *JCO*, submitted for publication in 1999).

2. Key Findings of P-1 QOL Study

- No excess depression in the TAM group;
- No difference in overall physical or emotional functioning on the SF-36;
- Participants in the TAM group reported a greater frequency of vasomotor and gynecological symptoms;
- No excess weight gain was reported in the TAM group; and,
- Participants in the TAM group reported a greater frequency of problems with sexual functioning, although overall rates of sexual functioning remained similar.

3. Problems in P-1 Study: The most serious problem for P-1 QOL was missing data, primarily for women who went off treatment. This occurred for several reasons:

- it was difficult for research staff in the clinical centers to administer the QOL questionnaire to women that did not present for follow-up visits;

- the NSABP failed to explicitly request QOL data on women who were off treatment; and,

- the early closure of the study.

B. *P-2 (STAR) QOL Study*

1. A revised version of the complete QOL questionnaire (CES-D, symptom checklist, SF-36, and sexual functioning scale) will be implemented in a sub-sample of women through the participating CCOP centers;

2. All participants in all centers will receive a revised symptom checklist at baseline and all follow-up examinations;

3. Plans are underway to target QOL in lower socioeconomic status and minority populations in the P-2 study.

III. **Progress Report: Treatment QOL Committee** (Jacek Kopec, M.D, Ph.D.)

A. *Primary Aims of NSABP Treatment QOL Studies*

1. B-23
   - Compare QOL in women with primary breast cancer treated with AC and CMF; and,
   - Study the effects of local treatment and TAM on QOL in women with primary breast cancer.

2. C-06
   - Compare QOL in patients with resected carcinoma of the colon who are treated with 5-FU+LV versus UFT+LV

3. B-30
   - Compare QOL in women with breast cancer who are receiving ATC, AC→T, or AT; and,
   - Examine differences in amenorrhea in premenopausal women in each treatment arm, and its relationship to symptoms, QOL, DFS and S.

B. *Instruments used in NSABP Treatment QOL Studies*

1. B-23:
   - Functional Assessment of Cancer Therapy – Breast (FACT-B);
   - Symptom checklist
   - SF-36 vitality scale
   - QOL linear rating scale
   - Additional questions (general health, functional limitations, return to normal activity).

2. C-06
   - Symptom distress scale
   - Symptom checklist
   - ECOG burden of care
   - FACT-C (colon)
   - SF-36 vitality scale
   - QOL linear rating scale
   - Return to normal activities

3. B-30
   - FACT-B
   - Symptom checklist
   - SF-36 vitality scale
   - QOL linear rating scale
   - Menopause and menstrual history questionnaire

C. *Accrual to NSABP Treatment QOL Studies*

1. B-23: n=169 (closed 12/31/98)

2. C-06: n=1462 (closed 3/31/99)

3. B-30: approved but not open

IV. **What is the Role of QOL Studies in Accomplishing NSABP's Mission?** (Samuel Wieand, Ph.D.)

A. Primary question is: What general QOL information is required for NSABP research?

B. There is a general pressure to the reduce the amount of information collected in clinical trials and we must operate within that framework. For example, trials with toxic treatments that are primarily looking for a survival benefit usually do not require a QOL component. The primary role of QOL is in equivalency trials where QOL may determine the choice between outcomes.

C. Potential downsides of QOL studies that need to be evaluated in the future:

1. Forms design and translation;
2. Impact on nurses and Biostatistical Center staff;
3. Impact on data managers in the field;
4. Impact on patient

D. In designing future QOL studies:

1. we must have specific hypotheses;

2. we need to take advantage of methodological devices like two-stage rules that may allow us to complete accrual to a QOL component earlier than the treatment component; and

3. in assessing the necessity of a QOL study, we need to ask not only "what do we expect to learn?"; but "what would we do differently if we learned it?"

V. Topics Discussed Following the Above Presentations

- The value of planning long-term morbidity studies in addition to studies with short-term QOL outcomes: e.g., late cardiac effects in B-31 or QOL following recurrence. Can the NSABP database be effectively utilized to study the impact of morbidity on long-term QOL?

- Utility assessments (standard gambles and time trade off studies) of specific toxicities or treatments: e.g., alopecia or treatment choice B-29.

- The necessity of involving QOL investigators in the protocol development process at the earliest possible point.

VI. Current Issues: The Process for Selecting QOL Studies

(Jennifer Aikin, RN, MSN, AOCN)

A. *Steps in the Protocol Development Process:*

1. Breast or colon protocol design committee generates research idea

2. Protocol officer and protocol chairman draft concept

3. Protocol team begins writing (protocol chairman, protocol officer, statistician, nurse, protocol specialist)

4. External Review (committee members, pharmaceutical company, headquarters staff)

5. Incorporate comments/ forms development

6. NCI submission

B. *Other Influencing Factors*

1. Pharmaceutical Company Issues (scientific agenda, type of trial [e.g. registration trial], contract issues)

2. Regulatory Issues (IND submission, adverse event reporting guidelines)

3. Competing priorities for staff time (other protocols in development, grant writing, meeting planning)

C. *Key reasons for including QOL in current treatment protocols:*

1. Equivalency trials (B-23, C-06);

2. Important clinical question with impact on QOL (B-30);

3. Pharmaceutical company request;

4. NCI cancer control credits for CCOPs.

D. *Issues for additional discussion:*

1. At what point in the protocol development process should the QOL Committee get involved?

2. What criteria are most important when selecting QOL questions to include in protocols?

3. What is the most effective way to write the QOL section of a protocol so that it does not slow down the approval process?

E. *Group Discussion*

1. It is important that the QOL Committee be involved at the concept phase of the protocol development process. Two possible strategies: forward all concepts to the chair of the Treatment QOL Committee at the time of NCI submission and/or invite the chair of the Treatment QOL Committee to protocol design committee meetings.

2. Quarterly QOL Committee conference calls may enhance the Committee's ability to make decisions about and design QOL studies in a timely fashion.

VII. **Current Issues: Relationship between Toxicity Reporting and QOL** (Richard Day, Ph.D.)

A. *Statement of Problem*: Why include a symptom checklist in the QOL assessment when we are already doing extensive toxicity reporting?

B. *Comparison of Methodologies*

1. Toxicity reporting forms and QOL symptom checklists have different content: i.e., objectively assessed signs vs. subjectively assessed symptoms;

2. Toxicity reporting forms and QOL symptom checklists answer different questions – i.e., presence or absence of a toxicity at a certain level vs. extent to which symptoms "bother" the participant;

3. Toxicity reporting forms and QOL symptom checklists reflect the point of view of different assessors – i.e., medical staff vs. participant.

4. Comprehensiveness of the assessment (potential bias) – i.e., the data for toxicity report forms are not collected in a uniform manner, whereas symptom checklists are always collected in the same way.

C. *Restatement of the problem for discussion*: How should QOL assessments and toxicity reporting be used to supplement each other?:

VIII. **Choosing and Streamlining QOL Instruments** (Jacek Kopec, M.D., Ph.D.)

A. *Objectives of presentation:*

1. Explain rationale for the selection of instruments in NSABP Treatment QOL studies; and,

2. Discuss opportunities for streamlining QOL data collection in future trials.

B. *Rationale for using the FACT in treatment trials:*

1. Comprehensive, multidimensional measure with a solid conceptual basis;

2. Thoroughly tested, valid, reliable and responsive to change; and,

3. Widely used in the US, including other cooperative groups (ECOG, RTOG).

C. *Rationale for additional measures in treatment QOL studies.* The FACT does not include:

1. Symptoms associated with specific treatments;

2. General health and disability questions; and,

3. Overall QOL rating scale (0 - 10).

D. *Opportunities and Challenges for Streamlining QOL Data Collection.* Critical topics:

1. Criteria for the selection of trials;

2. Careful formulation of study aims;

3. A focus on concepts to be measured and scale length in the selection of instruments;

4. Choosing the number of assessments;

5. Sample size considerations;

6. Questionnaire design and administration; and

7. Data transmission and entry.

IX. **Strategies for Reducing Missing Data** (Richard Day, Ph.D.)

A. *Key problems with QOL studies leading to missing data:*

1. QOL questionnaires are usually the only self-administered forms completed by research subjects; and

2. QOL questionnaires generally cannot be filled out retrospectively.

B. *Current experience from prevention and treatment studies*: if we can get the QOL questionnaire into the research subject's hands, they will fill it out completely.

C. *Primary cause of missing data*: research subjects who discontinue their assigned treatment and fail to appear for treatment or follow-up appointments.

D. *Potential solutions to missing data for discussion:*

1. Missing data forms (e.g., C-06): It may not be possible to prevent missing data, but the reasons for the missing data can be documented in order to assess the potential biases in the data set.

2. Real time monitoring: Missing data may be prevented by monitoring the data collection process in the clinical centers and by providing reminders to data managers concerning upcoming patient appointments.

3. Telephone and mail contacts: Patients who miss scheduled data collection points may be contacted by mail or by telephone by a data manager in order to collect missing data. This would have to be done promptly since QOL data cannot be collected retrospectively.

4. Fewer assessments carried out at follow-up examinations that are not related to treatment: Simplify the data collection procedure and only collect data at follow-up examinations that a patient will keep whether or not they are on treatment (e.g., baseline, 6 months, etc.).

5. Accept substantial levels of missing data and attempt to use statistical methods of imputation wherever possible: Imputation methods are complex and should simply be used as a supplement to items 1-4 above.

X. Responses and Observations from Consultants

A. *Claudette Varicchio, DSN, RN* – National Cancer Institute

1. It is important to educate PIs in the centers with regard to the importance of QOL studies.

2. The NCI looks for the following while assessing QOL studies as part of treatment protocols:

   - Independent sample size calculations to show a proper use of resources;
   - A clearly stated QOL question that requires an independent study instrument;
   - Does the choice of instruments fit the question?;
   - Is the QOL study integrated into the overall study protocol or is it a companion study?;
   - Is the QOL endpoint clearly measurable and clinically meaningful?;
   - Is there a compelling rationale for all the data points or is the data being collected simply because the participant is in the office?;
   - Is there a system of quality control for data collection?;
   - Dose the sequence of instrument administration avoid potential biases?;
   - Are validation studies built into protocols using new or untested instruments?; and
   - Does the study measure all relevant domains of QOL?.

3. Mailing QOL questionnaires to participants who are off treatment should be reasonably successful if a prior personal relationship has been established by the research staff at a center.

4. Delinquency reporting is an important tool in preventing missing data. Center's staff are proud of low delinquency levels and will try hard not to miss interviews.

B. *Carol Moinpour, Ph.D.*, Southwest Oncology Group

1. SWOG has not used a single QOL battery like NSABP, but tends to use a different battery for each study.

2. SWOG has had a positive experience with the use of videotapes as training devices.

3. Reminders sent out to clinical staff are valuable in avoiding missing data.

4. Quality control is absolutely necessary in QOL studies.

XI. Conclusions from Presentations and Discussions on March 6

A. Given the current stage of development of NSABP Treatment QOL studies, instruments should investigate a broad range of QOL domains. Initially, NSABP needs to build up a database of information on the QOL effects of standard clinical treatments. This can be done relatively quickly because of the large numbers of patients in NSABP trials. Subsequently, QOL assessments may become more focused and specific based on the background knowledge contained in the NSABP database. This is a strategy similar to that used in the evolution of QOL work from the P-1 to the P-2 trial.

B. There is no compelling evidence that patients in NSABP trials consider QOL research a burden. In fact, most reports suggest that patients are happy to complete QOL instruments. The single most important cause of missing data in NSABP QOL studies is not patient refusal or staff objections, but simply the inability to get the instrument into the patient's hands because they are off therapy and do not present for their scheduled follow-up visit.

C. Greater QOL training and outreach is required with doctors, nurses and research staff in the collaborating clinical centers. This outreach and training should be coupled with greater emphasis on the reduction of missing data through direct mailing and telephone calls to patients off treatment by the research staff. Instead of using the term "Quality of Life", it may be helpful to talk about "disease focused" versus "patient focused" measures of outcome.

D. A greater integration of the QOL Committees is required into the early development of treatment protocols. One possible solution to be implemented will be quarterly conference calls to review QOL study progress and potential new treatment protocols.

E. Increased consieration needs to be given to clinical epidemiological studies that use the NSABP database to assess long-term morbidity, QOL, and survival outcomes.

XII. Generating New QOL Research Ideas

A. *Companion/Ancillary Studies*

1. Characteristics:
   - separate from protocol
   - may be done at select institutions
   - may look at different kinds of outcomes (directly- or indirectly-related to protocol)
   - ideas may be generated by individual researchers or others (e.g. clinical trials nurse committee)

2. Examples of Companion/Ancillary Studies
   - NSABP ancillary studies (eye study [P-1E], bone mineral density [P-1B], tumor markers [B-27.1/B-27.2], B-23 QOL)
   - Other outcomes (psychosocial, pharmacoeconomic)
   - Interventions that impact outcomes (symptom management interventions, psychosocial interventions)
   - "Quality of Life Among Spouses of Breast Cancer Patients"- a proposal by Elizabeth Maunsell, Ph.D.

3. Questions to Consider
   - What types of companion studies should the NSABP support?
   - What is the process for reviewing proposals for companion studies?
   - What criteria should be considered when reviewing such studies?
   - What resources are available to fund companion studies?

4. Discussion
   - Studies with patient-focused outcomes are reviewed by the Department of Cancer Prevention (DCP), not the Cancer Therapy Evaluation Program (CTEP). Studies that focus on symptom management, prevention of toxicities and quality of life fall into this category and earn cancer control credits.
   - Protocols that are submitted should be "flagged" for both treatment and cancer control review.
   - Focus on CCOP institutions because they earn funding credits for participation in cancer control studies.
   - Need to establish a mechanism for other interested (non-CCOP) institutions to participate.
   - Dr. Varicchio noted that while intervention studies will be supported by the DCPC, descriptive studies are generally not. Funding for descriptive studies may come from other sources, such as R-01 applications. She suggested that these applications designate funds to go to the cooperative group for statistical center analysis, reimbursement for sites, and operations center support for grant submission. In this way a companion study is mutually beneficial for the cooperative group and the investigator.
   - The Treatment QOL Committee will review proposals submitted by individual investigators. If the research question has merit, the Committee will support the proposal in discussions with the NSABP leadership. Feasibility and funding issues will be determined in conjunction with the NSABP leadership.

XIII. **Future Directions**

A. *Priorities for Future QOL Studies*

1. B-32 (Sentinel Lymph Node Biopsy Protocol)
   - This protocol has been submitted to the NCI for final review. In addition to objective measures of functional ability, those present recommended that subjective data be obtained via a patient self-report of mobility following surgery
   - Dr. Ganz has a validated instrument that captures subjective information about movement, numbness and tingling.
   - Judy Negley will approach the B-32 protocol team about revising the protocol prior to its activation.

2. C-07 (5-FU/LV ± Oxaliplatin for Colon Cancer)
   - This protocol was recently distributed for external review. Since there is concern about neurotoxicity associated with oxaliplatin, the protocol includes an extensive neurotoxicity assessment checklist to be completed by the physician at each treatment and follow-up visit.
   - Those in attendance were concerned about the burden the checklist would be for investigators, and suggested that patient self-report data would be valuable.
   - Dr. David Cella and GOG developed a brief neurotoxicity scale that has been validated.
   - Jennifer Aikin will obtain a copy of the scale from Dr. Cella and speak with the protocol team about replacing the neurotoxicity checklist with the questionnaire.

3. B-31 (AC → Taxol ± Herceptin)
   - There was some discussion about looking at long-term cardiac toxicity for this trial that involves Adriamycin and Herceptin. Dr. Ganz offered to share a form used by SWOG to look at late cardiac effects.

4. Clodronate Study
   - A concept has been submitted to the NCI for a study that would randomize patients to clodronate or placebo as adjuvant therapy (with or without chemotherapy and/or tamoxifen). The endpoints of the study are 1) incidence of skeletal metastases, and 2) disease-free survival. The protocol design committee is interested in a QOL study that looks at symptoms associated with skeletal metastasis, and the resulting impact on QOL. This concept has been forwarded to Jacek Kopec, MD, Ph.D. for consideration by the Treatment QOL Committee.

5. QOL in Advanced Disease Trials
   - Since the NSABP will initiate a number of advanced-disease trials

in the future, this could be the ideal setting to evaluate the impact of specific chemotherapeutic agents on QOL. When comprehensive QOL data are available, then adjuvant QOL studies may focus on more targeted QOL issues.

B. *Organizational Issues*

1. Committee policies and procedures
   - Other cooperative groups have QOL Coordinators for each study that includes a QOL component. This Coordinator is involved in the design of the study, answers questions from the membership about QOL issues, and is then first author on any publication related to the QOL component of the study. While the Clinical Coordinating Section fields all clinical questions related to NSABP protocols, the group acknowledged the need to consider these issues.
   - A committee member who invests a great deal of time in designing a QOL study and analyzing the data should be recognized by being involved in the publication of the study.
   - These issues need to be discussed in greater detail during a Treatment QOL conference call with input from the NSABP leadership. A formal policy statement would be helpful.

2. Committee Communications with the NSABP Membership
   - With the closure of C-06 and B-23, there is a need to remind the membership of the importance of continuing to collect and submit QOL questionnaires for these studies. A memo will be drafted related to this.
   - With the initiation of B-30, there was discussion about the need to inform the NSABP membership about the importance of B-30 QOL to the overall success of this trial. A memo will also be drafted related to B-30.
   - The QOL Committees could also communicate with the NSABP membership via the web site and a standing newsletter column.

s:\sherick\aikin\report.wpd

Appendix 6

Draft Missing Data Form

Form QOLMD 3/17/98

Missing Data Form
Quality of Life Questionnaire
NSABP Protocol C-06

Instructions: Please submit this form to the NSABP Biostatistical Center whenever a scheduled Quality of Life Questionnaire (Version A or Version B) is not filled out by or not given to the patient. No Missing Data Form is required for partially completed questionnaires. (Partially completed questionnaires should be submitted to the NSABP Biostatistical Center per protocol specifications.)

Patient Name _____

Study Number (1-9) 2 6 _ _ _ _ _ _ _

First 3 Letters of Patient's Last Name (10-12) _ _ _

Date Form Completed (13-20) _ _ Mo _ _ Day _ _ Yr

Institution Name _____

The QOL Questionnaire was not filled out (or not given) at:
(Enter code in box at right.) (21-22) _ _

UFT + LV	5-FU + LV
21 - Baseline	11 - Baseline
22 - Day 1, Cycle 2	12 - Day 1, cycle 2
23 - Day 1, Cycle 3	13 - Day 1, cycle 3
24 - Day 1, Cycle 4	14 - 1-Year Follow-Up
25 - Day 1, Cycle 5	
26 - 1-Year Follow-Up	

Which version of the QOL Questionnaire was not completed?
(Enter code in box at right.) (23) _

1 - Version A
2 - Version B
3 - Both Version A and Version B

Patient was approached with QOL Questionnaire, but refused because:
(Check all items that apply, and enter sum of codes in boxes at right.) (24-25) _ _ **Sum of Codes**

- ☐ 0 - This question does not apply
- ☐ 1 - Patient felt too ill
- ☐ 2 - Patient lost interest or had too little time
- ☐ 4 - Patient dislikes or complains of burden
- ☐ 8 - Other reasons, specify _____

Patient was not approached with QOL Questionnaire because:
(Check all items that apply, and enter sum of codes in boxes at right.) (26-27) _ _ **Sum of Codes**

- ☐ 0 - This question does not apply
- ☐ 1 - M.D. thought patient felt too ill
- ☐ 2 - M.D. thought patient was emotionally unstable
- ☐ 4 - Staff oversight or due to understaffing
- ☐ 8 - Patient failed to appear for scheduled appointment
- ☐ 16 - Other reasons, specify _____

Appendix 7

Day R., Ganz P., Costantino J, Cronin W., Wickerham L and Fisher B., Health-related quality of life and tamoxifen in breast cancer prevention: A report from the National Surgical Adjuvant Breast and Bowel Project. J Clin Oncol 17, 1999, 2659-2669

Health-Related Quality of Life and Tamoxifen in Breast Cancer Prevention: A Report From the National Surgical Adjuvant Breast and Bowel Project P-1 Study

By Richard Day, Patricia A. Ganz, Joseph P. Costantino, Walter M. Cronin, D. Lawrence Wickerham, and Bernard Fisher

Purpose: **This is the initial report from the health-related quality of life (HRQL) component of the National Surgical Adjuvant Breast and Bowel Project Breast Cancer Prevention Trial. This report provides an overview of HRQL findings, comparing tamoxifen and placebo groups, and advice to clinicians counseling women about the use of tamoxifen in a prevention setting.**

Patients and Methods: **This report covers the baseline and the first 36 months of follow-up data on 11,064 women recruited over the first 24 months of the study. Findings are presented from the Center for Epidemiological Studies-Depression Scale (CES-D), the Medical Outcomes Study 36-Item Short Form Health Status Survey (MOS SF-36) and sexual functioning scale, and a symptom checklist.**

Results: **No differences were found between placebo and tamoxifen groups for the proportion of participants scoring above a clinically significant level on the CES-D. No differences were found between groups for the MOS SF-36 summary physical and mental scores. The mean number of symptoms reported was consistently higher in the tamoxifen group and was associated with vasomotor and gynecologic symptoms. Significant increases were found in the proportion of women on tamoxifen reporting problems of sexual functioning at a definite or serious level, although overall rates of sexual activity remained similar.**

Conclusion: **Women need to be informed of the increased frequency of vasomotor and gynecologic symptoms and problems of sexual functioning associated with tamoxifen use. Weight gain and depression, two clinical problems anecdotally associated with tamoxifen treatment, were not increased in frequency in this trial in healthy women, which is good news that also needs to be communicated.**

J Clin Oncol 17:2659-2669.

THIS IS THE INITIAL report of the findings from the health-related quality of life (HRQL) component of the National Surgical Adjuvant Breast and Bowel Project (NSABP) Breast Cancer Prevention Trial (P-1), a multicenter, double-blinded, placebo-controlled clinical trial. The purpose of this report is to provide a concise overview of the P-1 HRQL findings and an assessment of the effects of tamoxifen, when used as a preventative agent, on self-reported symptoms and everyday physical, emotional, and social functioning. Recommendations have been provided that may be helpful to physicians involved in counseling women considering the use of tamoxifen in the setting of prevention.

The primary objective of the P-1 study was to evaluate whether 5 years of tamoxifen therapy would reduce the incidence of invasive breast cancer in women at an increased risk for the disease. Secondary objectives were to assess the incidence of ischemic heart disease, bone fractures, and other events, such as depression, that might be associated with the use of tamoxifen. Eligible participants were randomized either to 20 mg daily of tamoxifen or to a placebo for a planned 5 years.

Detailed descriptions of the rationale, planning, and design of the of the Breast Cancer Prevention Trial and the HRQL component of the P-1 study, as well as specific instruments, have been provided in separate reports.[1-3]

PATIENTS AND METHODS

Participant Cohort and HRQL Data

This report covers the baseline HRQL examination and the first 36 months of follow-up data on 11,064 women recruited over the first 24 months (June 1, 1992, to May 31, 1994) of the study. This cohort of women represents 82.6% of the total P-1 accrual (n = 13,388). Restrictions were imposed on the initial HRQL report for two reasons. First, by limiting our attention to this cohort of women, we avoided the potential bias created by events beginning in March 1994,[4,5] which resulted in a suspension of accrual to the P-1 study. Second, a focus on the first 36 months of data collection permitted improved control over types of missing HRQL data because all 11,064 participants should have completed the eight scheduled examinations before the disclosure of the results of the trial in the spring of 1998.

From the National Surgical Adjuvant Breast and Bowel Project (NSABP) Operations and Biostatistical Centers, Pittsburgh, PA, and Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, CA.

Submitted December 7, 1998; accepted April 22, 1999.

Supported by public health service grants from the National Cancer Institute (NCI-U10-CA-37377/69974) and a career development award from the Department of Defense (DAMD17-97-1-7058).

Address reprint requests to Richard Day, PhD, Department of Biostatistics, Graduate School of Public Health, 130 DeSoto St, University of Pittsburgh, Pittsburgh, PA 15261; email rdfac@vms.cis.pitt.edu.

0732-183X/99/1709-2659

Instruments

The 104-item P-1 HRQL Questionnaire[3] was composed of the Center for Epidemiological Studies–Depression Scale (CES-D, 20 items), the Medical Outcomes Study (MOS) 36-Item Short Form Health Status Survey (SF-36, 36 items), the MOS sexual functioning scale (five items), and a symptom checklist (SCL, 43 items). The questionnaire was scheduled to be administered to all participants before randomization (baseline), at 3 months, at each succeeding 6-month examination for the planned 5 years of treatment, and for 1 year after treatment was completed.

Data Completeness

The P-1 study has multiple, complex levels of missing and incomplete data. In the case of self-administered instruments, such as the HRQL questionnaire, participants could leave items blank by error or because they did not wish to answer the question.[6] Beyond this, the staffs of collaborating centers were generally unable to collect self-administered instruments on participants who quit taking pills because they no longer appeared for follow-up examinations, although many of these participants can still be observed for primary end points (eg, breast cancer and fractures). In addition, there are participants who did not complete all of the scheduled follow-up HRQL questionnaires because of the disclosure of the trial results in the spring of 1998,[1] although they are still observed for primary end points. Finally, a small proportion of participants (1.7%) were lost to follow-up, even for primary end points.

Statistical Analysis

The P-1 HRQL data set is composed of multiple HRQL instruments, each with its own psychometric properties and research history.[3] This complexity is magnified by the fact that data distributions and patterns of missing data differ across the various instruments included in the HRQL questionnaire. In addition, sample sizes are large, resulting in the possibility of statistically significant findings for clinically negligible effects. All of these considerations argue for future detailed analyses of the data from each specific instrument. In this initial report, however, our aims were essentially descriptive in nature and emphasized basic comparisons of the two trial groups. In making these comparisons, we seek to identify consistent differences, between the trial groups, using simple nonparametric procedures. The sign test[7] is used to examine the consistency of binary differences (±) between the two trial groups across time, independent of the magnitude of these differences. A one-sided alternative is routinely used because tamoxifen is expected to have a negative effect on most short-term measures of HRQL. Friedman's test,[7] implemented as a generalization of the paired sign test,[8] was used as a nonparametric analog to the two-way analysis of variance when we wanted to block on a specific factor, such as age group. Positive findings, with regard to consistent differences between trial groups, were independently reviewed for magnitude to assess their clinical and functional significance for the participants' quality of life.

Clinical experience, as well as initial statistical investigations of the P-1 HRQL data set, suggested that the age of the study participants was a key factor contributing to the observed distribution of HRQL measures. Hence, the results presented here from various HRQL instruments were routinely stratified by three age groups (35 to 49 years, 50 to 59 years, and 60 years or older) that generally paralleled menopausal status. Relative risks (RRs) or absolute differences in mean counts are presented in the tables to estimate differences in effect size between the two groups.

Imputation procedures for missing items in otherwise complete scales were only used for eight SF-36 subscales, as recommended in the SF-36 scoring manual.[9] No data imputation was carried out for other scales, and incomplete scales were considered missing.[1]

RESULTS

Table 1 lists the demographic, medical, and behavioral characteristics of our participant cohort of 11,064 women by trial group. These data show that the women in the P-1 study were predominately white (96%), well educated (65% ≥ some college), married (70%), professional and technically trained (68.2%), currently employed (64.9%), and reported a middle- to upper-middle class family income (median, $35,000 to $49,999). None of the variables in Table 1 show a striking imbalance between the two trial groups.

Figure 1 charts the overall proportion and total numbers of women completing the HRQL questionnaire at each examination. It provides a general measure of comparative participant adherence with regard to the HRQL questionnaire in the two trial groups. Both trial groups showed a consistent decline in HRQL adherence across the first 36 months of the study, averaging 4.2% per examination in the placebo group and 4.6% per examination in the tamoxifen group. The proportion of HRQL-adherent participants was smaller in the tamoxifen than in the placebo group at every one of the seven follow-up examinations (sign test, $P = .0078$), with a maximum difference of 3.1% occurring at 36 months.

A number of demographic, clinical, and HRQL variables were examined to investigate whether differences could be detected between the women who failed to complete the HRQL questionnaire at 36 months in the tamoxifen and the placebo groups. These variables included mean age (tamoxifen = 53.1 years *v* placebo = 53.5 years) and mean RR (5.42 *v* 5.43), treatment status (10.1% *v* 10.5% on treatment), breast cancer in a first-degree relative (76.89% *v* 78.40%), prior estrogen use (32.5% *v* 33.3%), mean maximum CES-D score (12.52 *v* 12.46), and mean maximum number of reported symptoms on the SCL (14.2 *v* 13.9). These comparisons suggested that participants who failed to complete the HRQL questionnaire in each group were similar cohorts of women.

When, within a treatment group, the same variables were used to compare HRQL adherent and nonadherent women, only the treatment status variable was different between the two groups. A significantly greater proportion of HRQL-adherent women in both groups remained on treatment (87.0% *v* 89.6%) compared with HRQL-nonadherent women (10.1% *v* 10.5%). In other words, adherence in the HRQL component of P-1 was largely a reflection of treatment adherence. This was because most collaborating centers did not have the staff resources to administer the HRQL

Table 1. Demographic, Clinical, and Health Behavior Characteristics of P-1 HRQL Study Participants (N = 11,064)

Characteristic	Placebo		Tamoxifen		Total	
	No of Patients	%	No of Patients	%	No of Patients	%
Age, years						
Mean ± SD	53.83 ± 9.167		53.82 ± 9.184		53.83 ± 9.175	
Median	52		52		52	
Range	35-79		35-78		35-79	
Ethnicity						
White	5,290	95.54	5,282	95.57	10,572	95.55
Hispanic	63	1.14	49	0.89	112	1.01
Black	88	1.59	95	1.72	183	1.65
Asian	35	0.63	37	0.67	72	0.65
Other	47	0.84	39	0.71	86	0.78
Missing	14	0.25	25	0.45	39	0.35
Education						
Grade school	61	1.10	66	1.19	127	1.15
Some high school	248	4.48	218	3.94	466	4.21
High school graduate	1,003	18.11	1,009	18.26	2,012	18.19
Vocational school	593	10.71	614	11.11	1,207	10.91
Some college	1,180	21.31	1,194	21.60	2,374	21.46
Associate degree	349	6.30	349	6.31	698	6.31
College graduate	664	11.99	732	13.24	1,396	12.62
Professional school	546	9.86	519	9.39	1,065	9.63
Master's degree	726	13.11	684	12.38	1,410	12.74
Doctoral degree	133	2.40	106	1.92	239	2.16
Missing	34	0.61	36	0.65	70	0.63
Employment						
Unemployed	239	4.32	229	4.14	468	4.23
Retired	925	16.71	938	16.97	1,863	16.84
Full-time homemaker	660	11.92	670	12.12	1,330	12.02
Student	30	0.54	33	0.60	63	0.57
Employed full-time	2,713	49.00	2,682	48.53	5,395	48.76
Employed part-time	880	15.89	878	15.89	1,758	15.89
On medical leave	25	0.45	24	0.43	49	0.44
Permanently disabled	51	0.92	47	0.85	98	0.89
Missing	14	0.25	26	0.47	40	0.36
Occupation						
Homemaker	849	15.33	843	15.25	1,692	15.29
Professional	2,207	39.86	2,188	39.59	4,395	39.72
Technical	1,573	28.41	1,548	28.01	3,121	28.21
Services	487	8.80	487	8.81	974	8.80
Operators	92	1.66	94	1.70	186	1.68
Other	315	5.69	341	6.17	656	5.93
Missing	14	0.25	26	0.47	40	0.36
Income						
Under $10,000	211	3.81	161	2.91	372	3.36
$10,000-$19,999	549	9.91	571	10.33	1,120	10.12
$20,999-$34,999	1,127	21.35	1,170	21.17	2,297	20.76
$35,000-$49,999	936	16.90	984	17.80	1,920	17.35
$50,000-$74,999	1,153	20.82	1,151	20.83	2,304	20.82
$75,000-$99,000	511	9.23	478	8.65	989	8.94
$100,000 or more	564	10.19	521	9.43	1,085	9.81
Unanswered	296	5.35	301	5.45	597	5.40
Missing	190	3.43	190	3.44	380	3.43
RR of breast cancer						
1-2	416	7.51	416	7.53	832	7.52
2-3	929	16.78	865	15.65	1,794	16.21
3-5	2,074	37.46	2,154	38.97	4,228	38.21
5-10	1,618	29.22	1,605	29.04	3,223	29.13
10+	500	9.03	487	8.81	987	8.92
1st degree relatives w/breast cancer						
0	1,238	22.36	1,191	21.56	2,429	21.95
1	3,239	58.50	3,250	58.80	6,489	58.65
2	903	16.31	902	16.32	1,805	16.31
≥ 3	157	2.83	184	3.32	341	3.09
Marital status						
Never married	398	7.19	394	7.13	792	7.16
Presently married	3,843	69.41	3,876	70.43	7,719	69.77
Marriage-like	139	2.51	125	2.26	264	2.39
Divorced	748	13.51	707	12.79	1,455	13.15
Widowed	395	7.13	399	7.22	794	7.18
Unknown	0	0	1	0.02	1	0.01
Missing	14	0.25	25	0.45	39	0.35
Smoking						
Smoked at least 100 cigarettes in lifetime	2,697	48.83	2,729	49.60	5,470	50.39
Smoked at least 100 cigarettes in lifetime and currently smoke	705	12.76	712	12.94	1,417	12.85
Alcohol						
Never use	1,138	20.60	1,128	20.50	2,266	20.55
Some days	4,129	74.76	4,147	75.37	8,276	75.07
Every day	256	4.64	227	4.13	483	4.38
Previous estrogen use	1,171	31.98	1,838	33.25	3,609	32.62
Both ovaries removed	797	14.39	813	14.71	1,610	14.55
Menstrual period stopped	3,658	66.06	3,685	66.67	7,343	66.37

questionnaire via the telephone or mail to women who stopped treatment and failed to appear for their scheduled follow-up visits.

By the 36-month examination, 3,421 women had stopped their assigned treatment and failed to fill out the HRQL questionnaire for at least 6 months. Table 2 lists the primary reasons these women gave for stopping treatment. The placebo and tamoxifen groups did not differ with regard to protocol-specified events, such as invasive breast cancer, depression, or deep vein thrombosis, or other medical reasons, such as anxiety disorders or cardiovascular conditions. Hot flashes were clearly the most frequently reported sign or symptom that caused women to stop their assigned treatment (251 women); they occurred most often in the tamoxifen group (184 women). When stopping their as-

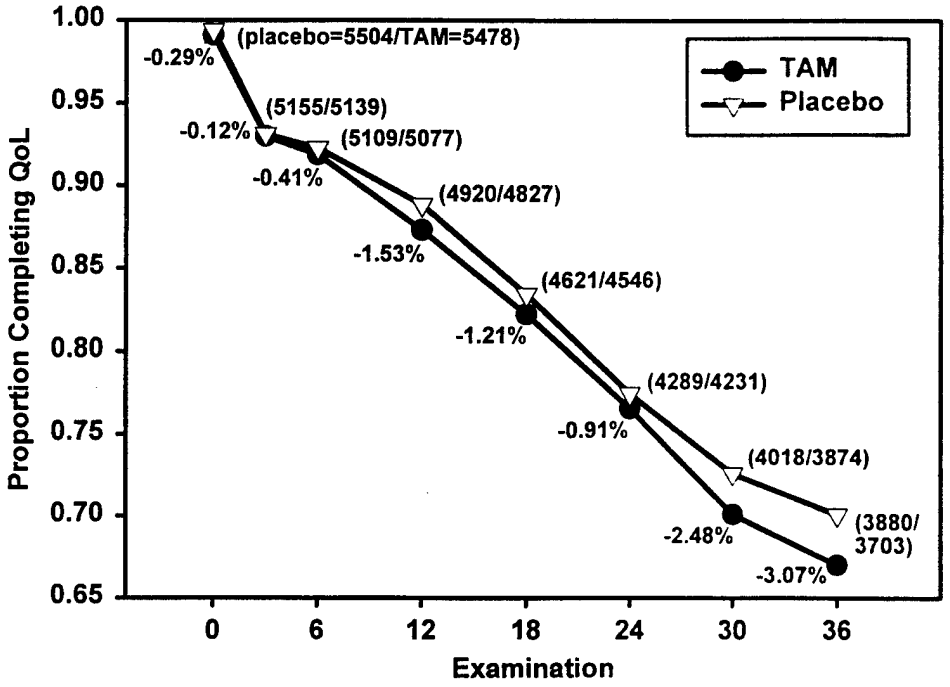


Fig 1. Proportion of participants in the tamoxifen group and placebo group completing HRQL questionnaire by examination (placebo, n = 5,537; tamoxifen, n = 5,527). Figures on chart are the number of women in the placebo/tamoxifen groups completing the HRQL questionnaire and the difference between TAM and placebo groups in terms of percent missing HRQL data.

signed treatment, participants in the placebo group were more likely to cite other nonmedical reasons, such as fear of side effects, change of mind, or desire to adopt an alternative therapy (eg, hormone replacement).

Table 3 shows the proportion of P-1 participants, by age group and examination, who scored above the most frequently used clinical cutoff (≥ 16) on the CES-D.[10,11] The youngest age group (35 to 49 years) in both trial groups consistently had the highest proportion of members scoring above the clinical cutoff, followed by the 50- to 59-year-old age group (Friedman test, $P = .001$ tamoxifen and placebo). The RRs listed in Table 3 show that, for all three age groups, the magnitude of the differences is small, and there was no consistent excess of participants in the tamoxifen group scoring above the clinical cutoff on the CES-D when compared with the placebo group. Similar findings with regard to the relationship between the two trial groups emerged from the analysis of the five-item mental health subscale on the MOS SF-36 (not shown).

The results of the SF-36 are summarized using the physical component summary (PCS) and mental component summary (MCS) scores[12] and the eight SF-36 subscales. The PCS and MCS scores represent aggregate measures that combine data from the eight subscales generally reported on the SF-36. The PCS aggregates data from the Physical Functioning, Role-Physical, Bodily Pain, and General Health subscales, while the MCS draws on data from the Vitality, Social Functioning, Role-Emotional, and Mental Health subscales. The PCS and MCS are scored using norm-based

Table 2. Reasons for Stopping Assigned Therapy by Participants Not Completing Quality of Life Questionnaire (Baseline to 36-Month Examination, n = 3421)

Reason for Stopping Assigned Therapy	Tamoxifen		Placebo		Total	
	No. of Patients	%	No. of Patients	%	No. of Patients	%
Protocol specified event	164	9.1	154	9.6	318	9.3
Reported signs or symptoms	545	30.2	336	20.8	881	25.8
Other medical	342	18.9	280	17.3	622	18.2
Other nonmedical	753	41.7	842	52.1	1595	46.6
Unknown	2	0.1	3	0.2	5	0.1
Total	1806	52.8	1615	47.2	3421	100.0

Table 3. Proportion of Participants in Tamoxifen Arm With a Clinically Significant Score (≥ 16) on the CES-D by Age Group and Examination

Examination	Age Group							
	35-49 Years		50-59 Years		≥ 60 Years		Overall	
	TAM	RR*	TAM	RR*	TAM	RR*	TAM	RR*
Baseline	0.074	1.03	0.082	1.28	0.058	0.918	0.071	1.07
3 months	0.122	1.10	0.104	1.05	0.085	1.08	0.105	1.08
6 months	0.138	1.06	0.114	1.00	0.093	0.910	0.117	1.00
12 months	0.128	0.937	0.122	0.999	0.096	0.989	0.116	0.968
18 months	0.139	0.892	0.126	0.918	0.101	0.929	0.123	0.908
24 months	0.143	1.02	0.124	0.980	0.095	0.924	0.122	0.980
30 months	0.142	0.978	0.107	0.961	0.104	0.934	0.120	0.959
36 months	0.135	0.898	0.111	1.04	0.097	0.887	0.116	0.930

Abbreviation: TAM, tamoxifen.
*RR = TAM/placebo.

methods; both component scores have a mean of 50 and a SD of 10 in the general United States (U.S.) population. This means that the PCS and MCS can be meaningfully compared with one another, and their scores have a direct interpretation in relation to the distribution of scores in the general U.S. population.

Figure 2 charts the PCS and MCS for the tamoxifen and placebo groups at each examination and by age group. As expected, mean PCS declines across the age groups. At follow-up examinations, the tamoxifen group was consistently lower on the PCS only in the 50- to 59-year-old age group (one-sided sign test, $P = .065$). However, the absolute differences were small, approximating one tenth of an SD. With regard to the MCS, all of the age groups scored above the mean MCS for the general U.S. population, and no consistent differences emerged between the two trial groups. Figure 3 summarizes the overall data from eight subscales on which the component subscores are based.

Table 4 lists the mean number of symptoms reported on the 43-item SCL by age group and examination. The mean number of symptoms reported was consistently highest in the 50- to 59-year-old age group, followed by the 35- to 49-year-old and 60 years or older age groups (Friedman test, $P = .001$ tamoxifen and placebo). The participants in the tamoxifen group also reported a small but consistent excess in the mean number of symptoms (< one) reported at 19 of the 21 age-stratified follow-up examinations (3 to 36 months; one-sided sign test, 35 to 49 years, $P = .0078$; 50 to 59 years and ≥ 60 years, $P = .065$) (Table 4).

Table 5 provides information on the proportion of women in the tamoxifen and placebo groups who reported symptoms on the SCL at least once during the treatment period, ie, the period excluding baseline but including the seven follow-up examinations. The five symptoms with the greatest relative difference between the two trial groups are given for each age group, and the 10 symptoms with the greatest relative difference are presented for all participants combined.

Tables 6 and 7 give detailed information, by age group and examination, on the reported frequency of hot flashes and vaginal discharge in the trial groups. The proportion of participants who reported hot flashes was elevated in all age groups of the tamoxifen group at every follow-up examination. Among the participants in the tamoxifen group, the 50- to 59-year-old age group had the largest proportion of women reporting hot flashes at each examination (median, 69.8%; Friedman test, $P = .001$), but the youngest age group (35 to 49 years) showed the greatest relative increase in proportion of women reporting hot flashes (median RR, 1.50; Friedman test, $P = .011$). Vaginal discharge was the most consistently elevated symptom in the tamoxifen group. The youngest age group (35 to 49 years) had the greatest proportion of participants reporting vaginal discharge at each examination (median, 35.5%; Friedman test, $P < .001$), and the oldest age group (≥ 60 years) reported the greatest increase of vaginal discharge relative to the placebo controls (median RR, 3.05; Friedman test, $P = .005$).

Figure 4 summarizes the information from the five items on the MOS sexual functioning scale. Figure 4A shows that a greater proportion of participants in the tamoxifen group, as compared with the placebo group, reported being sexually active during the 6 months before each follow-up examination. Although apparently consistent ($P = .031$), the absolute difference was small (mean, 0.78%) and may have been caused by chance. Figure 4B through 4E show that a small but consistently larger percentage of participants in the tamoxifen group reported a definite or serious problem in three of the four specific domains of sexual functioning during the follow-up period.

DISCUSSION

We observed in our earlier article[3] that measuring the impact of new treatments on HRQL is particularly important within the context of disease-prevention and health-promotion trials. Compared with patients suffering from clinically manifest disease, decrements in overall quality of life are likely to have a much greater impact on the subjective appraisal of treatment acceptability and the maintenance of long-term treatment adherence among high-risk but otherwise healthy individuals. This report covers the initial HRQL findings from a large, multicenter chemoprevention trial, which has shown that tamoxifen reduced the risk of invasive breast cancer in high-risk women by 49% during the first 5 years of administration. Given the apparent clinical efficacy of tamoxifen in the prevention setting, it is important to assess whether the various secondary effects of the drug might act to reduce this practical efficacy.[13-15]

The cohort of women taking part in the P-1 study clearly was not representative of the general population. They were predominately white, well educated, and middle class, with a strong professional and technical orientation. The initial HRQL findings presented in this report must be assessed within the context of the socioeconomic and cultural characteristics of the P-1 study cohort.

The subcohort of women discussed in this report represent 82.6% of the total study cohort. This subcohort was chosen to exclude potential biases, because of external factors eventuating in the suspension of accrual in P-1, and to control for the amount and types of missing data. Despite this, we still lost 31.5% of our participants by the 36-month follow-up examination. This proportion closely approximates the 10%-per-year loss to follow-up rate predicted at

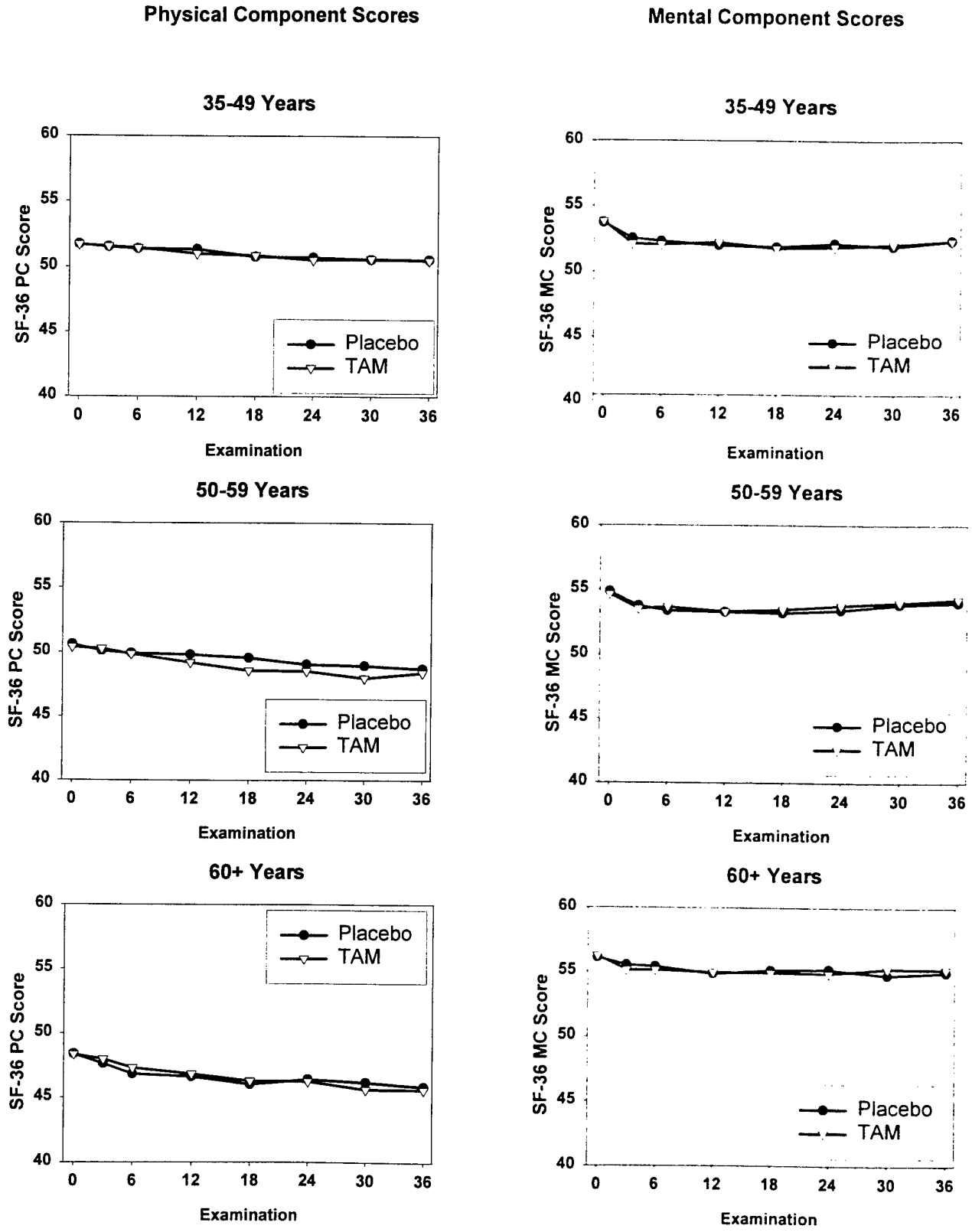


Fig 2. Mean scores by age group and examination on SF-36 physical and mental component scores (higher scores represent better quality of life).

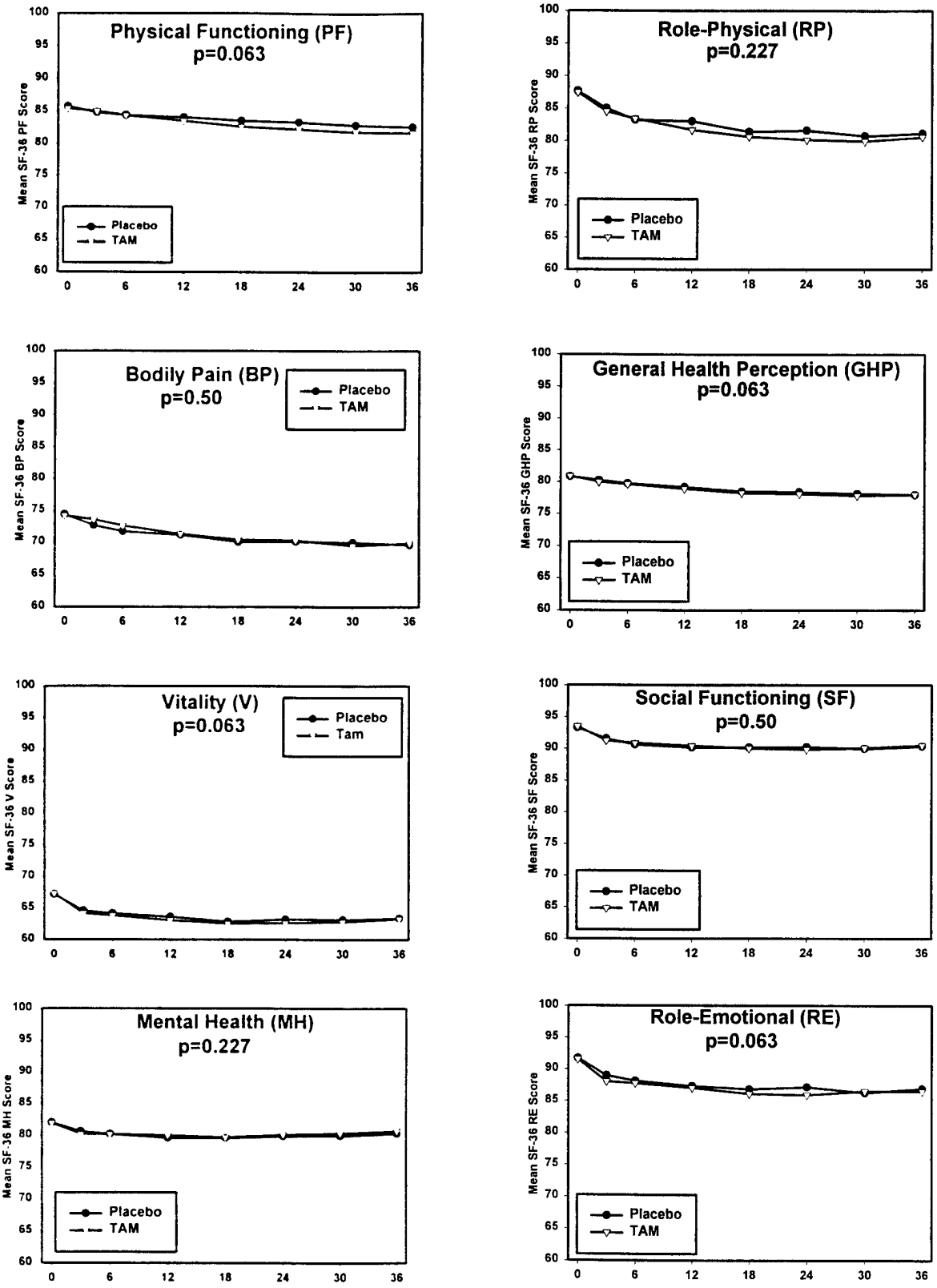


Fig 3. Mean SF-36 subscale scores by examination.

Table 4. Mean Number of Total Symptoms Reported on Symptom Checklist by Age Group and Examination

	Age Group							
	35-49 Years		50-59 Years		≥ 60 Years		Overall	
Examination	TAM	Difference*	TAM	Difference*	TAM	Difference*	TAM	Difference*
Baseline	8.84	+ 0.114	9.76	+ 0.236	8.89	- 0.030	9.14	+ 0.110
3 months	9.96	+ 0.319	10.54	− 0.006	9.63	- 0.166	10.04	+ 0.077
6 months	10.43	+ 0.564	11.06	+ 0.304	10.06	- 0.011	10.51	+ 0.322
12 months	10.87	+ 0.521	11.54	+ 0.655	10.43	+ 0.076	10.95	+ 0.429
18 months	11.08	+ 0.614	11.51	+ 0.452	10.65	+ 0.292	11.08	+ 0.469
24 months	11.05	+ 0.733	11.58	+ 0.549	10.68	+ 0.476	11.10	+ 0.602
30 months	10.27	+ 0.227	10.67	+ 0.547	10.15	+ 0.134	10.36	+ 0.299
36 months	10.79	+ 0.386	11.22	+ 0.700	10.50	+ 0.190	10.84	+ 0.426

Abbreviation: TAM, tamoxifen.

*Difference = tamoxifen minus placebo.

the beginning of the P-1 trial and is similar in pattern and number to the adherence data recently reported in a second large, multicenter chemoprevention trial of hormone replacement therapy for heart disease.[16] We have shown that there is only a small difference in the proportion of nonadherent participants in the tamoxifen and placebo groups and that the nonadherent women in both trial groups have generally similar key demographic, clinical, and HRQL variables. Given these considerations, it seems unlikely that a maximum difference of 3% in the HRQL follow-up rates between the two groups was sufficient to create a significant bias in our between-group comparisons.

HRQL adherence is closely related to treatment adherence. Based on the reasons for quitting treatment, it would seem that nonadherent women in both trial groups were those who were sensitive to the actual or possible occurrence of side effects caused by tamoxifen.

Much concern has been expressed about a potential relationship between tamoxifen use and the onset of depression.[17-21] Women who reported a history of depressive episodes or a history of treatment for nervous or mental disorders were not excluded from the trial. A brief eight-item affective screening questionnaire based on the CES-D and the Diagnostic Interview Schedule[22] was part of the baseline examination.[23] Using data from this brief screening instrument, local investigators were alerted to eligible participants showing signs of potentially serious affective distress at the baseline examination and caution was advised regarding their enrollment onto the trial. However, women who showed current signs of affective distress or depression were not routinely excluded from the trial.

With regard to the primary screening instrument used in the follow-up examinations, it has been pointed out that "the items in... (the CES-D) are generally related to affective distress but not to any particular psychiatric disorder."[11] For this reason, the numbers listed in Table 3 refer not to the prevalence of clinically diagnosable depressive disorders but, instead, to the prevalence of clinically significant affective distress that might be associated with a number of specific psychiatric disorders. However, if tamoxifen use was associated with the onset of clinically diagnosable depression, we would have expected to see a consistent excess of individuals scoring ≥ 16 on the CES-D in the tamoxifen group. No such consistent excess was observed. These findings agreed with the data from the mental health scale on the SF-36.

The MOS SF-36 served in this study as a measure of overall HRQL. For this initial report, we have presented data from the SF-36 in terms of two high-level component scores[12] and the eight basic subscales generally used in scoring this instrument.[9] Neither of these two methods of summarizing the SF-36 data demonstrated any clinically significant differences between the tamoxifen and placebo groups.

The first clear signs of consistent differences between the tamoxifen and placebo groups were observed in the SCL. In 19 out of 21 follow-up comparisons, the mean number of symptoms reported on the SCL were consistently different by age group (50 to 59 years > 35 to 49 years > 60+ years) and by trial group (tamoxifen > placebo). The absolute differences between the trial groups were relatively small and tended to be associated with the types of vasomotor, gynecologic, and sexual functioning symptoms previously reported for tamoxifen.[18,24,25]

The data from the MOS sexual functioning scale indicate that relatively small (< 4.0%) but consistent differences exist between the two groups in regard to the proportion of women reporting definite or serious problems in at least three specific domains of sexual functioning, sexual interest, arousal, and orgasm. These problems do not seem to be age group specific. Despite these findings for specific domains of functioning, there is no evidence that these problems result in a reduction of the overall proportion of women in the tamoxifen group who are sexually active.

Table 5. Symptoms Reported at Least Once Between Months 3 and 36 With the Largest Relative Difference Between Trial Arms

Age Group and Symptom	Placebo Arm Proportion (%)	Tamoxifen Arm Proportion (%)	RR (TAM/Placebo)
35-49 years			
Cold sweats	15.90	22.90	1.44
Vaginal discharge	46.29	62.55	1.35
Pain in intercourse	23.88	31.57	1.32
Night sweats	59.58	74.16	1.24
Hot flashes	65.54	81.28	1.24
50-59 years			
Cold sweats	16.11	27.00	1.68
Vaginal discharge	32.51	53.47	1.64
Genital itching	36.93	45.24	1.23
Night sweats	62.77	75.88	1.21
Bladder control (laugh)	47.67	56.94	1.19
≥ 60 years			
Vaginal bleeding	4.64	10.92	2.35
Vaginal discharge	19.82	45.81	2.31
Genital itching	32.05	40.96	1.28
Hot flashes	51.51	63.59	1.23
Bladder control (laugh)	49.88	56.49	1.13
Overall			
Vaginal discharge	34.13	54.77	1.60
Cold sweats	14.77	21.40	1.45
Genital itching	38.29	47.13	1.23
Night sweats	54.92	66.80	1.22
Hot flashes	65.04	77.66	1.19
Pain in intercourse	24.13	28.19	1.17
Bladder control (laugh)	46.65	52.51	1.13
Bladder control (other)	47.79	52.83	1.11
Weight loss	41.97	44.94	1.07
Vaginal bleeding	21.26	21.96	1.03

Abbreviation: TAM, tamoxifen.

Based on these data, we conclude that tamoxifen use is associated with an increase in specific vasomotor, gynecologic, and sexual functioning symptoms. At the same time, we did not observe any evidence that overall physical and emotional well being were significantly affected by these differences in the frequency of symptoms. We also found no evidence on the CES-D or the SF-36 mental health scale for an association in any age group between tamoxifen use and an increase in the proportion of women reporting clinically significant levels of affective distress and/or depression. How should clinicians integrate the results from the HRQL study data into decision-making and recommendations to women considering the use of tamoxifen in the setting of prevention? As demonstrated by the SCL data from the placebo group of the trial, many symptoms experienced by women who participated in this study are age and menopause related and exist independent of the use of tamoxifen. However, several symptoms are substantially more frequent in women using tamoxifen; these include vasomotor symptoms (cold sweats, night sweats, and hot flashes), vaginal discharge, and genital itching. Women need to be informed

Table 6. Proportion of Women Reporting Hot Flashes in Tamoxifen Arm and RR Compared to Placebo Arm by Age Group and Examination

	Age Group							
	35-49 Years		50-59 Years		≥ 60 Years		Overall	
Examination	TAM	RR*	TAM	RR*	TAM	RR*	TAM	RR*
Baseline	0.258	0.959	0.533	0.989	0.268	1.030	0.346	0.991
3 months	0.581	1.588	0.761	1.241	0.511	1.413	0.616	1.399
6 months	0.610	1.666	0.765	1.268	0.503	1.481	0.626	1.455
12 months	0.614	1.525	0.740	1.273	0.460	1.412	0.606	1.396
18 months	0.613	1.510	0.715	1.239	0.419	1.461	0.586	1.387
24 months	0.622	1.457	0.681	1.199	0.388	1.311	0.570	1.322
30 months	0.627	1.362	0.642	1.206	0.330	1.177	0.541	1.265
36 months	0.627	1.414	0.667	1.276	0.364	1.362	0.560	1.348

Abbreviation: TAM, tamoxifen.
*RR = TAM/placebo.

of these possible symptoms. Weight gain and depression, two clinical problems anecdotally associated with tamoxifen treatment in women with breast cancer, did not increase in frequency in this large placebo-controlled trial of healthy women. This is good news that must also be communicated to women. An informed discussion with a woman considering tamoxifen therapy should include these points in the risk/benefit discussion.

Disclosure of likely and unlikely symptoms should prepare a woman for what she might experience and reduce her anxiety or concerns should she begin preventive therapy. Without the detailed evaluation of HRQL data obtained in the P-1 trial, we would not be able to provide this level of information and reassurance to women considering preventive therapy. In addition, the setting of preventive therapy differs considerably from the treatment of breast cancer. Therefore, if a woman experiences untoward symptoms after starting tamoxifen treatment, the medication can be discontinued if the symptoms cannot be controlled or her personal assessment of the risks and benefits changes.

Table 7. Proportion of Women Reporting Vaginal Discharge in Tamoxifen Arm and RR Compared to Placebo Arm by Age Group and Examination

	Age Group							
	35-49 Years		50-59 Years		≥ 60 Years		Overall	
Examination	TAM	RR*	TAM	RR*	TAM	RR*	TAM	RR*
Baseline	0.201	0.957	0.135	1.041	0.058	0.907	0.138	0.975
3 months	0.379	1.549	0.308	2.023	0.275	3.665	0.326	1.972
6 months	0.391	1.686	0.302	1.931	0.269	3.057	0.327	1.973
12 months	0.380	1.700	0.304	1.973	0.262	3.333	0.321	2.020
18 months	0.363	1.558	0.278	2.251	0.252	3.029	0.303	1.961
24 months	0.341	1.797	0.272	1.991	0.238	2.994	0.288	2.052
30 months	0.325	1.633	0.282	2.404	0.246	3.075	0.288	2.083
36 months	0.316	1.671	0.264	2.332	0.241	3.096	0.277	2.095

Abbreviation: TAM, tamoxifen.
*RR = TAM/placebo.

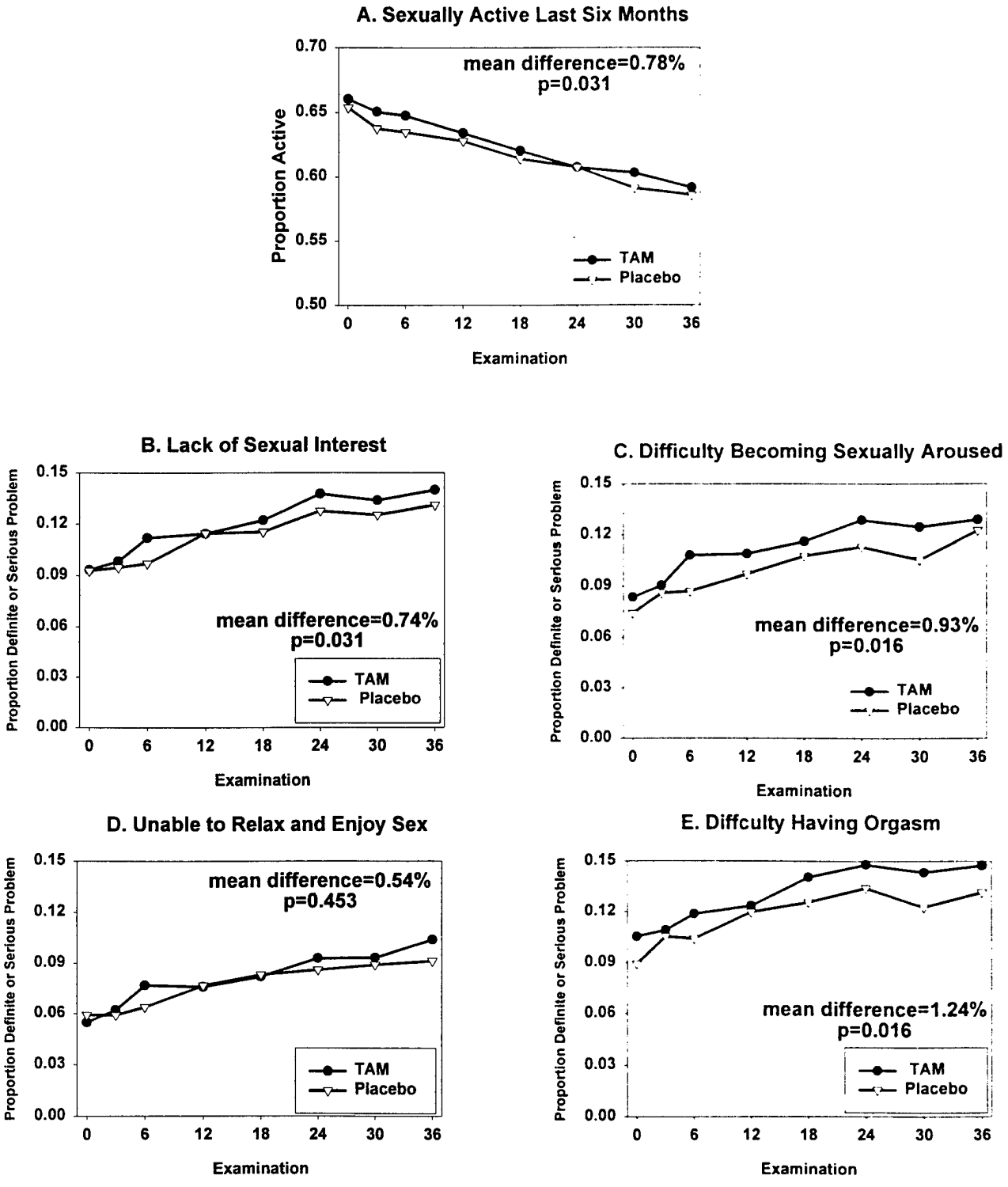


Fig 4. Proportion of women in the tamoxifen group and placebo group reporting a definite or serious problem in past 4 weeks on MOS sexual functioning scale (B through E, women who reported being sexually active in last 6 months).

The current report is a brief overview of the P-1 study HRQL data that focuses on important clinical and functional implications of tamoxifen use for women's overall HRQL. It will be supplemented in the future by a series of additional methodologic and clinical reports that will provide in-depth analyses of the data obtained from each one of the several P-1 study HRQL instruments.

ACKNOWLEDGMENT

We thank Carol Redmond, DSc, University of Pittsburgh; Leslie Ford, MD, National Cancer Institute, Bethesda, MD; Carol Moinpour, PhD, Southwest Oncology Group Statistical Center; John E. Ware, Jr, New England Medical Center, Boston, MA; David Cella, Northwestern University, Chicago, IL; Sheela Goshal and Wei Chen, NSABP Biostatistical Center; and members of the NSABP Prevention Quality of Life Committee.

REFERENCES

1. Fisher B, Costantino JP, Wickerham L, et al: Tamoxifen for the prevention of breast cancer: A report from the NSABP P-1 study. J Natl Cancer Inst 90:1371-1388, 1998

2. Fisher B, Costantino J: Highlights of the NSABP Breast Cancer Prevention Trial. Cancer Control 4:78-86, 1997

3. Ganz PA, Day R, Ware JE, et al: Base-line quality-of-life assessment in the National Surgical Adjuvant Breast and Bowel Project Breast Cancer Prevention Trial. J Natl Cancer Inst 87:1372-1382, 1995

4. Fisher B, Redmond C: Fraud in breast cancer trials. N Engl J Med 330:1458-1460, 1994

5. Fisher B, Anderson S, Redmond C, et al: Reanalysis and results after 12 years of follow-up in a randomized clinical trial comparing total mastectomy with lumpectomy with or without irradiation in the treatment of breast cancer. N Engl J Med 333:1456-1461, 1995

6. Ganz PA, Day R, Costantino JP: Compliance with quality of life data collection in the NSABP breast cancer prevention trial. Stat Med 17:613-622, 1998

7. Daniel WW: Applied Non-Parametric Statistics. Boston, MA, PWS-Kent Publishing Co, 1990

8. Deshpande JV, Gore AP, Shanubhougue A: Statistical Analysis of Non-Normal Data. New York, NY, John Wiley & Sons, 1995

9. International Resource Center for Health Care Assessment: How to Score the SF-36 Health Status Survey. Boston, MA, New England Medical Center, 1991

10. Radloff LS: The CES-D scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1:385-401, 1977

11. Roberts RE, Vernon SW: The Center for Epidemiologic Studies Depression Scale: Its use in a community sample. Am J Psychiatry 140:41-46, 1983

12. Ware JE, Kosinski M, Keller SD: SF-36 Physical and Mental Summary Scales: A User's Manual (3rd Printing Revised). Boston, MA, The Health Institute, New England Medical Center, 1994

13. Fisher B: A commentary on endometrial cancer deaths in tamoxifen-treated breast cancer patients. J Clin Oncol 14:1027-1039, 1996

14. Fisher B, Costantino JP, Redmond CK, et al: Endometrial cancer in tamoxifen-treated breast cancer patients: Findings from NSABP B-14. J Natl Cancer Inst 86:527-537, 1994

15. Gorin MB, Day R, Costantino JP, et al: Long term tamoxifen citrate and potential ocular toxicity. Am J Ophthalmol 125:493-501, 1998

16. Hulley S, Grady D, Bush T, et al: Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. JAMA 280:605-613, 1998

17. Cathacart CK, Jones SE, Pumroy CS, et al: Clinical recognition and management of depression in node negative breast cancer patients treated with tamoxifen. Breast Cancer Res Treat 27:277-281, 1993

18. Love RL, Cameron L, Connell BL, et al: Symptoms associated with tamoxifen treatment in postmenopausal women. Arch Intern Med 151:1842-1847, 1991

19. Shariff S, Cumming CE, Lees A, et al: Mood disorder in women with early breast cancer taking tamoxifen, an estradiol receptor antagonist: An unexpected effect? Ann N Y Acad Sci 761:365-368, 1995

20. Moredo Anelli T, Anelli A, Tran KN, et al: Tamoxifen administration is associated with a high rate of treatment-limiting symptoms in male breast cancer patients. Cancer 74:74-77, 1994

21. Pluss JL, Dibella NJ: Reversible central nervous system dysfunction due to tamoxifen in a patient with breast cancer. Ann Intern Med 101:652, 1984

22. Robins LN, Helzer JE, Croughan J, et al: National Institute of Health Diagnostic Interview Schedule: Its history, characteristics and validity. Arch Gen Psychiatry 35:837-846, 1978

23. Burnam MA, Wells KB, Leake B, et al: Development of a brief screening instrument for detecting depressive disorders. Med Care 26:775-789, 1988

24. Fisher B, Dignam J, Bryant J, et al: Five versus more than five years of tamoxifen therapy for breast cancer patients with negative lymph nodes and estrogen receptor-positive tumors. J Natl Cancer Inst 88:1529-1542, 1996

25. Fisher B, Costantino J, Redmond C, et al: A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. N Engl J Med 320:479-484, 1989