1. Introduction

Breast cancer patients receiving adjuvant endocrine treatment are faced with menopausal symptoms. These symptoms have an important impact on patients’ daily life, not only on physical but also on emotional well-being. For example, vasomotor symptoms are frequent and may be associated with sleeping difficulties and feelings of depression. Quality-of-life (QL) evaluation including menopausal symptoms is therefore an essential objective in the IBCSG Trial 24-02.

The two objectives of QL assessment in this protocol are: (i) to compare treatments in regard to QL; (ii) to compare QL in this trial to that in other adjuvant breast cancer trials.

These objectives will be addressed using the same basic approach to QL assessment that the IBCSG has used since 1986. It has been shown to be feasible for international breast cancer clinical trials. Previous methodological and clinical investigations were mainly related to chemo- and chemo-endocrine effects. An example relevant to the IBCSG Trial 24-02 is the analysis of the indicator for hot flushes in node-negative postmenopausal patients with operable breast cancer (IBCSG Trial IX). This indicator showed a high discriminative capacity between patients on chemotherapy and those on tamoxifen within the first three months.

Using the same approach to QL assessment will allow us to make comparisons across IBCSG trials, using the extensive QL database, which will be available from other trials.

2. Objectives

QL will be described in regard to intermediate- and long-term sequelae of treatment and disease, and the treatments will be compared in regard to QL. Specifically, the hypotheses to be investigated are

2.1. Primary hypotheses

2.1.1 Patients receiving tamoxifen plus ovarian function suppression will report more menopausal symptoms than those with tamoxifen only (see section 8.).

2.1.2 Patients receiving tamoxifen plus ovarian function suppression will report more sexual impairment than those with tamoxifen only (see section 8.).

2.1.3 Patients receiving exemestane plus ovarian function suppression will report more menopausal symptoms than those with tamoxifen plus ovarian function suppression (see section 8.).

2.1.4 Patients receiving exemestane plus ovarian function suppression will report more sexual impairment than those with tamoxifen plus ovarian function suppression (see section 8.).
These hypotheses will be tested by comparing the treatment groups using serial measurements of QL indicators over time (see 5.1). The indicator for *hot flushes* included in the IBCSG QL Core Form is selected as primary endpoint for the first and third hypotheses. The indicator for *loss of sexual interest* included in the trial-specific QL module is selected as primary endpoint for the second and fourth hypotheses.

### 2. Secondary hypotheses

2.2.1 The differences between treatments will persist over the whole treatment period (see 8.).

2.2.2 Those patients who report more severe menopausal symptoms during the first 6 months on endocrine therapy will also report delayed adaptation in other QL indicators both during and beyond that time (see 8.).

Physical well-being, mood and coping will be used to describe patients’ adaptation over time. The findings of this trial will be compared to that in other IBCSG adjuvant breast cancer trials.

Further exploratory analyses will address (i) the relationships among the QL measures (e.g., the impact of menopausal symptoms on global QL indicators, especially on overall treatment burden, physical well-being, coping and mood); (ii) bio-psychosocial interactions (e.g., the impact of initial prognostic factors on QL).

### 3. Patient Selection

For IBCSG centers, patients must have completed baseline Quality of Life (QL) Forms prior to randomization. The only exceptions are cognitive or physical impairment that interferes with QL assessment or inability to read any of the languages available on IBCSG QL forms.

For non-IBCSG centers, extent of participation in the QL study will be determined at the activation of the trial for each participating cooperative group.

Recruitment of 1872 patients will maximize the opportunity to detect clinically relevant treatment effects on QL.

### 4. Study Design

As in the other IBCSG trials with QL, a longitudinal design is used, including a baseline assessment, assessments to evaluate intermediate and long-term effects, and assessments following treatment failure to evaluate the impact of relapse. To the extent feasible, the assessment time schedule is compatible to that of the other trials to keep it as simple as possible and to allow comparisons across trials.

Patients are asked to complete a QL core form plus a trial-specific module

- at baseline, prior to randomization
- every 6 months during the first and second year,
- and annually in years 3 to 6.
All patients, regardless of disease status, are to be assessed on the same schedule. A detailed data collection schedule is displayed in Figure 1.

Figure 1: Quality of Life Assessment Time Points

<table>
<thead>
<tr>
<th>Months from Randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

To eliminate any differential anticipatory effects on baseline scores and to help insure compliance with the protocol requirements, a QL Core Form plus Module must be completed prior to randomization (i.e., it is an eligibility requirement for IBCSG participating centers).

5. Quality-of-Life Measures
5.1 Patient rated Quality of Life

The QL assessment consists of the *IBCGQ NL Core Form* and a trial specific *module*.

The QL Core Form was developed in 1986, and was subsequently revised for IBCSG Trials 10-93 through 14-93, which started on May 1, 1993. The revised form is designed to address the endpoints in adjuvant trials more specifically while still keeping the questionnaire simple and short. It includes global linear analogue self-assessment (LASA) indicators for physical well-being, mood; coping (PACIS); perceived social support and subjective health estimation (SHE). The indicators for physical well-being, mood and coping were confirmed to be responsive to cytotoxic side effects, mental distress and psychosocial dysfunction in patients with early breast cancer. They are suitable to describe patients' adaptation over time. Validation studies are summarized elsewhere.

In addition, LASA indicators specific to symptoms of nausea and vomiting, tiredness, *hot flushes*, and restrictions in arm movement are included, covering possible QL effects of all of the treatment modalities involved in these trials (surgery, chemos-, endocrine and radiation therapy).
For the IBCSG Trial 24-02, a one-page module for endocrine symptoms will be used in addition to the QL Core form. Besides non-experimental studies, the selection of symptoms was based on two recent breast cancer prevention trials which compared the effects of tamoxifen with those of placebo \(^5, 6\). All symptoms are assessed in the LASA response format. In addition, one global indicator is included for overall treatment burden (“Overall, how much are you bothered by any treatment related difficulties?”). This indicator has been validated regarding side-effects of anti-emetic and cytotoxic therapies \(^17\) and is expected to be similarly responsive to endocrine symptoms.

An expanded three-page IBCSG Trial 24-02 module will be used for participating centers having English as the primary language. The first page will be the same as the module completed at all participating centers in the study. The second and third page will contain additional questions from the Center for Epidemiologic Studies-Depression Scale (CES-D) \(^18\) and the Medical Outcomes Study (MOS) sexual problems measures \(^19\). These questionnaires have been used frequently for US Intergroup studies. The CES-D and the MOS are included to provide cross-validation between responses to selected IBCSG core and module questions and these comprehensive measures. These questionnaires will also be used as common QL reference data for the IBCSG and CALGB trial. Their assessment is restricted to baseline, months 6, 12 and 24.

In clinical trials, the distinction between indicators of specific symptoms and global indicators sensitive to treatment as well as disease-related problems in the broadest sense is very useful. In the global indicators, the score reflects a patient's subjective, intuitive choice and weighting of different aspects, summarized in a single response. Specific disease and treatment-related indicators can be used to examine the changing impact of symptoms on overall measures over time and in different situations. This can be done within a treatment group (e.g., for patients receiving tamoxifen plus ovarian function suppression, how much of the variation in the coping measure can be explained by the major menopausal symptoms over the first year) or to explain differences in the global measures among treatment groups (e.g., does hot flushes explain differences in physical well-being among the randomized treatments at 12 months).

The conceptual basis of the IBCSG approach to QL assessment, along with a description of methodological issues, clinical findings and planned steps for further development, have been summarized elsewhere \(^9\).

5.2 Sociodemographics and Co-morbidity
As in the other IBCSG QL trials, sociodemographic data and co-morbidity are part of the standard study documentation.

6. Timing Requirements, Data Collection and Local Data Management
6.1 Timing requirements
Assessment time points are determined by the interval from date of randomization, and coincide with the required clinical follow-up time points. The QL assessment time points are illustrated in Figure 1.
The schedule of QL assessment must be followed as closely as possible. The QL form has to be filled in always prior to diagnostic procedures. If exact timing is not possible, assessment should be done as close as possible to the required date.

For methodological reasons, the required schedule has to be followed exactly, with neither more nor fewer assessments. Shortly after randomization, the IBCSG operations office sends the local investigator a schedule of the dates of required QL assessments. This list should be put into each patient's chart to aid in the correct timing of the QL assessment.

6.2 Data Collection and Local Data Management
Within the first 6 years, every study patient is to fill in both the QL Core Form and the trial specific module at each scheduled assessment time point, as described in Figure 1; no form selection is acceptable.

If the patient does not complete the required QL Forms, then a Missing QL Assessment Form must be submitted for that assessment time point to provide the reason why the assessment was not completed.

The QL forms are to be filled in at the clinic. If the patient is being followed elsewhere, arrangements are to be made with the clinic or physician to have the patient fill in the forms as required. If, for administrative reasons, the form has not been presented to the patient, it may be filled in at home and mailed.

For the first assessment, the QL forms have to be explained to the patient, with particular emphasis on making sure the patient understands both LASA format (used in all questionnaires) and categorical response format (used only in supplemental questionnaires for centers with English as the primary language). For later assessments, the patient should be instructed to seek help only if she has problems in understanding any of the items in the form.

All questions on the QL Core Form and the QL Module should be answered. The forms should be checked after completion and, if necessary, the patients should be asked to fill in missing answers. Patients may wish to leave some questions unanswered if they make them feel very uncomfortable. They should be encouraged to answer all items, however, especially those concerning menopausal symptoms, as they represent a primary objective of the QL study.

Detailed instructions for the QL assessment are given in the IBCSG QL Manual. Copies are available from the IBCSG Coordinating Center in Bern.

7. Central Data Management
Computerized data quality control measures will be used to monitor the submission rates of the QL forms and the timing of assessment as required by the study protocol. Institutions will receive feedback on their performance and specific problems on a regular basis.
8. Statistical Considerations

8.1 Sample Size Calculations

This phase III randomized clinical trial is designed to compare differences in QL between patients who receive five years of tamoxifen and those who receive five years of ovarian function suppression (OFS) plus five years of tamoxifen. In addition, differences between OFS plus tamoxifen and OFS plus exemestane will be assessed. These two pairwise comparisons will each be assessed separately with no adjustment for multiple comparisons. According to investigator choice, patients can either receive no adjuvant chemotherapy or receive adjuvant chemotherapy and remain premenopausal and be randomized following completion of chemotherapy. The randomization is stratified according to chemotherapy use. QL will be described in regard to intermediate and long-term sequelae of treatment and disease. We will test the above hypotheses by comparing the treatment groups using serial measurements of QL indicators over time.

Many QL indicators are collected, but the indicators hot flushes and loss of sexual interest are selected to determine sample size. The indicator for hot flushes included in the IBCSG QL Core Form is the primary endpoint used to test whether patients receiving tamoxifen with ovarian function suppression will report more menopausal symptoms than those receiving tamoxifen alone. This measure will also be considered primary for the OFS plus tamoxifen versus OFS plus exemestane comparison. We will calculate the QL sample size to reflect both short and long term effects. First, to reflect the short-term effects we will base the QL sample size on the mean treatment difference of the change in hot flush scores from baseline to 6 months. Next, to reflect the long-term effects we will use the mean treatment difference of the change in hot flush scores from baseline to 24 months. Both sample size calculations were based on a two-sided 0.05 level test and respective standard deviations from IBCSG Trial IX were used to compute the 2 sample sizes.

To address the short-term effects, the common standard deviation, calculated from the pooled estimator of the variance of the change in hot flush scores from IBCSG Trial IX of Arm E (Tamoxifen) and Arm F (CMF -> Tamoxifen), was 35.1. 304 assessable patients per arm (608 total) are needed to achieve an 80% statistical power to detect a mean treatment difference in a change in hot flush score of 8 units on the original scale. In IBCSG Trial IX, 66% of the eligible patients filled out the indicator for hot flushes at baseline and 6 months post randomization. If we assume a similar compliance rate in Trial 24-02 we would need a total of 921 patients randomized to two-arms to detect an 8-unit difference in treatment means of the change in hot flush scores from baseline to 6 months with 80% power. Thus, 1382 patients are required for the three-arm trial.
Table 1 illustrates the total number of patients enrolled in two arms needed to detect a range of given treatment differences of the change in hot flush scores from baseline to 6 months with 80% power.

<table>
<thead>
<tr>
<th>Δ*</th>
<th>Total Assessed</th>
<th>Total Needed Assuming 66% Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>390</td>
<td>591</td>
</tr>
<tr>
<td>9</td>
<td>480</td>
<td>727</td>
</tr>
<tr>
<td>8</td>
<td>608</td>
<td>921</td>
</tr>
<tr>
<td>7</td>
<td>792</td>
<td>1200</td>
</tr>
</tbody>
</table>

*Mean treatment difference of the change in hot flush score from baseline to 6 month

To address the long-term effects, the pooled standard deviation from IBCSG Trial IX between the two treatment arms was 34.2. In order to detect an 8-unit mean treatment difference of the change in hot flush score from baseline to month 24 with 80% power we will need 288 assessable patients per treatment arm (576 total). Sixty percent of patients enrolled in IBCSG Trial IX responded to the hot flush indicator at baseline and at month 24. If we assume a 60% compliance rate in Trial 24-02, then we would need a total of 960 patients randomized to two-arms to detect an 8-unit mean treatment difference with 80% power. Thus, 1440 patients are required for the three-arm trial.

Table 2 shows the total number of patients enrolled in two arms needed to detect a range of given treatment differences of the change in hot flush scores from baseline to month 24 with 80% power.

<table>
<thead>
<tr>
<th>Δ*</th>
<th>Total Assessed</th>
<th>Total Needed Assuming 60% Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>370</td>
<td>617</td>
</tr>
<tr>
<td>9</td>
<td>456</td>
<td>760</td>
</tr>
<tr>
<td>8</td>
<td>576</td>
<td>960</td>
</tr>
<tr>
<td>7</td>
<td>752</td>
<td>1253</td>
</tr>
</tbody>
</table>

*Mean treatment difference of the change in hot flush score from baseline to 24 months

The indicator for loss of sexual interest included in the trial-specific QL Module is the primary endpoint for testing whether patients receiving tamoxifen plus ovarian function suppression will report more sexual impairment than those receiving tamoxifen alone. This endpoint will also be assessed for the OFS plus tamoxifen versus OFS plus exemestane comparison. The use of the QL Module for IBCSG Trial 15-95 was the first time an indicator for the loss of sexual interest was collected in an IBCSG trial. Trial 15-95 had two treatment arms: Arm A (standard dose EC/AC x 4 -> CMFx3 -> Tamoxifen) and Arm B (high dose ECx3 -> Tamoxifen). With regard to addressing the short-term effects, these QL data have matured and a common standard
deviation of 35.6 was observed from calculating the mean treatment difference of the change in loss of sexual interest from baseline to 6 months. To detect a mean treatment difference of 8 units on the original scale with 80% power, 312 assessable patients are required in each group (624 total).

In Trial 15-95, 50% of the patients answered the loss of sexual interest question at baseline and at month 6. Assuming a 50% compliance rate in Trial 24-02, we would need a total of 1248 patients randomized in two arms to detect a mean difference of 8 units with 80% power. Thus, 1872 patients are required for the three-arm trial.

Table 3 shows the total number of patients enrolled in two arms needed to detect a given range of treatment differences with 80% power.

<table>
<thead>
<tr>
<th>Δ*</th>
<th>Total Assessed</th>
<th>Total Needed Assuming 50% Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>400</td>
<td>800</td>
</tr>
<tr>
<td>9</td>
<td>494</td>
<td>988</td>
</tr>
<tr>
<td>8</td>
<td>624</td>
<td>1248</td>
</tr>
<tr>
<td>7</td>
<td>814</td>
<td>1628</td>
</tr>
</tbody>
</table>

*Mean treatment difference of the change in loss of sexual interest score from baseline to 6 months

Although IBCSG 24-02 and IBCSG 25-02 include slightly different patient populations and treatment programs, both studies include comparisons of OFS plus tamoxifen versus OFS plus exemestane. We, therefore, intend to perform an analysis of the primary hypotheses of this QL study combining the information from both trials. This will enhance the power to detect QL differences between the two treatment groups.

8.2 Statistical Analysis

The primary analyses will be based on treatment differences at each QL assessment time point. Wilcoxon Rank Sum tests will be used to test for statistical significance. In addition, longitudinal data analysis techniques will be used to examine QL change over time. The longitudinal model accommodates informative censoring (dropout due to events that also affect QL such as relapse or death) and adjusts for data missing at random. We will be modeling jointly the QL measure and time-to-event, such as overall survival. The survival component of the model acts as a “missing data mechanism”. Intermittent missingness (missing data that occurs before the patient’s time-to-event) is considered ignorable in this model (missing at random).

We will be collecting information on reasons why the patient did not fill out the QL assessment form. This information will be used to impute QL scores for intermittent data not missing at random. A sensitivity analysis will be done in parallel with various imputation techniques to validate the effect that imputing values has on the parameter estimates. A SAS macro has been made available to estimate these longitudinal models of informatively censored data.
REFERENCES


