

# Health-related quality of life in early breast cancer

## Methodological and clinical studies

### Mogens Groenvold

This review has been accepted as a thesis together with nine previously published papers by University of Copenhagen 30 January 2009 and defended on 1 May 2009.

Official opponents: Mikael Rørth, Erik Lykke Mortensen, Niels Kroman.

Correspondance: Department of Palliative Medicine, Bispebjerg Hospital, Bispebjerg Bakke 23, DK-2400 Copenhagen NV, or Department of Health Services Research Institute of Public Health, University of Copenhagen, Øster Farimagsgade 5, P. O. Box 2099, DK-1014 Copenhagen K.

E-mail: mold@sund.ku.dk

Dan Med Bull 2010; 57: B4184

#### **The thesis is based on the following papers**

- I. Groenvold M. Quality of life in breast cancer adjuvant therapy: validation and pilot testing of a combination of questionnaires. *The Breast* 1997; 6: 97-107.
- II. Groenvold M, Bjorner JB, Klee M, Kreiner S. Test for item bias in a quality of life questionnaire. *Journal of Clinical Epidemiology* 1995; 48: 805-816.
- III. Groenvold M, Klee M, Sprangers MAG, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient/observer agreement. *Journal of Clinical Epidemiology* 1997; 50 (4): 441-450.
- IV. Groenvold M, Fayers P. Testing for differences in multiple quality of life dimensions: generating hypotheses from the experience of hospital staff. *Quality of Life Research* 1998; 7: 479-486.
- V. Klee M, Groenvold M, Machin D. Quality of life of Danish women: population-based norms for the EORTC QLQ-C30. *Quality of Life Research* 1997; 6: 27-34.
- VI. Groenvold M, Fayers PM, Sprangers MAG, Bjorner JB, Klee MC, Aaronson NK, Bech P, Mouridsen HT. Anxiety and depression in breast cancer patients at low risk of recurrence compared with the general population - a valid comparison? *Journal of Clinical Epidemiology* 1999; 52: 523-530.
- VII. Groenvold M, Fayers PM, Pedersen MA, Sprangers MAG, Aaronson NK, Mouridsen HT. Breast cancer patients on adjuvant chemotherapy report a wide range of problems not identified by health-care staff. *Breast Cancer Research and Treatment* 2007; 103: 185-195.
- VIII. Groenvold M, Fayers PM, Petersen MA, Mouridsen HT. Chemotherapy versus ovarian ablation as adjuvant therapy for breast cancer: impact on health-related quality of life in a randomized trial. *Breast Cancer Research and Treatment* 2006; 98: 275-284.

- IX. Groenvold M, Petersen MA, Idler E, Bjorner JB, Fayers PM, Mouridsen HT. Psychological distress as a predictor of recurrence and survival in primary breast cancer. *Breast Cancer Research and Treatment* 2007; 105: 209-219.

Parts of papers I-III were included in the PhD thesis Validation of quality of life questionnaire for breast cancer patients (Groenvold M, University of Copenhagen, 1996).

#### **ABBREVIATIONS**

CEF	Cyclophosphamide, epirubicine, fluorouracil
CMF	Cyclophosphamide, methotrexate, fluorouracil
DBCg	Danish Breast Cancer Cooperative Group
EORTC	European Organization for Research and Treatment of Cancer
EORTC QLQ-C30	EORTC Quality of Life Questionnaire-Core-30
HADS/HAD Scale	Hospital Anxiety and Depression Scale
HRQL	Health-related quality of life
OS	Overall survival
RFS	Recurrence-free survival

#### **1. BACKGROUND AND INTRODUCTION**

A cancer diagnosis has tremendous consequences for most persons who experience it. In the case of breast cancer the initial treatment usually consists of surgery, and after the operation many patients are recommended one or more additional treatments including radiotherapy, chemotherapy, and hormonal treatment. All these factors may, of course, impact the patients' quality of life.

This thesis deals with the scientific challenges and clinical results of a study aiming at assessing the impact of breast cancer and its treatment on the patients' quality of life.

Studies of the nature, prevalence, and intensity of problems and symptoms experienced by the patients are often referred to as health-related quality of life (HRQL) research.

HRQL research deals with subjective experiences and poses many challenging scientific questions. Therefore, in the clinically motivated study reported here much attention was directed towards methodological issues.

##### **1.1 Epidemiology of breast cancer**

Breast cancer is the most common cancer in women and the incidence of the disease has been increasing for several years. In 2003 breast cancer was diagnosed in 4,044 women in Denmark [1] and in 2005 breast cancer accounted for the death of 1,255

women [2]. A woman living in Denmark has an 8.9% risk of breast cancer [3]. Breast cancer is rare in men: the prevalence is less than 1/100 of the prevalence in women, corresponding to about 30 new cases per year [1]. Most women diagnosed with breast cancer have 'locoregional disease' (as opposed to metastatic disease) meaning that the disease is still 'local' or 'regional'; there is no evidence of distant metastases. This does not, of course, preclude that there may be microscopic metastases. This thesis deals with locoregional breast cancer in women.

### 1.2 Treatment of breast cancer

The treatment of primary, locoregional breast cancer consists of surgery with or without additional *adjuvant therapy*. Surgery is performed to remove the breast tumour and metastases in local lymph nodes, and involves either tumourectomy (also called lumpectomy) or mastectomy (removal of the breast). Surgery is the most important part of the treatment of breast cancer.

Adjuvant therapy has the aim of curing some patients who would otherwise die from recurrence of breast cancer or delaying such recurrence. Clearly, there is no point in giving adjuvant therapy if the patient has already been cured via the operation. Therefore, the need for adjuvant therapy is elucidated by examination of a number of prognostic factors. During the last decades the prognostic factors used most widely have been metastatic spread to the axilla, tumour size, the tumour's content of hormonal receptors, and its malignancy. Patients who based on these variables have been classified as being at low risk of recurrence have not been offered any adjuvant therapy whereas high-risk patients have been offered such treatment.

Adjuvant therapy includes *local radiotherapy* against the breast area [4] and *systemic treatments* against (micro)metastases, which may have spread in the body. Systemic adjuvant therapy includes endocrine therapy (treatments aimed at suppressing the effect of oestrogen), chemotherapy (cytotoxic drugs, often given in combination) and, relatively recently, monoclonal antibodies such as trastuzumab [5-7]. In some instances neo-adjuvant therapy has been used before surgery but usually adjuvant therapy is given after the operation.

A range of clinical and pathological variables are used to guide the choice of adjuvant therapy, including the tumour's hormone receptor status, and HER-2 protein, and whether the woman is premenopausal or postmenopausal.

Around 1990, when this study was initiated, the value of combination chemotherapy was well proven [8]. In Denmark the combination CMF (cyclophosphamide, methotrexate, fluorouracil) was considered the standard therapy, mainly for premenopausal women at high risk of recurrence [9]. In other parts of the world alternative combinations, mainly those including anthracyclines, were considered the standard. Postmenopausal women at high risk of recurrence were generally offered tamoxifen although subgroups were offered chemotherapy.

Changes in chemotherapy since the initiation of the DBCG 89 Program [9] will be discussed in the two chapters dealing with chemotherapy studies.

The treatment of primarily metastatic breast cancer and recurrent breast cancer is different from that of primary locoregional breast cancer and is outside the scope of this thesis.

### 1.3 The DBCG 89 studies

The Danish Breast Cancer Co-operative Group (DBCG) was established in 1977. It is one of the first examples of a nationwide collaboration between the surgical, medical, oncological, patho-

logical, and radiological hospital departments involved in the treatment of a disease [10]. DBCG has developed guidelines and protocols for randomised trials, and was one of the first examples of the development and successful implementation of national guidelines standardising the treatment of a disease [10]. In 1989, when DBCG released its DBCG 89 Program, it included guidelines for diagnosis and treatment of primary breast cancer as well as three randomised trials [9].

These guidelines included the definitions of the group of patients considered low risk, i.e. those who were likely to have been cured through surgery, and those considered high risk, i.e. with a risk of breast cancer recurrence justifying additional, systemic treatment [9].

DBCG 89 A was the protocol describing the follow-up program for low risk patients not offered any systemic adjuvant therapy. Subgroups of the patients were offered local radiotherapy. The protocol did not involve randomisation.

Briefly, the three randomised trials had the following research questions. The DBCG 89 B trial randomised premenopausal women with receptor-positive tumours between standard CMF chemotherapy and ovarian ablation. It had been suggested that among premenopausal women with receptor-positive tumours the effect of chemotherapy was mediated via its reduction of hormone production in the ovaries [11, 12] rather than a cytotoxic effect. The research question was mainly whether ovarian ablation was as effective as chemotherapy [9].

The DBCG 89 C trial included postmenopausal women in a trial comparing three endocrine regimens. The standard at that time was tamoxifen for one year. This standard was compared with two years of tamoxifen and with six months of tamoxifen followed by six months of megestrol acetate. Thus, this trial compared two durations of tamoxifen therapy and compared the combination of two drugs against one drug.

The DBCG 89 D trial had a 2x2 design, i.e., it had two research questions and included a double randomisation resulting in a total of four treatment arms. The trial included premenopausal and postmenopausal patients who, in general, were at relatively higher risk of recurrence than the patients allocated to the two other trials. The first research question was whether the standard chemotherapy regimen CMF could be improved by exchanging one of the three drugs with another, i.e., CEF. The other research question was whether the drug pamidronate could reduce the risk of or the morbidity from bone metastases.

The DBCG 89 protocols are described in more detail in the Methods section.

### 1.4 Reasons for assessing HRQL in the DBCG 89 studies

It was well known that patients diagnosed with and treated for breast cancer might experience many different symptoms and problems. There were three main reasons for assessing HRQL in the DBCG 89 protocols.

#### 1.4.1 End-points in randomised trials

As outlined above, DBCG 89 B investigated whether ovarian ablation had the same effect on survival as chemotherapy. The idea was that if the treatments had a similar anti-tumour effect then it might be preferable for the patient to avoid chemotherapy. On the other hand, during the discussions when the HRQL was planned it was also suggested that ovarian ablation might be worse than chemotherapy: the menopause and sterility induced by ovarian ablation was permanent whereas in some patients treated with chemotherapy menstruation may persist or return.

Although ovarian ablation was briefer and thought to be associated with fewer side effects, it was argued that these patients might for example have a greater risk of depression in the following years. These considerations motivated a comparison of the HRQL outcomes in the two treatment arms.

No HRQL studies of tamoxifen or megestrol acetate had been conducted (paper I). The evidence concerning tamoxifen was about side effects (i.e., not from studies based on patient-report) and some was almost at the anecdotal level but nevertheless it was frequently mentioned that tamoxifen was associated with depression [13, 14]. If the combination of two drugs were shown to improve survival, this combination could be a new standard recommended to millions of future patients and it would be of great interest to know whether a gain in survival probabilities was accompanied by better or worse HRQL outcomes. The same could be said about the comparison of two durations of tamoxifen. Year two of the study where one group had completed tamoxifen treatment while the other was still on this treatment gave the opportunity to study the HRQL associated with tamoxifen treatment in a randomised trial. Thus, it was of interest to use the DBCG 98 c trial to investigate whether there was a difference in the HRQL impact between tamoxifen and megestrol acetate, and whether the patients randomised to two years of tamoxifen had worse HRQL in the second year than those treated for one year only.

Finally, the DBCG 89 d protocol made an HRQL study highly relevant. The trial compared the standard CMF chemotherapy with the CEF regimen. It was well known that CEF was more frequently associated with alopecia whereas it was unknown whether there were other differences in HRQL outcomes. Clearly, if it were shown that CEF was more effective than CMF it was important to know whether a potential gain in survival was 'paid for' by worse HRQL.

Two levels of use of HRQL data in the interpretation of randomised clinical trials can be listed. First, and simpler, the researchers can use this information in their interpretation of results. If, for example, the available research data show no difference in survival between two treatments with regard to survival but the HRQL data show a clear advantage, then researchers can conclude that this is an argument in favour of the mildest treatment. Second, the availability of HRQL can be used as a means of sharing the total information about treatments with patients and thus as a means of providing the patients access to more insight in the results forming the basis for treatment decisions. This is because the HRQL data may convey information about consequences of treatment that would otherwise be part of the doctors' overall evaluation of 'what is best for you'. In other words, HRQL data might be a way of obtaining a better basis for decision-making because important information – which otherwise would be undocumented – could become accessible for the patient. The research questions posed in the three trials of the DBCG 89 Program thus clearly motivated HRQL assessment but no study of HRQL was included when the protocols were designed. Shortly after the publication of the DBCG 89 Program I approached the DBCG and proposed to assess HRQL in the trials. The proposal was received very positively and a parallel 'add-on-study' of HRQL was launched when funding from the Danish Cancer Society had been obtained.

#### 1.4.2 Descriptive information

The second category of arguments for assessing HRQL in the DBCG 89 Program concerned the opportunity of using it as a

means of obtaining descriptive information about the longitudinal impact of both the disease and its treatment on HRQL. There were two main ways to use information about the frequency and course of the various symptoms and problems following breast cancer diagnosis and treatment:

- A. To be able to inform future patients about the consequences of the disease and the various treatments, and
- B. To give health care professionals insights which could be used to alleviate or prevent symptoms and problems (this also includes the potential use in continuous quality development).

Not only the randomised trials but also the very detailed guidelines standardising the treatment procedures across the country served to improve the opportunities for obtaining useful information.

Another important point is that knowledge about the patient-experienced consequences of treatments (and relevant treatment alternatives) is a necessary basis for the informed consent required by Danish law (Patientrettighedsloven, Lov om patienters retsstilling, lov nr. 482 af 01/07/1998; Sundhedsloven, Lov nr. 546 af 24/06/2005, [www.retsinfo.dk](http://www.retsinfo.dk) accessed June 2007).

Further, knowledge about likely consequences of treatments may make the patient feel safe because she knows what is going to happen and can prepare herself for this. By this it is not meant that all patients should always be given the maximal amount of information – this may neither be desirable nor practically possible – but the information given to each patient should be based on knowledge that is as scientifically sound as possible.

Thus, HRQL data might improve information to patients, might facilitate greater patient involvement in treatment decisions ('empowerment' via access to information), and might serve as a basis for better prevention or alleviation of symptoms and problems.

#### 1.4.3 Investigation of the psychosocial consequences of cancer

At a more general level, a longitudinal study of a large group of breast cancer patients using relevant questionnaires was anticipated to be able to elucidate questions of general scientific and clinical interest. Relatively little was known about the course of the various consequences of the disease and treatment over time. Little was known about differences between sub-groups of patients (e.g., younger versus older, more or less affluent patients, and between patients differing with regard to social network). Comparisons of sub-groups could clarify which patients managed the situation the best and the worst and information could be used to identify groups of patients in need of additional care.

Another, more basic research question, which could be elucidated via HRQL data, was whether there was any association between psychological distress and the risk of death from cancer. At the time of initiation of this study there was evidence of an association between self-rated health and survival in general population studies [15, 16]. Furthermore, Spiegel's randomised study published in 1989 [17], which indicated that metastatic breast cancer patients taking part in support groups had better survival, had generated renewed interest in the possible relationships between psychological distress and breast cancer survival.

In sum, there were strong arguments for assessment of HRQL in the DBCG 89 protocols.

### **1.5 Breast cancer from the patient perspective and HRQL evaluation**

Seen from the patient's perspective, a diagnosis of breast cancer may have multiple implications. It may be viewed as a sudden, unexpected threat to life, may cause acute hospitalisation, usually involves surgery with the removal of a breast or part of a breast, creates a need for medical decisions, may necessitate additional treatments, and may give rise to symptoms and practical problems. These and many other factors may cause an acute and severe disruption of the patient's daily life [18]. All this creates a strong need for mental adaptation, which it is hoped will lead to successful readjustment to a new situation. Thus, for many patients a diagnosis of cancer is a turning point in their life: habits and daily life activities are reviewed and are possibly changed. All these aspects may be investigated in various research projects but clearly a single study may elucidate only parts of the experience of breast cancer.

The present study falls within the category of 'health-related quality of life' (HRQL) research. Initially, the term 'quality of life research' was used when describing medical studies of patients' experiences of disease and treatment but recognising that many aspects of quality of life are unrelated to health, the term HRQL became preferred [19, 20].

There is no single, universally accepted definition of HRQL assessment but '... there seems to be an emerging consensus that generic HRQL takes into account levels of physical, mental, social, and role functioning, and includes abilities, relationships, perceptions, life satisfaction, and well being.' [20]. HRQL assessment is thus based on the WHO definition of health [19]. A fundamental characteristic of HRQL assessment (in contrast to 'toxicity rating' carried out by physicians) is that it is preferably based on patient self-report [21, 22].

When this study was initiated it was viewed as controversial whether the subjective experience resulting from breast cancer and breast cancer treatment could be investigated via questionnaires in a way that was sufficiently robust seen from a scientific point of view to allow such results to influence decision-making and clinical practice. I was often challenged when reading the scientific literature, following the debate in the field, and when presenting the project to colleagues and research partners. Some of the objections were:

- All patients react differently to cancer; their reactions are subjective and fluctuating; it is impossible to investigate this scientifically (clinicians).
- A questionnaire does not produce anything that can be used scientifically; we all know that when completing a questionnaire we tick some boxes but we could equally well have ticked other responses – much of it happens arbitrarily or at random, and the process is subject to all kinds of different and uncontrollable bias. A questionnaire cannot produce valid data (clinicians).
- Quantitative research methods such as questionnaires are not suitable for assessment of subjective experiences or, more generally, quality of life. Qualitative methods are needed; theoretical frameworks must be developed. Otherwise, results will be useless and potentially misleading (psychologists, etc.).
- It is practically impossible – with the resources potentially available to such a project – to carry out a longitudinal questionnaire study involving large numbers of patients across the entire country; it will not be feasible to identify the patients at the right time, to get their consent, or to organise

the collection of questionnaires at the right time (various colleagues).

- The current methodology applied to analysis of questionnaires is misleading and outdated; instead, newer statistical methods (which at that time were virtually unknown to almost all leading scientists in the field) have to be used (statisticians).

Given the many arguments in favour of conducting a large study of HRQL in the DBCG 89 Program, I took the objections seriously and discussed them and the methodological challenges with advisors and colleagues. The resulting research plan was an attempt at establishing a study that could provide results that were useful in relation to the research questions, that overcame the practical obstacles, and that at the same time investigated the scientific quality of the results, i.e. their validity and reliability. When initiating the study the problems around delineation of the field of enquiry (i.e., that it could rightly be argued that it was impossible to assess a huge and ill-defined concept such as 'quality of life') led to the following definition of aims in the clinical research protocol: '... to describe how, how much, and for how long the quality of life is affected by each kind of adjuvant treatment...' [23](p. 8). A quality of life study was defined as 'a mapping of treatment-related physical and psychological symptoms and effects on social, sexual, and work-related matters' [23](p. 8). It was added that 'The term 'quality of life' is thus used in a relatively narrow meaning. General investigation of the quality of life concept is not central to the research project. It is concerned with the assessment of a number of matters that are significant to quality of life' (p. 8).

As stated above, the concept HRQL became widely used at a later stage with the same motivation, i.e. to use a more specific and less pretentious term than 'quality of life' [24, 25]. As stated by Ferrans in a recent review, '... the term HRQL draws a line between those facets of life that are primarily health related and those that are not.' [24](p. 14-15). Thus, the initial conceptualisation made in the present study was in line with the subsequent development in the research field.

The study has resulted in publications investigating HRQL in a general population sample (paper V), a paper studying psychological distress in breast cancer patients compared to the general population (paper VI), and papers on the impact of chemotherapy compared to no adjuvant therapy (paper VII) or versus ovarian ablation on HRQL (paper VIII). It was also investigated whether psychological distress in newly diagnosed breast cancer patients was related to survival (paper IX). Based on this study an article investigating whether operation type (mastectomy or lumpectomy) was related to social class [26], a book chapter investigating whether there were social differences in the reactions to breast cancer chemotherapy [27], and a methodological article partly based on this study [28] were written; these publications are not included in the thesis. The same is the case, of course, for a Master's thesis [29] and a PhD thesis [30] using data from the study. The methodological parts of the study, which were added after the clinical HRQL study had been implemented, are introduced in the following sections.

### **1.6 Validity and reliability in HRQL research**

This section briefly reviews some of the concepts related to validity and reliability in HRQL research. The concepts were explored in more detail in my PhD thesis [31] and are extensively described in the literature [25].

Validity refers to the truth of scientific results or statements. All scientific fields have their approaches to assessment of validity and reliability. In HRQL a typical definition is 'Validation of instruments is the process of determining whether there are grounds for believing that the instrument measures what it is intended to measure, and that is useful for its intended purpose.' [32](p.45). Validity can be viewed as absence of systematic error. In contrast, reliability refers to absence of unsystematic error. This means that while validity problems will influence the results of a scientific study irrespective of its sample size, suboptimal reliability can be compensated for by a sufficient sample size. Many different terms are used to categorise the approaches used to validate questionnaires in HRQL research. Useful overall categories are content, construct, and criterion validity [25]. These terms are defined and discussed in detail in my PhD thesis [31].

### 1.6.1 Content validity

Content validity refers to the extent to which the questionnaire has the content needed to elucidate the research question. This implies that content validity (like other aspects of validity) is not an ability that a questionnaire can possess (it is often stated in the literature that 'this questionnaire has proven validity and reliability'); instead, content validity is related to a specific application of a questionnaire. For example, a questionnaire may have a high degree of content validity when used to assess the symptoms resulting from one chemotherapy regimen while it may have poor content validity when used to evaluate another chemotherapy regimen if it misses the main problem resulting from that chemotherapy regimen, e.g., neurotoxicity, as discussed in relation to palliative care trials [33].

The work aimed at assuring the content validity of a questionnaire usually includes a literature review combined with interviews with patients and health care professionals. The overall research question for the study should be used as the delineation of the literature review and as the basis for the questions asked in the interviews, for example 'which consequences do patients experience as a result of the disease or treatment?' Many consequences may be identified, and to select which of these to include in the questionnaire it is often desirable to obtain ratings of the relevance and importance of the issues from relevant patients. Paper I reports the work aimed at developing a content valid questionnaire for this study.

### 1.6.2 Construct validity

Construct validity concerns the constructs (concepts) used in the study or the research field. It is thus a theoretical way of approaching the validity discussion. However, in practice the theoretical questions are often not formulated and instead, standard statistical manoeuvres are often carried out and interpreted as numbers without proper acknowledgement of their meaning and theoretical justification.

Construct validity may concern important aspects related to the construction of multi-item scales. There are three main reasons for making multi-item scales: (a) to reduce measurement error (i.e., increase reliability), (b) to reduce the number of variables in the statistical analysis (often a careful attempt at obtaining good content validity results in a large number of items, which may result in an excessive number of results and problems resulting from multiple hypothesis testing), or (c) because the concept in question is best measured via multiple questions (e.g., one may want to capture various aspects of depression).

Irrespective of the reasons for construction of multi-item scales and the many advantages they may produce, there is a considerable risk that multi-item scales may lead to loss or distortion of information obtained by the items. It is problematic if important information about the research question disappears or is modified during the transition from items to scales. If, for example, we want to know the consequences of a new kind of chemotherapy and an item on dizziness shows that patients experience this problem, then it is problematic if this symptom is overlooked because we have analysed the dizziness item as part of a 'symptom scale' where the effect on dizziness is diluted and we therefore incorrectly conclude that the treatment is not associated with any symptoms.

There are several other potential problems associated with the creation and use of multi-item scales and still such multi-items scales are usually necessary. One of the newer approaches to the validation of multi-item scales is analysis for differential item functioning (DIF), previously called item bias analysis. In contrast to the traditional approach to construct validation, where one or more separate 'validation studies' are performed and are later referred to as proper justification of 'construct validity' or 'psychometric robustness' of the questionnaire, DIF analyses have the advantage of being able to examine the multi-item scales specifically in relation to particular research questions.

Paper II is an application of DIF analysis to one of the questionnaires used in this study and examines the ability of the questionnaire to compare groups varying with regard to treatment and age. DIF analysis was also used in the studies reported in papers VI, VII, and VIII. The results were not included in the published papers due to space restrictions but are included in this thesis. An entirely different way of approaching construct validity testing was also used in this study. While the researcher can make sure that the relevant items are included in the questionnaire and can make sure that multi-item scales do not distort the information obtained in the individual items, an additional, important question may be raised: do patients give the right answers when they complete the questionnaire?

Answers to questions about subjective matters do not exist before the question is asked; they are constructed by the individual through complicated processes [34]. Patients may misunderstand the questions asked, they may misunderstand the response categories or the way they relate to the question, or they may in error tick the wrong response options. Furthermore, patients may understand the questions and response options differently from that intended – not due to errors or misunderstandings, but simply because their reality is different from that of a healthy, academically trained researcher who has thought and read about the issues for months or years.

These considerations could be summarised into a basic question of whether questionnaire items are a valid way of obtaining information about the topics they are supposed to measure. Does our item on sleeplessness give valid insight into breast cancer patients' problems with sleeplessness? It is easy to imagine numerous sources of error. Paper III describes a method developed to elucidate whether patients understand questionnaire items in the same way as do the researchers conducting the study. If this were the case, it would be unlikely that major errors occurred during patient completion of the questionnaire. Additional results not included in Paper III are included in this thesis.

### 1.6.3 Criterion validity

Criterion validity is usually the third way of approaching the validity of HRQL questionnaires. The idea is that if an external criterion is available then the validity of the questionnaire can be measured directly against this criterion. However, when the questionnaire is used to measure symptoms and experiences such criteria are rarely available. However, if for example a questionnaire is used to determine whether patients are depressed, an interview with a psychiatrist can be used as a criterion.

Criterion validity in the traditional sense was not evaluated in this research project but the study comparing patients' responses to the questionnaire against data based on an interview (i.e., using the interview results as the criteria) can be viewed as an assessment of criterion validity.

### 1.7 Problems related to lack of a priori hypotheses and multiple significance testing

One of the basic principles of statistics is that the statistical methods should be used to test hypotheses – not to trawl the data searching for 'significant' associations. It follows from this that hypotheses should be formulated a priori, i.e., before the data is collected. A closely related principle is to limit the number of statistical tests carried out in a data set. Otherwise, problems of multiple hypothesis testing may occur (see also paper IV).

The present study is an example of the difficulties one may encounter when implementing statistical principles in clinical research. Many of the planned comparisons of groups had never been done before and therefore the basis for formulating a priori hypotheses was sparse. Furthermore, a questionnaire aiming at covering as many of the relevant symptoms and problems as possible would naturally contain a large number of variables. And on top of this, it was planned to follow patients over time, so six measurements of each variable would be available.

Two different approaches to these problems were applied in this study. Concerning one of the main research questions of the study – which aspects of HRQL are affected by chemotherapy? – there was considerable literature available and thus it was possible to use this literature to formulate hypotheses (papers I and IV). These hypotheses were not formulated in the original protocol as usually required in order to be a priori hypotheses, but were subsequently extracted from the literature review used to compose the questionnaire (paper I). Thus, they were a priori formulated in the sense that they were based on data collected before the study.

Another solution was explored in relation to some of the other planned comparisons. Given that there were no published studies having compared for example chemotherapy to ovarian ablation, it was difficult to formulate well-motivated hypotheses. A staff survey was conducted to elucidate whether health care professionals treating breast cancer patients had expectations that could be used to formulate hypotheses that could guide the statistical analysis (paper IV).

## 2. AIMS

The overall aims of this study were to evaluate the impact of early breast cancer and adjuvant therapy on health-related quality of life (HRQL) and to assess whether psychological distress had prognostic significance. This involved the following specific aims:

- 1) To compose a questionnaire measuring the impact of early breast cancer and adjuvant therapy on health-related quality of life (paper I) and to employ this questionnaire longitudinally in breast cancer patients.

- 2) To investigate whether the multi-item scales included in the questionnaire were adequate representations of the information collected through their items (paper II).
- 3) To investigate whether patients understood and responded to the items of the questionnaire in the same way as did the researchers (paper III).
- 4) To investigate whether the views and experiences of health care professionals are useful in handling problems related to hypothesis testing in the analysis and interpretation of health-related quality of life data (paper IV).
- 5) To facilitate the interpretation of results from breast cancer patients: to use the same questionnaire(s) to investigate the HRQL of a sample of women from the general population (papers V and VI).
- 6) To investigate the prevalence of anxiety and depression in newly diagnosed breast cancer patients as compared to women selected randomly from the general population (paper VI).
- 7) To investigate whether there are differences in HRQL between premenopausal low-risk patients not offered any systemic therapy and patients on chemotherapy (paper VII).
- 8) To investigate whether there are differences in HRQL between premenopausal patients with receptor-positive tumours randomised to chemotherapy or ovarian ablation (paper VIII).
- 9) To investigate whether psychological distress and other HRQL variables carry prognostic information independent of biological variables (paper IX).

## 3. PATIENTS AND METHODS

### 3.1 Design

This was a prospective, longitudinal questionnaire-based study of (1) consecutive patients included in the DBCG 89 A protocol for follow-up of low-risk patients, and (2) consecutive patients randomised in the trials in DBCG 89 protocols B, C, D [9]. A cross-sectional study of Danish women randomly selected from the general population and a small, cross-sectional survey of health-care professionals were also included.

### 3.2 The DBCG-89 Protocols

The DBCG 89 Program for Treatment and Follow-Up of Patients with Primary, Operable Breast Cancer [9] contains guidelines for the surgical, medical, and oncological therapy of breast cancer. It also includes guidelines for follow-up, for pathological procedures, and a detailed description of the various tests and examinations involved in the diagnosis of early breast cancer.

#### 3.2.1 Inclusion criteria

The protocol had the following general inclusion criteria [9, 10]:

- 1) Female less than 75 years
- 2) Primary, unilateral, histologically proven breast cancer, excluding in situ carcinomas and inflammatory cancer, treated with lumpectomy or mastectomy and axillary dissection
- 3) No prior neoplastic disease (except cutaneous cancer and cervical cancer *in situ*).

For patients fulfilling the general inclusion criteria, the DBCG 89 Program provided a decision-sheet to determine risk of recurrence and the adjuvant systemic therapy. Using this sheet any patient could be placed in one of four categories. Protocols A, B,

C, and D determined the treatment and follow-up of these patients.

In contrast to systemic therapy, local treatment did not depend on protocol allocation but was determined by common guidelines. Local radiotherapy against the residual breast was offered to patients who had undergone lumpectomy (breast-conserving therapy with removal of the tumour). Local radiotherapy was additionally offered to patients who were up to 45 years old and had four or more positive lymph nodes, and to all patients whose tumour had not been radically removed.

Patients allocated to protocol A were viewed as low-risk patients and were not offered any systemic therapy. These patients had tumour-negative axillary nodes and tumours up to 50 mm. Most hospitals also required that premenopausal women had histological grade I (low-grade malignancy) tumours. Of the 59 hospitals reporting patients to DBCG, 50 agreed to inform patients about the present study and only Protocol A patients from these hospitals were included in the HRQL study.

Each of the three other protocols described the standard systemic adjuvant therapy for the particular sub-group of breast cancer patients and included a randomised trial comparing this standard therapy to one or more other treatment regimens. Patients allocated to one of these protocols were informed about the randomised trial at the department taking care of adjuvant therapy. The patients accepting randomisation were subsequently randomised by telephoning the DBCG Secretariat. Patients not accepting randomisation were offered the standard therapy.

In addition to the general inclusion criteria, the specific inclusion criteria for the protocols were:

DBCG 89: Premenopausal, node-positive, and receptor-positive

DBCG 89 C: Postmenopausal, node-positive, and receptor-positive/unknown

DBCG 89 D: Premenopausal, node-positive, and receptor-negative/unknown, premenopausal, node-negative, and histological grade II-III (medium-high grade malignancy) (most hospitals), or postmenopausal, node-positive, and receptor-negative.

### 3.2.2 Treatments

The randomised trials in the three protocols [9] were:

DBCG 89 B: (1) Standard CMF (cyclophosphamide, methotrexate, fluorouracil) chemotherapy versus (2) ovarian ablation. CMF was given as nine cycles of intravenous cyclophosphamide 600 mg/m<sup>2</sup>, methotrexate 40 mg/m<sup>2</sup>, 5-fluorouracil 600 mg/m<sup>2</sup> every three weeks. Ovarian ablation was irradiation (five doses of three Gy against the pelvic region) or (rarely) surgical oophorectomy

DBCG 89 C: (1) Standard tamoxifen 30 mg daily for one year versus (2) tamoxifen for 2 years versus (3) tamoxifen for 6 months followed by megestrol acetate 160 mg daily for 6 months

DBCG 89 D: (1) Standard CMF (cyclophosphamide, methotrexate, fluorouracil) chemotherapy versus (2) CEF (cyclophosphamide, epirubicin, fluorouracil) chemotherapy. The CMF regimen was the same as in Protocol b. CEF was given as CMF with methotrexate substituted by epirubicin 60 mg/m<sup>2</sup>. In addition, this protocol randomised patients between no additional therapy (1 or 2) versus oral pamidronate 150 mg twice daily for four years (arms 3 or 4).

### 3.3 Relationship between study populations and the nine papers

The relationship between the nine papers included in this thesis and the study populations is shown in Fig. 1. Thus, papers II (DIF analyses), III (validation), and IX (survival) were based on patients from all DBCG 89 protocols. Paper VI (low-risk patients versus general population sample) included breast cancer patients from Protocol A. Paper VII (CMF chemotherapy versus no chemotherapy) included premenopausal patients from Protocol A (control group) and premenopausal patients randomised to CMF chemotherapy in Protocols B and D. Finally, paper VIII included patients from Protocol B.

### 3.4 Development, composition, and pilot testing of questionnaire

The development of the questionnaire to be used in this study is the subject of paper I, which includes a detailed description. The development took place as summarised below.

#### 3.4.1 Literature review

The literature was searched for publications describing the quality of life impact of breast cancer adjuvant therapy. The review was based on MEDLINE searches, reference lists of identified articles, and other sources. Papers dealing with chemotherapy, endocrine therapy, and ovarian ablation were identified. From each article, data about patient-reported negative effects of the treatments were extracted, and a list summarising the results was made. Because no articles dealing with ovarian ablation or endocrine therapies were identified in the literature review a gynaecologist was consulted about whether any likely effects of these treatments were missing from the list of issues made from the literature review.

In order to avoid unimportant issues, the list resulting from the literature review was examined in the interviews described below, and issues not considered severe or frequent were removed. The literature review also included a review of existing questionnaires that could be used for breast cancer patients.

#### 3.4.2 Interviews with patients

A convenience sample of 14 breast cancer patients attending the outpatient clinic at the Department of Oncology, State University Hospital (Rigshospitalet) was interviewed. The interviews consisted of two parts. First, in an open (qualitative) part, interviewees were asked about how they experienced adjuvant therapy and how it affected their daily lives. After having completed this description, they were asked to nominate the three most important negative effects of adjuvant therapy. In a second, structured part of the interview, interviewees were asked to what extent they had been bothered by each of the issues on the list developed in the literature review. Finally, 8 of the 14 patients who had filled in the preliminary version of the questionnaire (described below) were interviewed about the acceptability of the questionnaire.

#### 3.4.3 Construction of questionnaire

The construction of the questionnaire was based on the review of the HRQL impact of adjuvant therapy as well as the interviews. The existing questionnaires were reviewed and new items were developed. When constructing the items the same simple and brief structure as used in the EORTC QLQ-C30 was used when possible. However, changes to the structure or the response categories were made if this was thought to improve the items. It quickly became clear that the questionnaire would become rela-

Table 1

Relationship between study population and the nine papers.

Paper	DBCG 89 protocol and treatment arm											General population	Convenience sample	Nurses, doctors			
	A pre	A post	B1	B2	C1	C2	C3	D1	D2	D3	D4						
I																	
II	•	•	•	•	•	•	•	•	•	•	•	•					
III	•	•	•	•	•	•	•	•	•	•	•	•					
IV																	•
V														•			
VI	•	•												•			
VII	•							•									
VIII			•	•													
IX	•	•	•	•	•	•	•	•	•	•	•	•					

Pre: premenopausal; Post: postmenopausal

tively long, and therefore, each issue was represented with one item only, except if it was judged that more items were needed to measure the concept adequately.

#### 3.4.4 Pilot study

The preliminary questionnaire was pilot tested in 84 breast cancer patients at the outpatient clinic at the Department of Oncology, State University Hospital (Rigshospitalet). All patients with a planned visit within a one-week pilot study period were eligible. Terminal patients, patients visiting the clinic for the first time, and patients above 75 years of age were, however, not included. In addition to the 74 patients fulfilling these criteria, 8 patients from the clinic, who were interviewed later, were also included. The questionnaire was sent to the patients by post with an accompanying letter asking them to complete the questionnaire at home and bring it with them to their visit to the clinic a few days later. The questionnaire also included a 'debriefing form' containing questions about the questionnaire.

The pilot study was also used as the basis for a small 'known-groups comparison' [25, 35] in order to test whether the questionnaire could detect differences between patients in chemotherapy (N=23) and patients not receiving any treatment (N = 23).

#### 3.4.5 Sociodemographic variables

In addition to the HRQL questionnaire described above a brief questionnaire was constructed to collect information on marital and cohabitation status, number of children, and education. The social class classification developed by the Danish Social Research Institute was used, and items to collect the relevant information for this were made [36]. Based on these data social class was assigned 'manually' to each participant in the breast cancer and the general population (see below) studies. The social class classification has five levels ranging from V (unskilled worker) to I (the most affluent; includes academics and groups of self-employed and employed persons) [36, 37]. In the coding, the 'family social class' (as recommended in [36](p. 15)) was used for married, cohabiting, or widowed women: social class was determined both for the woman and for her husband/cohabitant. Each woman was then assigned the higher of the two values [36](p. 14).

#### 3.4.6 Adaptation of the questionnaire to the general population study

The basis for the general population study was the questionnaire developed for breast cancer patients. However, items that were obviously related to cancer treatment and might give the respon-

dents an impression that they were suspected of having a disease were omitted.

### 3.5 The questionnaire study in DBCG 89

#### 3.5.1 Inclusion of patients

The present study included consecutive patients registered in protocol A, as well as consecutive women randomised in the three protocols 89 B, 89 C, and 89 D. Accrual to the questionnaire study was initiated on 1 June 1991 and the goal was to include 100-150 fully evaluable patients in each of the 11 protocol arms [23, 31].

It could be problematic to send a letter with a questionnaire to a patient who was not prepared for this and who might be worried about how the information about her disease and treatments had become available to researchers at the University of Copenhagen. To prevent this problem, I contacted all surgical, medical, and oncological departments in Denmark who were involved in treatment of patients with primary breast cancer and asked them to hand out a written information sheet to all patients diagnosed with breast cancer. The departments involved in adjuvant therapy did this by adding the information about the questionnaire study to the standard information used to give information about the relevant DBCG protocol. The surgical departments, which were the vast majority, organised to hand out the information sheet as part of their routine. A total of 59 departments reported patients to the DBCG during the study period and 50 of these agreed to distribute this information, and Protocol A patients from these departments were included.

The 'initial information letter' explained that a questionnaire study was going on and that some patients would receive a letter with more details about this. It emphasised that the patient was not asked to make a decision as to whether she would participate at that time – the letter was informing about the possibility that the patient could be contacted only. The letter included the address of the office of the HRQL study and the information that if the patient did not want to receive the more detailed letter about the study she could indicate this and would thus not be contacted.

Every weekday during the inclusion phase, the DBCG Secretariat mailed a list of all patients registered in Protocol A or randomised in one of the three protocols to the office of the HRQL study at the Department of Social Medicine, University of Copenhagen. The design of the study determined that in order to get comparable results across the different protocols and treatment arms,

questionnaires had to be completed by the patients at the same point in time, measured from the date of diagnosis, irrespective of protocol. Therefore, patients registered by the DBCG Secretariat later than the planned date for the first questionnaire, i.e., 7 weeks postoperatively, were excluded from the HRQL study.

### 3.5.2 Questionnaire administration

#### Information letter

The questionnaire was sent to the patients by post. A patient information sheet explaining the purpose of the study, emphasising that participation was voluntary, that the patient could withdraw at any time without any consequences, and that the information they provided would be kept confidential, accompanied the first questionnaire. It was also stated that no information would be released from the questionnaire to the hospitals involved in the treatment and care of patients. Finally, the letter contained instructions about when the questionnaire was to be completed (see below). A stamped, addressed response envelope was enclosed.

#### Timing of questionnaires

The questionnaires were sent to the patients to be completed at 1, 3, 5, 9, 15, and 24 months after the date of randomisation. The questionnaires to patients in protocol A, who were not randomised, were sent at the same points in time, measured from the operation. To do this, the average time from operation to randomisation in protocols B, C, and D was determined.

The patients in chemotherapy were asked to complete the questionnaires seven days after they had their chemotherapy. The letters were sent out a few days before the estimated date of completion. All other patients were asked to complete the questionnaire as soon as possible. As a result, all patients in the study completed the questionnaires at the same number of days after their operation irrespective of which protocol they were allocated.

In the beginning of the study, the questionnaires were sent out based on preliminary estimations. After about two months and again 2-3 months later the schedule was reviewed by examining the data for all patients entered. The preliminary schedule was found to be very accurate in achieving 'simultaneous' completion of questionnaires across protocols, but a few, small revisions were made to optimise the schedule.

#### Reminders

Patients who did not return the questionnaires were sent reminders after two, four, and six weeks. The reminders were carefully written to emphasise that study participation was voluntary, and to take into account that some chemotherapy patients would have to wait for some time before completing the questionnaire. A questionnaire and a response envelope was enclosed with the first and third reminders.

#### Ethical committee approval

The Danish ethics committees approved the HRQL study (V.200.1873/90, V.200.2067/91).

## 3.6 General population study

### 3.6.1 Identification of study sample

A random sample of women living in Denmark was obtained from the Danish Central Population Register (CPR). All women who were born on a particular date in all odd years from 1913 to 1971 were identified. As described in paper V, a colleague conducted a parallel study, and the women identified from the CPR were randomly distributed between the two studies. Up to 200 patients in each 10-year age stratum were included in the present study.

### 3.6.2 Questionnaire administration

The women were contacted by post in April 1992 following the same procedures as for the breast cancer study (see above) except that they were sent only one questionnaire. Of course, the information was different and emphasised that we did not contact them because we thought they were ill. However, the women were encouraged to participate even if they were ill.

## 3.7 Analysis for differential item functioning (DIF)

The multi-item scales of the EORTC QLQ-C30 were analysed for DIF in relation to age and treatment (chemotherapy) using three-way contingency tables (paper II). A table was made for each combination of item and exogenous variable, controlling for scale score. The null-hypothesis of no association between item and exogenous variable after control for scale score was tested by calculation of the partial gamma [38]. The two-sided test probability for partial gamma equal to zero was found via Monte Carlo simulation (1000 simulations) using a computer program [39]. The same approach to DIF testing was used in each of the papers comparing groups of patients differing as to treatment (papers VI-VIII). The grouping variable used in each study was used as exogenous variable. In addition, age was also included as exogenous variable in paper VI. However, due to space restrictions these results were removed from the papers during the peer-review process. The results are summarised in this thesis (Appendix A).

## 3.8 Testing whether patients and researchers understand questionnaire items in the same way

The method was developed for this and a parallel study [40-43] in response to our concerns about the validity of patient-completed questionnaires (paper III). The principle was to compare patient responses to the questionnaire against an observer's rating of the same patients' open-ended responses to the same questions. The observer was the researcher who had composed the questionnaire. A high extent of agreement between the patient responses to the questionnaire given before the interview and the observer ratings would indicate that, in general, patients had understood the items in the same way as the observer and thus that the items were not to a large extent misunderstood or erroneously completed.

The study was carried out in collaboration between two studies, the present study and one including gynaecological cancer patients conducted by Marianne Klee. From the present study 57 patients, who had already completed one or two of the six sequential questionnaires were randomly selected. In addition, 88 gynaecological cancer patients were invited to take part. The EORTC QLQ-C30 was used in both studies, and all patients could therefore be used in the analysis of this questionnaire (paper III). In contrast, the HAD Scale and the DBCG 89 Questionnaire were used in breast cancer patients only (Appendix B).

Between 1 and 24 hours after having completed the questionnaire at home (and having put the questionnaire in a sealed envelope) the participants were interviewed by a nurse via telephone. The interviewer asked the same questions as in the questionnaire but the patients were asked to respond using their own words and to avoid using the response categories used in the questionnaire. The interviews were tape-recorded and were subsequently rated by an observer (M. Groenvold for the breast cancer patients, M. Klee for the gynaecological cancer patients). The observer made qualitative comments during the rating. The questionnaires completed by the patients before the interview were compared to the observer rating based on the interview. For each item, the overall agreement (i.e., the proportion of cases where the patient and the observer had given identical responses) and the (weighted) kappa were estimated. A priori it was decided that kappa values equal to or below 0.40 indicated potential validity problems, values up to 0.60 also deserved attention, whereas values of 0.61-1.00 indicated acceptable results [44]. A detailed description of the methodology is provided in paper III.

### 3.9 Staff survey

#### 3.9.1 Identification of study sample

Almost all patients in DBCG 89 protocols B, C, and D were treated at one of the five comprehensive cancer centres or at one of four regional oncological departments. We contacted 46 health care professionals working at these nine centres/departments, 19 physicians and 27 nurses (paper IV). These included the consultant and head nurse in charge of breast cancer treatment, who were asked to identify their most experienced colleagues.

#### 3.9.2 Questionnaire

A staff questionnaire was constructed by selecting 18 HRQL dimensions from the patient questionnaire. We selected the dimensions we thought were most likely to be affected by adjuvant therapy and were most important, as based on the pilot study and literature review. The staff questionnaire consisted of six almost identical parts. Each concerned a comparison of two groups selected from the DBCG 89 protocols. For each of the 18 HRQL dimensions it was asked: 'Which group – all things being equal – has the problem/symptom to the largest extent?'

#### 3.9.3 Questionnaire administration

Each health care professional received a package consisting of a staff questionnaire, a patient questionnaire, an information letter, and a stamped return envelope.

### 3.10 Comparison of participants and non-participants

In the clinical studies (papers VI-VIII) the characteristics of the final groups of participants in the study were compared against larger subsets of the target populations to determine whether the patients actually included were similar to the target groups (details in each paper). Age and tumour size were compared using Wilcoxon's rank sum test; proportions defined by other clinical variables were compared using Fisher's exact test or  $\chi^2$  test. The same was done to compare participants and non-participants in the validation study (paper III) and to compare the groups within papers VII and VIII. In paper VI the demographic characteristics of breast cancer patients and the general population sample were compared using ordinal logistic regression controlling for age.

### 3.11 Analysis of HRQL data

#### 3.11.1 Scoring of questionnaires

The EORTC QLQ-C30 was scored according to the Scoring Manual [45]. A high score on one of the five functional scales or on the global health status/quality of life scale indicates a good function, whereas a high score on one of the three symptom scales or the six single items indicates a high level of symptoms/problems. Two of the early papers used simpler methods. In paper I the scores were dichotomised. In paper II a linear transformation of EORTC QLQ-C30 scores was used but the scores were not transformed to 0-100 (footnote to Table 2 in paper II).

The HADS was scored according to guidelines: scores for each of the two sub-scales were constructed by summation of its seven items [46] when at least 6 of the 7 items were not missing. The DBCG 89 Questionnaire was analyzed as single items. Items using the same four response as in the EORTC QLQ-C30 were transformed to 0-100 scales as for EORTC items [45], except in paper I where the scores were dichotomised.

#### 3.11.2 Group comparisons

In paper I the proportions experiencing symptoms in the two 'known' groups were compared using Fisher's exact test. In papers II, III, V, VII, and VIII scores were compared between groups using Mann-Whitney (Wilcoxon) rank sum test (two-tailed) [47]. In paper VI the HADS scores were compared using age as covariate in an analysis of covariance (ANCOVA) model. The same comparisons were carried using the non-parametric partial gamma [38, 48] with age grouped in 10-year intervals, and this method was also used to compare the proportions of HADS cases. The level of significance was 0.05 in all the analyses listed, except in paper VII where it was 0.01 and where at least two significant findings in the treatment period were required to confirm a hypothesis. The SAS statistical analysis program (SAS Institute Inc., Cary, NC, USA [49]; versions 6 to 9.1) was used for all analyses unless otherwise specified.

### 3.12 Prognostic factor analysis

In addition to a range of clinical and biological variables (paper IX), six 'HRQL' variables were selected for analysis. The EORTC QLQ-C30 emotional function scale and global quality of life item and the anxiety and depression subscales of the HADS were selected as indicators of psychological distress. The EORTC QLQ-C30 physical function and fatigue scales and the global health item were selected as indicators of physical health. Social class was included to control for possible confounding (social class may be related to HRQL as well as to prognosis).

Patients were followed until 1 March 2005 resulting in a median follow-up time of 12.9 years.

The multivariate Cox proportional hazards regression analysis was used to predict recurrence-free survival (RFS) and overall survival (OS). The categorisation of clinical and biological variables was described in the article (paper IX). To avoid over-estimation of effect resulting from categorisations derived from exploratory analyses of the data, all patient-rated variables were dichotomised at the median. In addition, to take the clinical definitions of 'case' vs. 'non-case' into account, the HADS subscales were analysed using the recommended cut-points 7/8 and 10/11 [46].

The analysis took place in three steps. First, multivariate 'biological models' for RFS and OS were made based on the clinical and pathological variables. Second, each of the patient-rated variables and social class were added to the biological models, and the risk

ratios for that variable in combination with all variables in the biological models were estimated. Third, all the 'self-rated' variables and social class were added to the biological model and a stepwise selection ( $p < 0.05$ ) was carried out, keeping all biological variables.

In addition, we carried out the final multivariate analysis resulting from the procedure described above in low-risk patients (Protocol A), only ( $N=432$ ). These patients had not received any systemic adjuvant therapy but some had radiotherapy; this variable was included in the model.

When the proportional hazards assumption was not fully satisfied we compared the results using the variable against an analysis stratified by that variable. Two-sided  $p$ -values based on the Wald test statistic were estimated. The SAS software package version 9.1 was used.

## 4 RESULTS & DISCUSSION

### 4.1 Questionnaire development, composition, and pilot testing (paper I)

Based on the literature review a list of issues was made. Because no articles dealing with ovarian ablation or endocrine therapies were identified in the literature review a gynaecologist was consulted and asked whether any likely effects of these treatments were missing in the list. Two issues were added in order to assess consequences of low levels of oestrogen; 'vaginal dryness' and 'urinary incontinence'. Based on recommendations in the literature [50], the issues about sexuality were supplemented with 'sexual satisfaction'. A number of issues were removed from the list because they were not considered severe or frequent or were difficult to operationalize.

As a result of the review of available instruments, two questionnaires, which were widely used internationally, were selected for this study. The EORTC QLQ-C30 [35, 45, 51], a 30-item questionnaire developed by the European Organisation for Research and Treatment of Cancer Quality of Life Group [52] was selected because it covered many of the issues identified in the literature review, because it was considered to be well-structured (consisting of brief multi-item scales as well as single items), because of its format with simple questions and response options, and because it was developed in a cross-cultural, mainly European context.

To assess anxiety and depression, the two psychological constructs reported most frequently in the literature, the Hospital Anxiety and Depression Scale (HAD Scale) [46] was selected. This questionnaire was widely used [22, 53-56] and was recommended for cancer studies [57, 58]. It consists of 14 items constituting two seven-item scales for anxiety and depression, respectively.

To assess social network/contact, four items from the Danish Glostrup Population Studies were selected [59].

In addition, 19 items, including one open-ended item for supplementary comments, were developed.

The 14 interviews with patients generally confirmed the decisions made during the choice of issues for the questionnaire, and the pilot testing with 58 patients confirmed that the questionnaire was acceptable. However, a few revisions of questionnaire developed for the study were made. Three items about vaginal discharge, weight gain, and wearing wig and two 'administrative' items about dates for treatment and questionnaire completion were added. Two items on consequences of surgery were removed because they were considered out of focus, and one item

on cohabitation was moved to the questionnaire on demographics. The wordings of a few of the newly developed items were modified.

Thus, the 69-item questionnaire used in this study consisted of the EORTC QLQ-C30 (30 items), 21 items developed for the study, four items on social network/contact, and the HAD Scale (14 items).

In the analysis for papers VII and VIII not all of the items were reported. The four items on social network/contact had been included in order to be used as covariates in analyses, not as outcome variables. The item on sexual satisfaction was excluded due to ambiguous interpretation. Finally, the two 'administrative' items and the item for comments were not used as outcome variables. In papers VII and VIII the 17 remaining items developed for this study have been named the DBCG 89 Questionnaire (the English translation is shown in the Appendix of paper VIII).

In the general population study we used the same questionnaire except that eight obviously cancer-related items (e.g., the items on hair loss) were removed (paper V).

### 4.2 The questionnaire(s) used in this study compared to other questionnaires

How does the content of the questionnaire combination used in this study compare with questionnaires used in other studies? Table 4 (section 4.9.6) shows the content of the three questionnaires used in this study. Table 4 also includes findings about chemotherapy in this and other studies, organised according to the structure of our questionnaire as further discussed in section 4.9.6. Obviously, other studies have used other questionnaire combinations. The results, which could not be organised according to the content of our questionnaire(s), are summarised in Table 5.

Thus, taken together, Tables 4 and 5 illustrate the extent of sufficiency of the questionnaires used in the literature as measured according to their ability to reflect the HRQL of breast cancer patients in adjuvant chemotherapy. The two tables show that the questionnaire combination used in the present study is the most complete. This is further discussed in section 4.9.6.

The two standard questionnaires, the EORTC QLQ-C30 and the HAD Scale, have become widely used. The EORTC QLQ-C30 has been used in thousands of studies and is the most frequently used instrument in European and Canadian HRQL studies in oncology [60]. The HADS has also been used extensively [61, 62]. Thus, our choice of these two instruments turned out to be congruent with decisions made in many subsequent studies. Therefore, a considerable part of our results have become comparable with a large part of the literature. The DBCG 89 results are not comparable to other studies, but our study (paper VII) showed that a questionnaire with at least part of that content is necessary. Questionnaires are also discussed in section 4.9.6.

### 4.3 Study participation

#### 4.3.1 Inclusion of patients

The inclusion periods for the protocols are listed in Table 2. The table shows that the planned number of patients (100-150 fully evaluable participants per protocol arm) was reached quickly in protocols A and C, whereas protocols D and particularly B had slower accrual. The age limits for protocols A and C were increased shortly after the study was initiated and therefore inclusion of patients in the oldest old groups started a few months later.

During the inclusion periods a total of 1,950 patients were registered and contacted about the quality of life study. Of these, DBCG later determined that 50 did not fulfil all inclusion criteria; two were contacted outside the inclusion periods. These patients were excluded leaving 1,898 eligible patients (Table 2).

**Table 2**

**Inclusion periods for each of the DBCG 89 protocols and the number of patients included. For protocol A the dates refer to dates of operation, for the other protocols the dates are randomisation dates.**

Protocol	Inclusion started	Inclusion stopped	Patients included
A1	1 June 1991	31 October 1992	
A2 (< 70 years)	1 June 1991	31 March 1992	
A2 (70-74 years)	1 December 1991	30 September 1992	538
B	1 June 1991	6 February 1996	317
C (< 70 years)	1 June 1991	31 October 1992	
C (70-74 years)	1 October 1991	31 December 1992	469
D	1 June 1991	8 June 1995	574
<b>Total</b>			<b>1,898</b>

#### 4.3.2 Patient participation

Of the 1,898 patients, 1,713 (90.3%) filled in the first of the six questionnaires (Table 3). The table shows that the number of patients participating declined modestly over time; the figures are not adjusted for the fact that some of the missing patients at the later assessments had died.

**Table 3**

**Number of patients completing each of the six questionnaires of the study.**

	Month					
	1	2	3	4	5	6
Participants	1,713	1,644	1,599	1,561	1,502	1,404
Per cent	90.3%	86.6%	84.2%	82.2%	79.1%	74.0%

#### 4.4 Testing multi-item scales for differential item functioning

##### 4.4.1 Results concerning all treatment arms (paper II)

The analyses of the nine multi-item scales of the EORTC QLQ-C30 using data from the first 1,189 patients completing the questionnaire showed DIF in three scales (paper II). DIF in the physical function scale was found in relation to both age and treatment (+/- chemotherapy) for the item 'Do you have to stay in a bed or a chair for most of the day?' After control for the physical function scale score, higher proportions of younger patients and patients in chemotherapy answered 'Yes' to this item. There was a relatively strong association between scores on this item and the 'vomiting' item. Our interpretation was that in addition to being a

measure of physical function this item is a measure of nausea/vomiting: some patients who had severe nausea after chemotherapy had to stay in bed (paper II).

The results of paper II led to two conclusions. First, it confirmed the potential relevance of DIF analyses. Second, at the methodological level, we found support for a relatively strict significance criterion of  $p < 0.001$ , which reduces the number of 'DIF cases' identified. Therefore, this approach was used in the analyses related to papers VI-VIII.

##### 4.4.2 Results concerning anxiety and depression in breast cancer patients compared to the general population (Appendix A)

The main finding was that there was DIF in relation to group in the depression scale mainly for item 10: 'I have lost interest in my appearance'; at a given level of depression there were much higher scores on this item in the general population than in the breast cancer patients (Appendix A). This item therefore influenced comparisons between the two groups: depression scores in the general population sample were biased upward. If this item was removed from the HAD depression scale, the magnitude of the difference in mean scores between the two groups (described in section 4.8.2) was diminished, but still significant.

An analysis of the wordings of the items of the depression scale exhibiting DIF suggested an explanation. The item, 'I have lost interest in my appearance', has response categories including 'I take just as much care as ever' (corresponding to '0' depression). The analysis indicated that – at a given level of depression – women from the population sample had a much greater likelihood of being 'depressed' on this item. The context of response may explain this bias. A cancer patient may perceive this item as relating to her disease ('I have lost interest in my appearance' - since I became ill). On the other hand, a woman from the general population sample had no clear point in time or event with which to compare her 'interest in ... appearance' and may have responded by comparing her actual 'interest' to some prior level, e.g. when she was younger. If this difference in perception actually took place, it might explain why – when controlling for depression score – the score on this item was higher in the general population sample.

The HADS was developed for and validated in hospital patients [46]. We must therefore assume that the cancer patients' perception of this item better reflects the intended meaning. If the interpretation of the DIF is correct, this item overestimates depression in population samples. As stated above, correcting for this bias by excluding the item from the scale score diminished the difference between groups but did not change the conclusion of significantly higher depression scores in the general population sample than in the patient sample.

However, similar DIF may affect scores on four other depression (but no anxiety) items, which have such diffuse references to prior states. Unfortunately, as the test for DIF examines one item at a time in relation to the other items of a scale, it cannot detect effects that affect the majority of items in a scale. The analyses thus suggested, but could not demonstrate, that the problem of 'diffuse back reference' affects additional items in the depression sub-scale. It was not possible to correct fully for this bias.

The depression sub-scale also showed DIF with regard to age for 'I feel as if I am slowed down' (item 8) and 'I look forward with enjoyment to things' (item 12) (Appendix A). This DIF does not directly affect the outcome of the comparison of the two samples as the analysis was age-stratified, but it explains at least part of the apparent increase in depression score with age. It is interest-

ing that the validation study found independent evidence for the age-bias in item 8 and thus supports the interpretation above (section 4.5.2).

In conclusion, the DIF analyses suggest that the HAD depression sub-scale over-estimates depression in a general population sample and – of less importance here – in older women compared to younger women.

The results of the DIF analyses had to be omitted to shorten paper VI but the interpretations of the DIF results were used in the paper where they appear as ‘speculations’ rather than results. An Australian study carried out DIF analyses in a similar project comparing breast cancer patients to a general population sample using the HAD Scale [63]. That study found the same overall results (lower anxiety and depression in breast cancer patients than in the general population sample). They explicitly tested our interpretation of the findings, particularly the hypothesis that ‘diffuse back references’ as discussed above for item 10 and other items might contribute to the counter-intuitive findings discussed in section 4.5.3. Using different methods to test for DIF the Australian study found the same DIF for item 10 as described above. For the depression scale all DIF findings were compatible with the hypothesis that items having ‘diffuse back references’ tend to underestimate depression in breast cancer patients relative to women from the general population. However, in contrast to our study DIF was found for anxiety items as well, and the overall conclusion by Osborne et al. was that the finding of lower levels of anxiety and depression in breast cancer patients could not be explained by DIF [63]. This may be correct but as discussed above the main problem may be that the DIF analysis method is not well suited for bias affecting several items in a scale.

#### 4.4.3 Results concerning chemotherapy versus no chemotherapy and chemotherapy versus ovarian ablation (Appendix A)

The DIF found in paper II was not found in the data reported in papers VII and VIII even though the data were overlapping. However, DIF of similar magnitude and direction, but not meeting the  $p < 0.001$  significance criterion, was found in the data reported in both papers. The two studies found that patients in chemotherapy had poorer cognitive function than patients not in chemotherapy (paper VII, Table 2 and Fig. 1) and than patients who had undergone ovarian ablation (paper VIII, Fig. 2). The DIF implies that in these cases the cognitive function scale score was inappropriate as a description of the effect of chemotherapy: chemotherapy had a relatively strong effect on concentration and a much weaker effect on memory.

#### 4.4.4 Discussion

Our article (paper II) was the first publication presenting and applying DIF analyses to HRQL research. Is this new ‘technology’ relevant and worthwhile?

The results discussed above suggest that it may indeed be worthwhile to carry out DIF analyses: the results of DIF analyses contributed to the detection of possible problems associated with the use of the HAD Scale. If undetected, these problems could have led to misinterpretations.

In the subsequent analyses we used the relatively strict criterion of  $p < 0.001$  (Bonferroni-correction) based on paper II and did not find any DIF. With the relatively small samples in particularly paper VIII it is clear that the power to detect DIF was limited. The replication of the findings of DIF in the cognitive function scale suggests that this DIF was indeed missed in both studies due to the strict significance criterion (i.e., a type II error). Therefore,

one could consider being less restrictive in smaller studies but of course this will increase the number of ‘false positive’ findings (type I errors). We reviewed the remaining DIF results related to the studies reported in papers VII and VIII, which did not meet our criterion for significance. There were 15-20 findings and most were difficult to interpret and were probably random fluctuations, or were at least without clinical importance.

In later research we continued using the Bonferroni adjusted significance level and based on Bjorner’s study [64] we added a further requirement to reduce the number of findings: that the partial gamma coefficient had to be numerically larger than 0.30 (the criterion was later adapted to logistic regression analysis) [65, 66].

As previously discussed, an Australian study found DIF in the HAD depression scale supporting our interpretation (i.e., bias due to ‘diffuse back references’) but concluded that DIF could not explain the unexpected finding of lower levels of anxiety and depression in breast cancer patients compared to the general population. This highlights an important limitation of the DIF method: that it cannot detect or correct for effects affecting several items in a scale in the same way. Theoretically, this limitation can be avoided if the total scale score, which is used to control for the level of the attribute, is replaced with an unbiased variable measuring the same attribute [67]. However, in practice this is often not possible because such a variable is not available (questionnaires are kept as brief as possible and measuring the same dimensions twice or more is not feasible). Instead, it is sometimes recommended to substitute the total score with a score based on item response theory (IRT) scoring of the same scale [68-70]. This may, however, not be relevant for the EORTC QLQ-C30, as our later research showed that there was little difference between traditional and IRT scoring of these scales [71]. In our recent study comparing translations of the EORTC QLQ-C30, DIF analyses using IRT scoring were compared to those based on traditional scoring: DIF results were similar [72].

DIF analyses were uncommon in the HRQL field (whereas they have been widely used in educational testing for many years [73]) until quite recently. Now they have become relatively widely used: according to a PubMed search (January, 2007) 108 out of approximately 180 articles mentioning DIF were published in 2004-2006.

This increasing use makes it even more important to critically discuss the usefulness of DIF analyses. Our results in this thesis and in recent papers showing that DIF analyses are effective in detecting problems in the translation of multi-item scales as well as in examining for cultural differences [66, 74] illustrate the relevance of the method. Many other examples are found in recent applications in HRQL research.

However, despite the potential relevance of the method it should also be evaluated in the light of its ‘side effects’. DIF analyses cost time, complicate the reporting of results, and can be seen as a ‘problem generation mechanism’, which delays the research process. Do the benefits outweigh the costs? Are DIF analyses necessary in routine analyses of HRQL data?

We recently reviewed the use of DIF analyses, particularly in the analysis of clinical trials, in a book chapter. Given that it is not always feasible to carry out DIF analyses we made the following recommendations about when DIF analysis is of particular importance [67]:

- ‘In analyses of great clinical importance (to make sure that the conclusions as to content are correct)

- If the unidimensionality of the scale is questionable, e.g. scales covering broad domains of HRQL
- If DIF has previously been detected in the scale in relation to the variable
- In analyses focusing on variables, which may be associated with DIF, e.g. ethnicity or sociodemographic variables such as sex and age
- In studies using questionnaire versions in more than one language.'

#### 4.4.5 Conclusions

- Several cases of DIF were found in the multi-item scales of the EORTC QLQ-C30 and HADS.
- DIF analyses are relevant when important analyses of multi-item scales are made.
- The usefulness and interpretability of DIF analyses may be increased through parallel investigations using qualitative or cognitive interviewing techniques, as seen by the congruence of findings in this and the next chapter (section 4.5).

### 4.5 Testing whether patients and researchers understand questionnaire items in the same way

#### 4.5.1 EORTC QLQ-C30 (paper III)

Of the 57 breast cancer patients, 46 (81%) were successfully interviewed. Of 88 gynaecological cancer patients, 49 (56%) were successfully interviewed. The overall participation was 66% (paper III).

In general, the agreement between patients' responses to the EORTC QLQ-C30 questionnaire before the interview and the observer's rating based on the interview was remarkably high. The mean scores did not differ between patients and observers. The median overall agreement for the 30 items was 0.85 (range 0.47-1.00). It was above 0.80 for 21 of the 30 items. The median kappa/weighted kappa was also 0.85 (range 0.49-1.00). No items had kappa values at or below the a priori chosen threshold 0.40, and only three items were in the range 0.41-0.60 (paper III). While the overall result was that the agreement was very high, the analysis of both the quantitative and qualitative data suggested some potential problems. The most important of these problems was what we labelled 'selective reporting': patient replies such as 'Yes, I had the symptom but it was due to something else' were observed for the item about pain and other symptoms. We had not instructed the patients to report only symptoms caused by the breast cancer or its treatment. Despite this, several patients clearly wanted to prevent symptoms caused by other factors, e.g. arthritis, being misclassified by the researchers as results of breast cancer. Therefore, they reported selectively, carefully distinguishing between causes of symptoms.

Selective reporting may invalidate comparisons of groups of persons who differ as to their perception of the research aims. This was particularly relevant to the present study where great effort had been put into establishing an optimal control group based on a general population sample. For example, we must expect that equal proportions of breast cancer patients and women from the general population suffer from pain due to arthritis. However, if the breast cancer patients reported only parts of this pain, whereas participants from the general population sample reported their arthritis symptoms more completely, then the comparison of pain is biased.

In addition, a specific problem in a single item was detected. Some patients seemed to overlook the word 'difficulties' in the

item 'Has your physical condition or medical treatment caused you financial difficulties?' Instead they answered whether the disease or treatment had had any economic consequences. This item may therefore lead to over-reporting.

#### 4.5.2 HAD Scale (Appendix B)

The quantitative results from the breast cancer patients are shown in Table B1 (Appendix B). The median overall agreement was 0.80 (range 0.61-0.91). The median weighted kappa was 0.79 (range 0.65-0.95).

Thus, the median values were slightly lower than for the EORTC QLQ-C30, but in contrast to the results for the EORTC QLQ-C30, all items were above the a priori defined threshold for acceptable results (0.61). The plan (paper III) was to use the qualitative data in the cases where the agreement indicated potential problems ((kappa up to 0.40) or was in the intermediate range (0.41-0.60). The agreement was higher than these limits but the qualitative data will still be briefly examined.

The comments are listed in Table B2 (Appendix B), and showed the following:

The comments to items 1 and 12 can be viewed as corrections to the negative way items in symptom questionnaires are generally phrased. These patients report positive effects (changes) due to breast cancer. This 'positive adaptation/adjustment' [75-78] is discussed in paper VI.

The answers to item 3 showed that it was understood by some patients as concerning fear of breast cancer recurrence and death. The same point was made for item 5. This finding supports the hypothesis of selective reporting discussed above (section 4.5.1 and paper III): some breast cancer patients specifically understood these items as dealing with breast cancer-related worries; therefore, they may have left out some problems that were unrelated cancer.

Patients who understand the item in this way may feel that the response categories are inappropriate. A patient experiencing a moderate extent of fear of recurrence is likely to feel that 'Very definitely and quite badly' is not suitable, but the two following options 'Yes, but not too badly' and 'A little, but it doesn't worry me' may also be unsuitable: there seems to be a contradiction between the seriousness of the question and the response categories.

The two comments to item 8 suggest that this item about being 'slowed down' may capture an effect of ageing in addition to being a measure of depression. This finding was also made based on the DIF analyses reported in section 4.4.2, and the qualitative results thus corroborate the statistical findings. The effect of this possible age bias is that depression in older persons is over-estimated compared to younger individuals having the same level of depression.

Other comments address difficulties in discriminating between adjacent response categories (item 5) and minor technical or specific issues.

An overall view of the qualitative comments does not suggest major validity problems affecting the sub-scale scores to a large extent. Despite this, the following results are important:

The finding that some patients lacked positive response options emphasises that it is too simple to expect that disease and treatment have negative consequences only.

Selective reporting was found, confirming results from the EORTC QLQ-C30.

Possible age-bias was found in some depression items confirming DIF analysis results.

Specific problems were found in some response categories. Most of these problems (except selective reporting) could probably be handled through relatively small modifications of particularly the response categories. However, one of the virtues of standard questionnaires is that they are seldom changed (thus making results comparable across studies) so it is doubtful whether this will be done.

The interpretation of the results for the HAD Scale is somewhat different compared to the EORTC QLQ-C30 and the DBCG 89 Questionnaire items in the sense that our estimates of agreement on individual items may be less important: the HAD Scale is linked to the relatively well-established concepts, anxiety and depression. It was developed as a screening questionnaire and can be deemed valid if it accurately predicts these diagnoses. Numerous validation studies have, in general, shown that the HAD Scale is in agreement with professionals' diagnoses (summarised by Bjelland [62]). However, the very high level of agreement observed in this validation study is encouraging as it confirms that items are generally understood as intended. In addition, relevant findings that can be used to nuance the interpretation of HAD Scale results were made from the qualitative data. This study thus adds important information to the literature about the validity of the HAD Scale.

#### 4.5.3 The DBCG 89 Questionnaire (Appendix B)

The results concerning agreement are shown in Table B3 (Appendix B). The median overall agreement was 0.91 (range 0.48-1.00). The median weighted kappa was 0.92 (range 0.51-1.00). The observer rated mouth soreness significantly higher and rated desire for intercourse significantly lower but both differences were relatively small. Thus, agreement was generally excellent. The one exception exhibiting 'moderate agreement' [44] only was the item 'satisfaction with appearance', which used a seven-point scale, and had overall agreement 0.48 and weighted kappa 0.51. There was no difference in mean scores for this item. The qualitative data are shown in Table B4 (Appendix B). The two comments related to the 'satisfaction with appearance' item both concerned the ambiguity these patients had felt: they were strongly affected by their recent breast operation but did not know whether to take this into consideration because it was not visible to others. More generally, these observations indicate that the item is relatively broad and unspecific, and this – together with the many response options – results in less agreement. In fact the overall agreement for this item was almost identical to that observed for the two 'overall items' 29 and 30 in the EORTC QLQ-C30 (Table 2, paper III), which use seven similar response categories, whereas the kappa was somewhat lower. It is usually advisable to be specific in item formulation (when possible) in order to limit ambiguity but on the other hand, as for this and the two similar EORTC items, one may want an overall evaluation. In summary, the results indicate that the responses to this item may vary more strongly with the individual patient's perception than do more specific items. However, there is no evidence that the item was misunderstood, and scores for patients and observers were similar.

Even though the agreement for the remaining DBCG 89 items was excellent, some potentially important issues appeared in the qualitative data (Table B4, Appendix B). Concerning the item 'interest in sexual intercourse' the standard response categories (taken from the EORTC QLQ-C30) appear suboptimal. The usual EORTC QLQ-C30 interpretation of the response categories would correspond to a range from complete loss of libido ('not at all') to

no loss of libido ('very much'). However, the results showed that some patients perceived 'a little'/'quite a bit' as expressions of 'normal' sexual desire. This observation may also explain that observers rated this item slightly lower than patients. The item appears valid in the sense that patients understood the content correctly, but when results of the item are interpreted it should be kept in mind that 'normality' is not equal to 'very much'. The item might probably be improved through response categories better matching the question.

Concerning the item on work ('... worked outside your home...'), two weaknesses were detected. First, respondents having paid work at home are not eligible to report their work with this item. Second, child minders in Denmark usually have working hours above the usual level, and an answer result such as '40 hours' (about 20% reduced time) may therefore be misinterpreted as full time work. However, the significance of these weaknesses appears to be minor.

#### 4.5.4 Discussion

As outlined above, the present validation study was designed mainly to validate the EORTC QLQ-C30 and the DBCG 89 Questionnaire items, not the HAD Scale, which could be (and had been) validated against clinical diagnoses. The validation study was designed to elucidate whether the patients' open-ended responses were in agreement with the wordings of items as understood by the researcher.

The main finding of this validation study was that the agreement was very good, and considerably higher than expected, thus indicating that despite the scepticism one may (and should) have towards questionnaires, breast cancer patients' completion of these questionnaires seems to be valid to a remarkably high extent.

The levels of agreement were much higher than usually observed in inter-rater studies where patients' responses have been compared to health care professionals' or significant others' evaluations [79, 80]. Although not fully comparable due to different statistics, it is also remarkable that the values observed here for the EORTC QLQ-C30 do not appear to be lower than test-retest correlations ranging 0.70-0.90 in a Norwegian study [81]. In other words, there does not seem to be more discrepancy between patients and observers than between patients' own assessments separated by a few days.

The most important problem found was that of selective reporting, which is probably a phenomenon affecting all or at least most HRQL questionnaires that do not specifically ask the patients to indicate whether the symptoms are caused by the disease or treatment. Additional analyses of other data based on the same method have confirmed that selective reporting is found in other questionnaires as well [82]. There can be little doubt that the phenomenon is real and that it may affect comparisons of groups who have different perceptions of the study.

The magnitude of effect of selective reporting could be elucidated experimentally if patients were first given the usual instructions, and then, after completion of the questionnaire, were informed that it was important that they reported all symptoms including those caused by factors other than cancer. They could then complete the questionnaire again after the instructions, and the results could be compared. Alternatively, cancer patients who had completed a questionnaire could be interviewed about what they had reported.

As described later, paper VI showed unexpected results indicating that breast cancer patients had less psychological morbidity than

the women from the general population sample. It seems likely that selective reporting was part of the explanation of this probably incorrect result. As described above, some breast cancer patients perceived questions in the HAD Scale as relating to whether they were worried or sad due to the cancer. They may thus have excluded 'everyday worries or sadness' from their replies. However, other factors may also have contributed (section 4.8.3).

Concerning selective reporting it is important to note that in many studies the groups compared will not have different perceptions of the study, and therefore selective reporting does not influence the outcome of a comparison. Selective reporting is described and discussed in the recent version of Fayers & Machin's textbook [25] (p. 438-441), where it is concluded that it may be a general mechanism that may lead to under-estimation of symptoms/problems.

The finding of a phenomenon such as selective reporting, which is interesting and important but is probably still in most cases a limited source of bias, should not distract attention from the answer to the overall research question posed in this validation study: do patients understand the items in the way they are intended? The remarkable result of this study was the generally very high level of agreement found for the two standard questionnaires as well as for the newly developed DBCG 89 Questionnaire.

The present validation method was also used to validate newly developed items in the questionnaire used in the parallel study of gynaecological patients [82, 83]. The level of agreement found for the two parts of the questionnaire was similar to those found here, median weighted kappas were 0.88 (range 0.71-1.00) [82] and 0.80 (range 0.52-1.00) [83]. Selective reporting was observed in the questionnaire measuring urological and gynaecological morbidity [82].

During the recent years a new scientific field, cognitive aspects of survey methodology (CASM), which is highly relevant to HRQL questionnaire validation, has evolved [34, 84, 85] (see also section 7.1.2). In a review of the use of cognitive techniques Willis et al. described the present study as '...an interesting example of cognitive techniques... Such an approach points the way for other cancer-related HRQOL researchers to evaluate their instruments.' [34] (p. 616).

#### 4.5.5 Conclusions

- This new method for validation of questionnaires showed high levels of agreement between patients' written answers and observers who had listened to an open-ended interview. This indicates that, in general, patients respond to the questions as intended. This finding supports the validity of the EORTC QLQ-C30, the HAD Scale, and the DBCG 89 Questionnaire.
- A mechanism we termed selective reporting was identified via the qualitative comments. Selective reporting may lead to under-estimation of the levels of symptoms and problems and may lead to bias if groups of respondents who are in different circumstances and therefore perceive the questions differently are compared. Selective reporting is likely to be a general, methodological problem, which is not specific for the questionnaires investigated here.
- The identification of the selective reporting mechanism probably contributed to avoiding misinterpretations of results in other parts of this study (particularly in the compar-

ison of breast cancer patients to women from the general population).

- The method also proved useful in providing other insights into the complex process of questionnaire completion, and appears to be a widely applicable questionnaire validation technique that can be used in and adapted to all kinds of questionnaire.
- In the recent years extensive work based on cognitive psychology has taken place mainly in the USA, and this method can now be seen as one of many 'cognitive interviewing techniques'.

### 4.6 The staff survey (paper IV)

#### 4.6.1 Participation

Of the 46 health care professionals contacted, 36 (78%) responded.

#### 4.6.2 Staff survey results for the comparison chemotherapy versus ovarian ablation (Appendix C)

The staff survey data were removed from the final version of paper VIII to reduce the length of the paper and because it was difficult to integrate it into the text without complicating the structure.

Table C1 shows the staff survey data for the comparison of patients on chemotherapy versus ovarian ablation. A priori we decided that if at least half the staff members expected that a quality of life issue would be more affected by one of the treatments, this was a 'staff hypothesis'. The majority of professionals expected more nausea and vomiting, hair loss and fatigue in patients receiving chemotherapy (Table C1, top). As shown in paper VIII and in section 4.10.2 of this thesis, this was found in the patients' responses.

A majority of staff respondents expected more hot flushes and more irregularity of bleedings/menostasia in patients having undergone ovarian ablation (Table C1, bottom). Again, this was seen in the quality of life data (paper VIII), section 4.10.2. Thus, for these five quality of life issues, we can regard the staff expectations as independent hypotheses confirmed in the data. For the remaining issues, the majority of staff members either did not expect a difference or the picture was mixed. This means that the data could not be used as a basis for formulating and testing hypotheses.

If we look at the patient data and compare these to the staff's expectations (i.e., taking the same point of view as in the next section, 4.6.3), there are surprising discrepancies: significant minorities of physicians and nurses did not expect that nausea and vomiting, hair loss, and fatigue were more prevalent in patients receiving chemotherapy.

#### 4.6.3 Staff survey results for the comparison chemotherapy versus control

The staff survey was not carried out with the comparison of control patients and patients on chemotherapy in mind (paper IV, page 484). When initiating the HRQL study, a relatively large amount of literature was available and the literature review (in paper I) could in fact be seen as a compilation of the hypotheses available in the research field at that time. Therefore, when reporting the results of the comparison of control patients and patients on chemotherapy in paper VII we did not use the data from the staff survey to generate hypotheses. Instead, the original literature review (in paper I) was used to generate such hy-

potheses. Thus, the staff survey was used in a way that was not anticipated when designing the study: the results were seen as a description of the level of knowledge about HRQL consequences of chemotherapy in the staff treating the patients.

As elaborated in section 4.9.8 and paper VII, the staff study suggested that the information given to patients about HRQL consequences of chemotherapy was insufficient.

#### 4.6.4 Discussion

The staff survey was used as intended in relation to the DBCG 89 trial comparing chemotherapy and ovarian ablation (paper VIII) and as a way of elucidating staff knowledge about patient-experienced consequences of chemotherapy (paper VII). Concerning the comparison of chemotherapy and ovarian ablation (paper VIII) five hypotheses could be extracted from the staff survey data and these were confirmed in the patient data. Thus, it proved possible to use the staff survey data as intended. However, these data were not included in the published paper (section 4.6.2). If they had been incorporated in the paper, would this manoeuvre have led to the anticipated increase in the scientific credibility of the results? The five hypotheses concern findings that are not surprising. Three of them are extremely well known consequences of chemotherapy: nausea and vomiting, hair loss and fatigue. Our reporting that these symptoms/problems were more frequent in patients on chemotherapy has high credibility even without the 'support from the staff survey'. The same is the cases for hot flushes and irregularity of bleedings/menostasia. In other words, the staff survey allowed us to generate hypotheses only in areas where we did not need the support of such hypotheses.

One could have hoped that the staff had picked up previously unknown consequences of one of the treatments; such findings could have been used to form hypotheses, and the patient data could have confirmed the hypotheses, but this was not the case. In fact, the finding that even HRQL consequences of chemotherapy, which should be widely known by the staff, were not expected, suggests that the likelihood that staff surveys can be used to pick up such problems is limited.

Later research seems to support this interpretation: even in a palliative care setting where the attention towards patients' symptomatology should be very high, we showed that many symptoms and problems were not detected by doctors [86] or nurses [87]. This is in line with other studies [88]. And related to this, we found little agreement between patients and doctors on the assessment of the patients' HRQL [89].

In this thesis, the staff survey data have been used as intended in relation to one trial only (paper VIII). However, preliminary analyses of the data concerning the other comparisons seem to give the same picture: only the most evident and pronounced HRQL consequences of treatments were expected by the majority of staff members. And as these effects are also convincingly demonstrated in the patient data, the staff surveys had little added value.

In contrast, the 'alternative' use of the staff survey data to elucidate the staff knowledge and thus (indirectly) their information to patients (paper VII) was useful as it suggested serious deficits and highlighted a need for additional research (and probably also for quality improvement) in this area.

More generally, the staff survey was an attempt to address both the problems resulting from the fact that a priori hypotheses were not formulated and the problems of multiple hypothesis

testing resulting from the large questionnaire and the repeated measurements. As such, the staff survey could be seen as an example of 'methods triangulation', i.e. the widely recommended parallel use of more than one scientific method. Ideally, with the addition of a simple, low-cost study, evidence with increased scientific robustness would have come out from the first study reporting newly detected HRQL consequences. However, that was not the case here.

An alternative solution to the problem of multiple hypothesis testing was that used in paper VII, i.e. the generation of hypotheses from a literature review. This was done after the original protocol was written, and thus it cannot be termed true a priori hypotheses, but because the basis was data collected prior to the study, it can still be defended. In paper VII the issue of multiple hypothesis testing was also handled through a decision rule (to confirm a hypothesis, a between-groups difference had to be found at least twice during the first three assessments) and through a reduction of the significance level to 1%. This approach seemed to work well. However, as no literature was available concerning chemotherapy vs. ovarian ablation, this method could not be used there.

There are at least four other solutions to the problem of multiple hypothesis testing, which were carefully considered but rejected. First, one could have limited the number of statistical tests by reducing the number of variables through the creation of overall endpoints. This would mean that instead of having more than 30 outcome variables, one would combine these to e.g. a single summary score, which could be called an 'HRQL score', or maybe two or three summary scores, e.g. for physical symptoms, psychological distress, and functional impact. It is not unusual to see such overall scores in the literature. For example, for the widely used 36-item SF-36 questionnaire [90] the two Physical and Mental Component Scores may be used instead of separate scores for physical function, bodily pain, vitality, etc. [91]. Similarly, a so-called Trial Outcome Index (TOI) composed of 23 items has been developed from the FACT-B breast cancer questionnaire [92]. However, such overall outcome variables were not suitable for the present study because they would not have produced a detailed picture of the HRQL in the different groups. Furthermore, as demonstrated in DIF analyses, there may be significant DIF (and thus loss or distortion of information) even in brief scales addressing relatively well-defined concepts. It is likely that the DIF problems would increase if so-called higher-order summary variables were created. This implies that important information about group differences might be lost.

Second, even if one refuses to reduce the number of outcome variables, the six measurement points during the two-year study period needed not to be used as separate outcomes. An alternative would be to base the comparison of groups on one overall test for each variable summarising the information from the six points in time. Such an overall test could be made in numerous ways ranging from simple tests (e.g. based on the mean value of six observations or the area under the curve) to sophisticated, statistical models, e.g. growth curve models [93], multilevel models, etc. [25, 94]. Such procedures could have added precision to the estimates and could thereby have increased the likelihood of detecting between-group differences. As the overall concern in the present study was to conserve the complexity of the original data, it was found valuable to determine the temporal patterns of the various symptoms and problems (e.g., 'how long is anticipatory nausea a problem?'). Therefore, in the longitudinal studies (papers VII and VIII) the decision was to carry out separate tests for the separate points in time.

Third, the problem of multiple hypothesis testing could be handled by adjusting the level of significance, e.g. via the Bonferoni method: the level of significance is adjusted by dividing the p value by the number of tests carried out [95] (see also paper IV). The problem resulting from such adjustment is that it reduces the power of the study to detect between-group differences [96, 97]. Therefore, in this study with its many variables and a relatively modest sample size, a full Bonferoni adjustment would mean that many of the findings of differences between groups would not be detected. Variations of the Bonferoni method, which take into account that the variables are not independent, reduce but do not eliminate this problem. In paper VIII the level of significance was kept at the traditional 5% but this was accompanied by 'warnings' in the text and with a graphical presentation of the level of significance, which made it possible to see from the figures how the results would have been if other levels of significance had been used. A partial adjustment of the level of significance was made in paper VII.

A fourth approach is to carry out an overall test for a number of variables followed by analyses of the individual variables only if the overall test is positive. A recent example of this was a comparison of two adjuvant chemotherapy regimens in early breast cancer using the EORTC QLQ-C30 and the breast cancer module QLQ-BR23 [98]. First, O'Brien's global rank procedure [99] was used. An overall difference was found for the EORTC QLQ-BR23, and subsequent analyses of the individual subscales showed differences in 'systemic side effects' and 'upset by hair loss' [98]. This technique reduces the risk of false positive findings but an important assumption is that similar effect is seen across the variables [100]. Differences seen in a few variables only may be missed. Thus, this approach has some of the same advantages and disadvantages as the use of overall endpoints. It is therefore not surprising that the study found few differences [98].

#### 4.6.5 Conclusions

- It is a generally acknowledged statistical principle that a priori hypotheses should be formulated whenever possible. This approach may also limit problems of multiple hypothesis testing. When a priori hypotheses are not available, there are different options that can be used. The formal use of a literature review to establish hypotheses is one example (paper VII).
- The idea of staff surveys to support the analyses of HRQL data is basically sound. However, the case was not proven here. The staff survey did not add significantly to the analysis of HRQL results. Staff surveys may be useful in other contexts, particularly when it is important to be able to make relatively definitive conclusions from a single study in a new area.
- More traditional ways of tackling the problem of multiple hypothesis testing (e.g., to collapse variables into overall summary scores, to combine information from longitudinal assessments into a single variable, or to adjust the p-value for the number of tests) are justified in some cases but the risk of loss of information should be considered.
- The staff survey proved useful in an unexpected way. Many staff members seemed to be unaware of common HRQL consequences of treatments. This indicates that there may have been deficits in the information given to patients, and calls for research concerning the sufficiency of information given to cancer patients today.

## 4.7 The general population study

### 4.7.1 Participation

The participation in the two parts of the general population study (which were part of the present study and a parallel study investigating gynaecological cancer, respectively) was strongly different. Of the 872 women contacted as part of the present study, 608 (70%) responded (Table 1, paper V) (unfortunately, in the table '872' was erroneously replaced with '860' and '70%' with '71%'.) The table also shows the participation for the different age strata; there was no significant association with age (Chi-square=0.96, p=0.34).

In contrast, the participation in the parallel study intended to produce reference values for gynaecological cancer patients was much lower (49%) (paper V); therefore, the data from the two studies were not merged in the reporting. Paper V therefore includes EORTC QLQ-C30 data from the participants selected to produce a control group for the present breast cancer study only, and only this part of the general population study is discussed here.

### 4.7.2 Results for EORTC QLQ-C30 (paper V)

The mean scores (overall and age-stratified) for each of the 30 items of the EORTC QLQ-C30 are shown in Fig. 1 of paper V. Fig. 2 of paper V shows means scores for the 9 multi-item scales. The study showed that women in the general population sample reported functional limitations as well as symptoms. For example, the mean score for emotional function was 77 (i.e., 23 points lower than 'best possible emotional function'), and the score for pain was 21 (no pain=0).

Scores on the first two items of the physical function scale (strenuous activities, taking a long walk) clearly declined with age corresponding to increasing limitations in older women. The large differences in these two items translated into an overall age association for the five-item physical function scale. A similar tendency was seen for the role function scale (particularly item 6, limited in work) although a more positive score was seen in the oldest age group. However, the latter was probably an artefact reflecting that these women did not experience work-related limitations because they were above the retirement age. Other items showing increasing problems with age included item 10 (need to rest), item 25 (difficulty remembering), and the pain items 9 and 19.

In contrast, the emotional function scale (items 21-24) showed a trend towards increasing (better) scores with age.

### 4.7.3 Results for HAD Scale (paper VI)

Results for the Hospital Anxiety and Depression Scale are shown in Table 1 of paper VI. The overall mean score for anxiety was 6.0. According to the authors' criteria [46], only 68.9% had scores classified as 'non-cases', 18.7% were classified as 'doubtful cases', and 12.4% as 'definite cases'. In other words, almost a third were classified as having possible psychological morbidity. The depression scores were lower. The mean score was 3.4, and 8.0% were classified as 'doubtful cases', and 3.5% as 'definite cases'.

Figure 1 in paper VI shows that the associations with age were not perfectly linear. However, statistically significant, opposite trends were seen for anxiety and depression; anxiety tended to decrease with age, whereas depression increased, at least in women above 40 years of age.

#### 4.7.4 Results for DBCG 89 Questionnaire (Appendix D)

Mean values for 12 items of the DBCG 89 Questionnaire are shown in Table D1 (Appendix D). Items referring to chemotherapy (e.g., hair loss) and surgery (e.g., current level of energy compared to before the operation) were not used in the general population study, as they were not meaningful.

As for the two other questionnaires the study showed at least some symptomatology and/or functional limitation for all items. The menopause-related items such as hot flushes, presence of menstrual bleedings, regularity of bleedings, incontinence, and vaginal dryness showed clear relationships with menopause. The same is the case for the two items on sexuality. Employment declined with age.

#### 4.7.5 Discussion

One may view a general population sample as 'healthy controls' and might thus anticipate absence of problems. However, as illustrated in these data, quite significant levels of symptoms and functional limitations were seen.

This is of course not surprising. The Danish health and Morbidity Survey 2000 showed that 49% of the population reported at least one disease [101]. In Swedish and Norwegian general population studies, using other measures of disease, only 23% and 32% of the participants, respectively, did not report any health problems [102, 103].

Therefore, also participants in studies of e.g. breast cancer are likely to have health problems other than cancer and this emphasises that general population studies are indeed needed in order to interpret results from studies of specific groups of patients. Otherwise, all symptomatology reported by the patients might be interpreted as if caused by the disease or its treatment.

Our study was the first to collect responses to the EORTC QLQ-C30 from a general population sample. Since then similar studies have been carried out in Norway [102, 104], Sweden [103, 105], Germany [106], and Korea [107].

The participation in the Danish sample (70%) was relatively high; slightly higher than in studies using the EORTC QLQ-C30 in Norway (68%) [104] and Germany (67%) [106] but lower than the impressive participation obtained in Sweden (78%) [105] (the participation in the Korean study was not described). A value of 70% is probably satisfactory for a general population study, although it is markedly lower than the extremely high participation seen among breast cancer patients in this study. It is not surprising that a clinical study asking about the patient's wellbeing in the context of breast cancer may appear more relevant than a general population study. A subsequent Danish study using a different questionnaire was carried out in exactly the same way, and had a slightly lower response rate of 67% [108].

This relatively high level of participation reduces the risk of selection bias. However, it is still possible that there is insufficient representation of individuals who do not read and write well, are severely ill or handicapped, have psychiatric disease, etc. On average, individuals in these groups probably have poorer quality of life than the participants. While opposite effects (e.g., the most healthy are busy with other things and might not prioritise participation in a survey like this) might also be present, it seems likely that more non-participants experience quality of life below the average, and that general population surveys such as this therefore over-estimate quality of life.

Fayers compared the Danish, Norwegian, Swedish, and German general population studies using the EORTC QLQ-C30, and found notable differences [109]. Whereas the Scandinavian data were

fairly similar, he noted a markedly stronger decline with age in global quality of life in Germans than in Scandinavians. Similarly, fatigue scores increased more with age in Germans. The same tendencies could be seen in international reference data for the SF-36 questionnaire thus confirming the finding [109]. The Korean data [107] are not directly comparable due to adjustment for a number of background values.

General population studies have also been conducted for the HAD Scale in Sweden [110], the Netherlands [111], Germany [112], the United Kingdom [113], and Australia [63]. The anxiety mean scores were 6.0 in our study against 4.8 in Sweden, 5.1 in the Netherlands, 5.0 in Germany, 6.1 in the UK, and 8.2 in Australia. For depression the corresponding scores were 3.4 against 3.8 in Sweden, 3.4 in the Netherlands, 4.7 in Germany, 3.7 in the UK, and 4.2 in Australia. However, differences in sampling methods, age distributions, and in the reporting of results make it difficult to compare the results directly. However, anxiety scores seem to exhibit larger international differences than depression scores, which are remarkably similar. The Danish anxiety score is similar to or lower than the two English-speaking countries and higher than the other countries. In contrast the Danish depression score is one of the lowest. The important point – which will be further discussed in section 4.8.3, which compares the data against breast cancer patients' scores – is that the Danish general population scores appear to be similar to those from other countries. Comparing the age-relations in the Danish data it is notable that while the EORTC emotional function scale shows a trend towards increasing emotional function with age, the HAD anxiety scale showed a similar decrease in anxiety with age, whereas the depression score increased with age. One interpretation of this could be that the EORTC emotional function scale is functionally closer to anxiety than depression. Another interpretation, which probably explains part of the increase with age, is the age bias in the HAD depression subscale discussed in section 4.4.2.

#### 4.7.6 Conclusions

- The general population study was carried out with a relatively high level of participation (70%). This high participation limits but does not exclude the possibility of selection bias.
- Both the two standard questionnaires and the newly developed DBCG 89 Questionnaire showed considerable 'morbidity' (functional limitations and symptoms) in women randomly selected from the general population. In broad terms, the results are similar to those from the subsequent international studies using the EORTC QLQ-C30 and the HAD Scale.
- These results, which show that parts of the symptomatology reported by cancer patients may have causes other than cancer and cancer treatment, constitute a good justification for the use of general population samples in the interpretation of data from cancer patients.
- As shown in the next section, the interpretation of HRQL data from general population 'control groups' may be more difficult than anticipated.

### 4.8 Anxiety and depression in low-risk breast cancer patients compared to a general population sample (paper VI)

#### 4.8.1 Participation

Patients (pre- and postmenopausal women in DBCG 89 protocol A)

The selection of patients is described in paper VI. Of the 538 patients invited to participate, 468 (87.0%) completed the questionnaire (89.2% in the age group 30-39 years, 77.9% in patients aged 70-75 years).

#### General population sample

The participation in the general population sample was described in section 4.7.1. The women in the general population sample and the breast cancer patients were not significantly different with regard to marital status, number of children, employment status or social class (paper VI).

#### 4.8.2 Results

The surprising main finding was that the scores for anxiety and depression were significantly higher in the general population sample than among low-risk breast cancer patients. DIF analysis identified DIF in one depression item. Correction of this DIF diminished the difference but it was still significant (Appendix A and section 4.4.2). The interpretation of the DIF analyses suggested that similar bias may affect four other items of the depression scale but such bias affecting the majority of items in a scale cannot be detected with the DIF method used here (section 4.4.2).

The validation of the HAD Scale (Appendix B and section 4.5.2) confirmed that as discussed in paper VI, selective reporting may be one of the likely explanations; see further discussion below.

#### 4.8.3 Discussion

Originally, we wrote a manuscript reporting general population 'norms data' only. Just before submission of the manuscript the analyses for the next paper, the comparison of the general population data to low-risk breast cancer patients, were carried out. As these analyses revealed the surprising results reported in paper VI, submission of the manuscript was deferred until these results had been interpreted. Given that the subsequent interpretation seriously questioned the use of 'norms data' we decided to drop the separate publication of general population data and to report these data along with the data from low-risk breast cancer patients (paper VI) instead.

Breast cancer is probably the disease that has been most extensively investigated by psychosocially interested researchers [114] and anxiety and depression are some of the most frequent outcomes in such studies. The belief that breast cancer patients suffer from anxiety and depression is well-established [18]. When writing paper VI numerous papers reporting that breast cancer patients have anxiety and depression could be identified [115-128].

Therefore, the findings in paper VI were extremely controversial. We discussed a range of possible explanations of the unexpected findings and concluded that the comparison of responses from a general population sample against breast cancer patients was probably not valid due to primarily selective reporting and response shift (paper VI).

The DIF analyses (Appendix A and section 4.4.2) and the validation of the HAD Scale (Appendix B, table B1 and B2, and section 4.5.2) were included in the originally submitted version of paper VI but had to be omitted to shorten the paper according to comments from reviewers; brief conclusions from these studies were instead used as 'speculations' in the interpretation of results (paper VI, p. 527-8). Thus, these results do not add new aspects to the discussion as they were already taken into account in our conclusion.

The HAD Scale has been widely used in the last few years. In paper VI we reported that according to a MEDLINE search up until August 1998 there were 267 publications with the HAD Scale in the title or abstract. By October 2006 the number was 1,800 publications, and 83 of these concerned breast cancer. However, only one other study in primary breast cancer including a general population sample has been identified. An Australian study also used the HAD Scale and replicated our findings of higher levels of anxiety and depression in a general population sample than in a cross-sectional study of breast cancer patients 2-43 months after diagnosis [63, 129]. Furthermore, they carried out DIF analyses. Osborne et al. carefully discussed their findings taking our discussion and conclusions into consideration. They ended up with two options, (1) 'The present study and that of Groenvold, each with a reasonably strong experimental design, have demonstrated that women with breast cancer tend to have lower anxiety and depression than population women. This observation remains to be explained.' and (2) 'An alternative explanation is there may be mechanisms whereby women diagnosed with breast cancer may interpret all or most items on the HADS from a different reference point to women without breast cancer.' [63].

Thus, in contrast to our more unequivocal conclusion stating that the results obtained with the HAD Scale in a general population sample are not directly comparable with results from breast cancer patients (i.e. Osborne's option 2), Osborne et al. were open to the possibility that breast cancer patients may be experiencing less anxiety and depression.

At the present point in time it is probably not possible to fully resolve this discussion. In my view, there is strong arguments suggesting that the HAD Scale may not be a valid method for comparing anxiety and depression in breast cancer patients against that of general population samples. To further clarify this question it would be highly relevant to carry out further validation of the HAD Scale when applied to general population samples.

#### 4.8.4 Conclusions

- This study made the unexpected finding that breast cancer patients had lower levels of anxiety and depression than women selected at random from the general population.
- After careful considerations we concluded that this finding is probably incorrect.
- The most likely explanation is that the use of the HAD Scale to compare the two groups of women, who are in markedly different situations, may not be valid. Several potential sources of bias have been identified, including the wording of particular HAD Scale items, the phenomenon 'selective reporting', and the response-shift problem.
- Further validation of the HAD Scale, particularly in healthy respondents and among participants in general population surveys, is needed.
- Intuitively, general population samples are attractive in the interpretation of HRQL from patients because they allow estimation of 'excess morbidity caused by the disease compared to controls'. However, before such use can be recommended, the comparability of HRQL data from patient populations against general population samples must be further evaluated.

#### 4.9 Impact of chemotherapy on quality of life (paper VII)

##### 4.9.1 Participation

Control group (the premenopausal patients in DBCG 89 protocol A). Of the 199 low-risk breast cancer patients contacted, 181 were alive and recurrence-free two years postoperatively, and 148 of the 181 patients (81.8%) had completed all six questionnaires.

Chemotherapy group (the premenopausal patients in DBCG 89 protocols B (arm 1) and D (arm 1)). Of the 242 patients invited to participate, 204 were alive and recurrence-free two years postoperatively, and 159 of these 204 patients (77.9%) completed all six questionnaires.

The women in the chemotherapy group and the control group were not significantly different with regard to age, marital status, number of children, employment status, or social class. The patients in chemotherapy differed from the control group by having breast cancer with less favourable prognosis (paper VII, Table 1).

##### 4.9.2 Results – HRQL in the chemotherapy period

Table 2 in paper VII summarises the results, and Figs. 1 and 2 in paper VII present the longitudinal comparison of HRQL. Based on the literature we hypothesised that patients in chemotherapy had a worse quality of life on 30 variables. Worse HRQL was defined as significantly ( $p < 0.01$ ) worse scores at two or three points in time during chemotherapy. Twenty-three of these hypotheses were confirmed.

##### Confirmed hypotheses

Figs. 1 and 2 (paper VII) present the mean scores over time in both groups. There were many large differences between groups – the differences of 15 points for cognitive function, 17 for social function, 20 on overall health status/quality of life, 28 for fatigue, and 27 for nausea and vomiting exceed the 10 point difference, which has been suggested to represent a clinically meaningful difference on the EORTC QLQ-C30 [130]. Similarly, large differences were seen in Fig. 2 for hot flushes/sweats, anticipatory nausea, energy, (ir)regular bleedings, weight gain, hair loss, sexual interest, and vaginal dryness.

Table 2 in paper VII also shows the differences in frequencies of 'symptoms'/'impairments' for the variables for which our hypotheses were confirmed. Again, many of the differences were relatively large, for example 43% compared to 16% reported that their cognitive function was below the chosen threshold; 80% compared to 31% were fatigued, and half of the patients had developed some degree of anticipatory nausea.

The results concerning psychological distress were mixed. Contrary to expectation, the EORTC emotional function and the HAD anxiety scales did not show significant differences between groups, whereas patients in chemotherapy were significantly more depressed according to the HAD depression scale.

The results concerning sexuality were also mixed. The hypothesis that patients in chemotherapy had less sexual interest was confirmed but the frequency of sex was not significantly lower in the chemotherapy group. Many patients in chemotherapy reported vaginal dryness.

##### Hypotheses not confirmed

Significant differences were not seen for financial difficulties, employment, or full-time work. The expected difference in vaginal flux was also not found.

##### 4.9.3 Results – HRQL after the chemotherapy period

Specific hypotheses were not put forward concerning the HRQL assessments after the patients in chemotherapy had completed their treatment, i.e. for the assessments at 9, 15, and 24 months. For seven variables significantly poorer HRQL was seen in the chemotherapy group at all three 'off-chemotherapy' assessments: hot flushes/sweats, anticipatory nausea, irregular bleedings, hair loss, sexual interest, vaginal dryness, and weight gain. Significant differences in the same direction were found at some of the assessments for cognitive function (at nine months), difficulties sleeping (9, 24 months), financial difficulties (15 months), amenorrhoea (15, 24 months), and urinary incontinence (9, 24 months) (paper VII, Figs. 1 and 2).

##### 4.9.4 Results – staff survey

In this part of the study, the staff survey was used not to generate hypotheses but to elucidate the health care professionals' knowledge and experiences. The upper half of the middle column of numbers in Table 3 in paper VII shows the proportion of respondents who did not expect a difference between groups but where such difference was found (and was hypothesised based on the literature review). For example, 13 of 35 staff members (37%) responding to the first question answered that there was no difference between patients in chemotherapy and patients not in chemotherapy in their ability to work and do household jobs. If the results from the HRQL study are viewed as the truth (at least for the variables where a difference was hypothesised and confirmed) then the staff members not expecting such differences between groups can be viewed as being wrong. Thus, almost two thirds of the respondents failed to recognise that patients in chemotherapy may be dissatisfied with their appearance. Other marked discrepancies concern cognitive function, social function, and weight gain<sup>1</sup>.

##### 4.9.5 Discussion – challenges in reviewing the literature for 'effects of chemotherapy'

A large amount of literature concerning adjuvant chemotherapy and HRQL is available. For example, a PubMed search using the search words 'breast neoplasms AND adjuvant AND chemotherapy AND quality of life' resulted in 337 references (1 February 2007). Before discussing our results against the literature some points about the challenges of reviewing such research will be made.

When results of a study like ours are to be interpreted a key question is: what is the current knowledge concerning effects of breast cancer adjuvant chemotherapy on HRQL? 'Effects' imply causality. Randomized, double blind trials may be optimal for detecting 'effects' but are not available. Blinding is not possible for obvious reasons, and few studies randomizing patients between chemotherapy and no treatment have involved HRQL assessment. This means that optimal data elucidating the impact of chemotherapy on HRQL are sparse.

A review of the available literature was undertaken with two aims:

- 1) If possible, to formulate clear inclusion criteria characterizing eligible studies, i.e. studies that could appropriately claim to describe 'effects of chemotherapy'.

<sup>1</sup> The results for financial difficulties should not have been presented in the section of Table 3 (paper VII) dealing with confirmed hypotheses because the hypothesis was not confirmed.

- 2) To extract the results from the eligible studies: what 'effects' had been reported?

The formulation of methodological inclusion criteria proved difficult. Many studies were ineligible for obvious reasons but many could make reasonable claims about likely effects of chemotherapy. Few studies were ideal, and relying on these would exclude most of the literature. Therefore, the 'middle group' of 'reasonable' studies had to be scrutinized and these proved to be heterogeneous.

The basis for statements of effect can be summarized as follows:

- Within-patient change over time, e.g. an increase in the intensity of a symptom from a pre-treatment assessment to an on-treatment assessment. Advantages of such studies include that compared to other designs the risk of confounding may be smaller. An important disadvantage is that the pre-treatment ('baseline') assessment is rarely a true baseline because patients may already be severely affected by the diagnosis and other treatment (typically surgery). Therefore the change from pre-treatment to on-treatment is likely to under-estimate effect and can lead to false-negative conclusions.
- Comparisons of groups of patients varying as to exposure to chemotherapy. In some cases multivariate analyses have been used to control for possible confounders, in other cases patients were matched with 'controls'. Most studies were cross-sectional but some were longitudinal.
- Studies of chemotherapy patients only. Most of such studies cannot be used because one does not know whether problems are caused by chemotherapy. However, there may be exceptions, e.g. when patients have been carefully interviewed throughout the treatment and have been asked to report the symptoms they experienced due to the treatment. The study by Love [131] is a good example; its results appear reasonably robust despite the lack of a control group.
- Studies comparing two chemotherapy regimens. Unless, as described above, the study reports a comparison with a baseline assessment, the evidence that can be drawn from such studies depends on whether differences are found between groups. When no differences are found, no evidence as to chemotherapy effects can be extracted unless it is evident that the results differ from what is 'normal' (e.g., if 50% of the patients in both groups report vomiting). Studies finding differences between groups may be useful. For example, if two durations of chemotherapy are compared, the period where one group is on treatment and the other group is off treatment may provide valuable information. Particularly if it is a randomised trial, such a design may provide relatively solid evidence concerning effects of chemotherapy because the risk of confounding is minimal. However, it is also evident that such a design has its weaknesses because both groups have had chemotherapy and thus its sensitivity and ability to describe the course of problems/symptoms over time is limited.

In addition to the problems related to definition of clear criteria defining whether sufficient evidence for effect is available, a number of other difficulties are encountered in reviewing the literature. Chemotherapy regimens vary as to drugs, drug doses, drug administration (e.g. intravenous vs. oral), intervals between cycles (e.g., weekly, three-weekly, four-weekly) and number of cycles (and thus duration of treatment), etc. Many articles do not describe the nature of regimens sufficiently. Furthermore, patients treated according to different regimens may be included in

the same study but separate results per regimen are not presented. Again, if strict requirements are used, many articles will be excluded.

Another concern is the HRQL methods used. Were they sufficiently validated? Were they comprehensive? As documented in section 4.2 most, if not all, instruments used are not comprehensive: they assess subsets of the relevant HRQL aspects only.

A long list of additional criteria describing the 'usual' epidemiological criteria concerning the conduct of the studies could be made and used in the selection of studies, e.g. criteria describing the representativity of the samples, possible sources of bias, etc. Thus, the available studies are extremely heterogeneous. If strict, uniform inclusion criteria were applied, a large proportion of the studies would be excluded. Thus one has to choose between making a 'strict' review of very few papers using well-defined inclusion criteria (e.g., including the papers investigating a particular treatment regimen only or including only papers with a very strong design in terms of causality) and making a more inclusive review of a broader range of papers.

Based on these observations it was decided to carry out the review 'manually' without strict methodological criteria. Instead each paper was individually reviewed and articles fulfilling the following overall question were included '*does this study provide reasonably solid evidence concerning the presence (or absence) of HRQL effects of commonly used chemotherapy regimens?*'

The only relatively strict criteria used were that studies of breast cancer adjuvant chemotherapy were excluded if

- They reported 'cases only' – unless other characteristics of the studies (e.g., a change over time) made inference about a 'chemotherapy effect' plausible.
- Patients were treated with regimens other than CMF, CEF, CAF, AC, etc. Thus, studies of taxanes were excluded but a few studies of mixed regimens (with a small proportion of patients treated with taxanes) were included. Studies of high dose chemotherapy were also excluded. A single exception was made concerning a study including several HRQL variables; these patients were treated with NCF (mitoxantrone instead of M or E or A) [132]

If subgroups of patients treated with drugs having specific, well-known side effects (such as neuropathic side effects of vincristine) were included in a study, such side effects were omitted when extracting results from the study.

When studies judged to be 'borderline' were encountered, the originality of the findings was also considered. If, for example, similar results had been found in several studies that should definitely be included in the review, the threshold for including a study was higher than if it reported new findings.

Another point noted during the review was that many studies used several extensive questionnaires in parallel but placed their focus on positive findings. Such possible publication bias is well known [133, 134], and in a review like this there are at least two mechanisms that may exaggerate such bias towards identifying positive effects only. First, for practical reasons it is virtually impossible to extract all the negative findings, which have often not been reported in detail. Second, when negative results are found, the power of the study to detect differences between groups is crucial. Negative findings in small studies having low power are of little interest. The present review of the literature should therefore be seen in the light of there being a risk of false-positive findings.

Finally, the delineation of 'HRQL' must also be considered. In general, only patient self-reported data concerning the patients'

experiences were included. 'Toxicity ratings' made by the physicians in many clinical trials were not included. Similarly, side effects such as neutropenia, infections, cardiotoxicity, second malignancies, etc. were excluded. However, a few exceptions were made for relatively 'objective' issues such as weight gain and amenorrhea, which were included in our questionnaire but could equally well be measured objectively or assessed by a physician.

#### 4.9.6 Discussion – HRQL in the treatment period

Tables 4 and 5 represent the results of a review undertaken as described in section 4.9.3: have similar results been found in other studies? First, each of the findings of our study will be discussed in relation to the literature – in the order they are listed in Table 2, paper VII. Table 4 is a graphical summary of these findings. The terms (names of multi-item scales and the words used in single items) used in our study will be used, but it will be pointed out when only partially overlapping terms were used in the other studies. This will be followed by a discussion of HRQL aspects investigated in other studies that were not assessed in our study (Table 5).

##### Physical function (PF)

We did not hypothesise an effect on PF but found slightly lower PF in the chemotherapy group at three months. A similar effect (a decrease from the pre-treatment assessment) was observed after standard CEF chemotherapy and (more pronounced) after dose intensive CE [141].

##### Role function (RF)

RF assesses limitations in relation to work and household jobs, and we found a clear difference between groups during chemotherapy. Impact on the RF scale was also observed in the aforementioned international trial [141] and in a small German study [151]. Similar results have been reported in some of the first studies from the United States [164, 165] and Canada [142], and again recently from Canada [168], although the latter study employed the FACT-G functional scale, which, in addition to work, household jobs, etc. includes sleep, acceptance of illness, enjoyment of life, and overall quality of life [175]. In interviews with 21 women selected from the same study [168] it was reported that fatigue, nausea, and cognitive problems interfered with the ability to work full-time; about 30% were able to work part-time during chemotherapy [148].

##### Emotional function (EF)

Contrary to expectation we did not find a difference between groups on the EF scale (whereas we found increased depression according to the HAD Scale). This contrasts with an American interview study where 10 of 21 patients reported emotional problems due to chemotherapy [138] and a German study, which found a almost dramatically reduced EF scale score of 50 in patients in combined chemotherapy and radiotherapy compared to 68 (endocrine therapy and radiotherapy) or 62 (radiotherapy alone) [151]. This latter study is limited due to small numbers of participants (N=41 in the chemotherapy group) but this does not explain the low mean scores. EF scores in our chemotherapy group were not below 75.

Whereas the EORTC EF scale used in the studies quoted above includes tenseness, worries, irritability, and feeling depressed, a Canadian study reported reduced emotional wellbeing according to the FACT-G (sadness, coping with illness, hope, nervousness,

worry about dying, and worry that the condition gets worse) [175]. Two large International Breast Cancer Study Group (IBCSG) trials convincingly showed an effect of CMF on mood (a scale ranging from 'happy' to 'miserable') and emotional wellbeing (the 28-item Befindlichkeits-Skala) [156]. Lowered mood associated with chemotherapy was also found in a later trial from the same group randomising between CMF and tamoxifen [140]. It is possible that this 'mood scale' is conceptually closer to the EORTC overall quality of life scale (which showed clearly scores in the chemotherapy group, see below) than to the EF scale. A small study reported higher psychological distress (Psychological Adjustment to Illness Scale [176]) in relation to chemotherapy [154]. Two early American studies reported impaired mood [158] and described that patients in chemotherapy experienced nervousness, irritability, tearfulness, etc. [164, 165]. A Swedish study, however, found no effect of chemotherapy on anxiety or depression as measured by the HADS Scale [167].

These somewhat conflicting results may suggest that the methods used in our study (particularly the EORTC emotional function scale and the HAD anxiety scale) may not capture the psychological consequences of chemotherapy (whereas the overall QL scale did). However, the concept 'psychological consequences' may not be sufficiently precise. This term, along with others used here and elsewhere, e.g., psychological distress, (effect on) mood, etc. need further clarification.

Along with the other results, our findings of no difference between groups in anxiety but a small difference in depression suggest that while chemotherapy is clearly distressing, it relatively rarely leads to major psychological distress (the incidence of depression was 9% in chemotherapy patients compared to 5% in controls). This may seem paradoxical but may be explained by the fact that the treatment may be experienced as meaningful – as a way of fighting the risk of recurrence.

##### Cognitive function (CF)

We found lower (self-reported) CF in the chemotherapy group during treatment (and at three months after completion of chemotherapy). The scale comprises two items on the ability to concentrate and memory, and this was the only scale showing differential item functioning (DIF): most of the difference reflected concentration difficulties; the difference in memory problems was much smaller (section 4.4.3). A difference in the EORTC CF scale was also found in a German study [151]. An early study also based on self-report found impaired ability to concentrate and ability to write following CMF [142]. A Canadian study found moderate to severe cognitive impairment in 15 of 31 patients in chemotherapy compared to four of 36 healthy controls [143]. In a recent American study patients in chemotherapy reported more forgetfulness, difficulty concentrating, and being easily distracted [152]. Australian patients named 'trouble concentrating' as the sixth most troublesome aspect of chemotherapy [149]. A recent Canadian study found that 16 of 100 patients in chemotherapy compared to four of 100 controls had moderate to severe cognitive dysfunction [168]. A sub-group of 21 patients from the study was interviewed in detail about their perception of the impact of chemotherapy on cognitive function: 'Patients reported changes in short-term memory, concentration, verbal fluency, processing speed and to a lesser degree planning and visual-spatial abilities. Almost all patients (20 of 21) reported difficulty with recent memory. Participants described increased forgetfulness (of names, words, places, and appointments) and slower memory retrieval.' [148]. In contrast, a Swedish study found no effect of

chemotherapy on self-reported cognitive function [167]. Cognitive function is further discussed below in the part of this section dealing with possible chemotherapy effects after treatment.

#### Social function (SF)

As expected, we found that SF (physical condition or medical treatment interfered with family life or social activities) was considerably lower in the chemotherapy group. This was also found in a small German study [151]. Relatively small, early studies have found that CMF affected family and marital relationships [164, 165] and family life and social life [142].

#### Global physical condition/quality of life (QL)

Our study showed markedly lower scores for 'overall physical condition and quality of life' during chemotherapy; the same was found in a recent German study [151] and in the international study of CEF, where overall QL dropped from baseline [141]. One of the first studies of HRQL, from Canada, also found worse 'global quality of life' in patients receiving CMF [142], and a recent, large, randomised study comparing CMF to tamoxifen found poorer 'subjective health' in the chemotherapy group [140].

#### Fatigue

Fatigue, nausea, and hair loss are the most well documented HRQL consequences of chemotherapy. Increased fatigue after chemotherapy was found in all studies assessing fatigue. We showed a marked impact on fatigue (80% reported 'a little' or more compared to 31% in the control group). Some of the studies investigated fatigue in detail. A recent American study [136] used a diagnostic interview to identify cancer-related fatigue according to four proposed ICD-10 criteria (duration, impact, cancer-related aetiology, and that fatigue is not primarily a consequence of a psychiatric condition) [177, 178]. The overall prevalence of cancer-related fatigue in breast cancer patients undergoing adjuvant therapy was 26%, and with an OR of 2.23 it was higher in patients receiving chemotherapy [136]. A Dutch study investigated the course of fatigue during and after chemotherapy [146]. Increasing fatigue during chemotherapy was found, whereas another study reported that the level of fatigue was constant [157]. Australian breast cancer patients identified fatigue as the second most troublesome aspect of chemotherapy [149]. A study of mixed cancer patients (44% received adjuvant chemotherapy for breast cancer) showed that while 86% had experienced tiredness only 8% had expected this symptom after chemotherapy [131]. Surprisingly, a Swedish study found no effect of chemotherapy on fatigue [167].

Recent Danish research is evaluating whether physical exercise reduces fatigue; encouraging results have been found [179]. A recent Cochrane review concerning exercise in breast cancer adjuvant therapy found a non-significant improvement of fatigue. The authors concluded that additional research was needed to clarify the effect of exercise [180].

#### Nausea and vomiting (NV)

All studies assessing nausea and vomiting after chemotherapy found this. In addition to the studies reported here many articles reporting results of antiemetic treatment in breast cancer adjuvant therapy have been published. The issue is therefore not whether patients experience NV but to what extent. The medical treatment and prevention of NV (antiemetics) has been strongly improved the last years following the launch of the drug ondansetron in the beginning of the 1990s [181, 182] Additional drugs

have been developed, and combinations with corticosteroids have proven effective [182-185]. As a result, the problem of NV has decreased (and thus, the problem today is probably not as pronounced as shown in paper VII), and the relative importance of this side effect of chemotherapy has probably diminished. This is an important achievement that can be seen in the light of the situation described in the interviews of the initial pilot study (paper I): of 12 patients in chemotherapy, eight reported frequent vomiting and constant nausea for 48 hours or more (up to 10 days) following each treatment; of the remaining four patients, only one did not vomit but still suffered from severe nausea. However, the problem of nausea and vomiting is still important. For example, a study with 49% breast cancer patients enrolled patients in 2001-2002; patients received antiemetic treatment according to guidelines; 37% had acute nausea and 52% had delayed nausea after moderately emetogenic chemotherapy. The study also showed that nausea affected several aspects of the patients' daily lives [186].

#### Pain

No difference in pain scores was observed between groups in our study but pain in relation to chemotherapy has been reported in other studies [132, 141, 169] and in a recent review [173]. The nature of pain is usually not described but could be mucositis (e.g., sore mouth as found in our study) or problems with needle injections, which has been rated as the 9th most troublesome aspect of chemotherapy [149]. In addition to sore mouth, a French study evaluating NCF chemotherapy (corresponding to CMF with mitoxantrone instead of cyclophosphamide) reported headache (44%), stomach pain (38%), muscular pain (30%), and articular pain (28%). Each of these was found distressing by at least 40% of the patients experiencing it; stomach pain being described as distressing by 71%. However, the patients also received tamoxifen, which has been associated with muscular and articular pain [132].

#### Dyspnoea

We did not hypothesise a difference in dyspnoea but found this; 36% of the patients in chemotherapy compared to 12% in the control reported 'a little' or more dyspnoea at five months (Table 2, paper VII). A relatively old Dutch study found that 30% of patients in adjuvant chemotherapy compared to 5% of controls reported shortness of breath [170]. A more recent American study reported that patients in chemotherapy had more dyspnoea than those not in chemotherapy [137], and an increase in dyspnoea scores was also seen during standard dose CEF [141]. Only one of 36 participants in our staff survey expected that patients in chemotherapy had increased levels of dyspnoea (Table 3, paper VII).

#### Trouble sleeping

In our study 70% of patients on chemotherapy compared to 35% in the control groups reported 'a little' or more trouble sleeping (Table 2, paper VII). Similar results have been found in several studies (Table 4), and Australian patients rated trouble sleeping as the 8th most troublesome aspect of chemotherapy [149]. There may be several mechanisms involved in sleep disturbances; one of which is insomnia due to hot flushes and sweats resulting from the chemotherapy-induced menopause observed in many premenopausal women [187].

Table 4

'Effects of chemotherapy' during the treatment period. Summary of results extracted using the categories of paper VII. Results that could not be included in the categories are shown in Table 5.

- 'Effect'
- ☒ Differently formulated but probably almost the same topic

Author, year	EORTC QLQ-C30													HA D	DBCG 89 Questionnaire																						
	Physical function	Role function	Emotional function	Cognitive function	Social function	Global health/QL	Fatigue	Nausea/vomiting	Pain	Dyspnoea	Trouble sleeping	Loss of appetite	Constipation		Diarrhoea	Financial difficulties	Anxiety	Depression	Energy	Weight gain	Hair loss	Wig	Anticipatory nausea	Sore mouth	Hot flushes/sweats	Urinary incontinence	Vaginal dryness	Vaginal flux	Amenorrhoea	Irregular bleedings	Frequency of sex	Interest in sex	Satisfaction with app.	Employment	Full-time job		
Paper VII	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		
Andrykowski, 1988 [135]																						●															
Andrykowski, 2005 [136]							●																														
Arora 2001 [137]										●																											
Beisecker, 1997 [138]			☒				●	●								●	●	●	●	●	●			●													
Berglund, 2001 [139]																																				●	
Bernhard, 2004 [140]						☒	●	●									☒																				
Bottomley, 2005 [141]	●	●				●	●	●	●	●	●	●	●	●	●																						
Boyd, 1988 [142]		☒		☒	☒	☒	●	●												●				●												☒	
Brezden, 2000 [143]																																					
Burwell, 2006 [144]																																					
Camoriano, 1990 [145]																			●																	☒	
De Jong, 2004 [146]							●																														
Demark, 1993 [147]																			●																		
Downie 2006 [148]				●			●	●																													
Duric, 2005 [149]				☒			●	●			●					●				●																	
Fairclough, 1999 [150]							●	●																													
Fetting, 1998							●	●																													
Galalae 2005 [151]	●	●	●	●	●	●	●	●			●	●	●	●	●																					●	
Ganz, 2004 [152]				●																				●		●											
Goodwin, 1999 [153]																																					
Hoskins, 1997 [154]			☒																																		
Hughson, 1986 [155]																	●																				
Hurny, 1996 [156]				☒							●					☒																					
Jacobson, 1999 [157]							●																														
Knopf, 1986 [158]							●	●			●	●	●	●	●		☒		●																		
Land, 2004 [159]														●																						☒	
Levine, 1988 [160]							●	●			●									●																	
Love, 1989 [131]							●	●			●	●	●	●	●				●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
Lu, 2007 [161]						☒																															
Macquart, 1997 [132]2							●	●																													
Malinowszky, 2006 [162]																																					
McArdle, 1981 [163]																																					
Meyerowitz, [164, 165]	☒	☒		☒			●	●			●		●					●	●																		
Montgomery, 1997 [166]																																					
Tchen, 2003 [168]	☒	☒		●			●	●															●														
Tierney, 1991 [169]							●	●	●		●													●	●												
Van Dam, 1980 [170]							●	●		●	●	●	●	●	●																						
<b>Reviews</b>																																					
NIH Cons. 2000 [171]				●			●	●											●	●																	
Shapiro, 2001 [172]				●		●	●	●			●						●								●		●										
Moore, 2007 [173]				●			●	●	●																												
Partridge, 2001 [174]				●			●	●											●	●																	

Loss of appetite  
Loss of appetite was seen in 45% of our patients receiving chemotherapy compared to 8% in the control group, and has been

found in many previous studies (Table 4). It seems closely linked to nausea and vomiting.

Constipation  
We did not anticipate constipation but found this in 52% of the chemotherapy group compared to 12% of the controls. Only one of the other studies reviewed has addressed this symptom, and

<sup>2</sup> This study evaluated NCF chemotherapy; tamoxifen was also given.

found it in 19% of the patients [131]. Constipation is a well-known side effect of ondansetron [188], which was used frequently as an antiemetic.

#### Diarrhoea

As expected, diarrhoea was more frequent in the chemotherapy group where 35% reported at least 'a little' diarrhoea compared to 12% in the control group. EORTC QLQ-C30 diarrhoea scores of around 28, that is considerably higher than in our study, were observed following standard dose CEF [189]. Increased frequency of diarrhoea was also reported in a German study [151]. In an American study 37% experienced diarrhoea after chemotherapy [131], and in a French study the prevalence was 17% [132].

#### Financial difficulties

The expected differences concerning financial difficulties were not found. The hypothesis was based on an American study [164, 165] first published in 1979, which reported direct costs (most insurances policies did not cover all costs), lost income, lack of job mobility, and adverse effects on employment status [165]. However, later studies concerning the treatment phase have not addressed this issue. Taken together with the lack of differences concerning work, our results suggest that while some patients may have been on sick leave during treatment, the Danish employment system secures full wages during temporary absence for most patients, and thus the chemotherapy seemed to have no major economic consequences. Furthermore, the chemotherapy did not seem to affect the patients' employment situation, at least not in the short term (see below).

#### Anxiety and depression (HAD Scale)

We did not find the expected increase in anxiety whereas patients in the chemotherapy group were more often depressed according to the HAD Scale (9% vs. 5%). A Swedish study found no effect of chemotherapy on HAD anxiety or depression [167]. A study focused on anxiety did not find higher anxiety in women randomised to chemotherapy [190]. A small group of patients may thus become clinically depressed, whereas for the vast majority – while distressing – the treatment is not associated with psychiatric morbidity. Psychological distress was further discussed in the section on emotional function above.

#### Energy

As shown in Table 4 several studies have assessed fatigue, whereas 'lack of energy' has not been the term used to communicate the findings, except in an interview study [138]. Energy can be seen as the opposite of fatigue; thus, the discussion in the section on fatigue (above) also applies to energy. One important observation made in our study was that, although significantly lower than in the chemotherapy group, the proportion of women not undergoing chemotherapy who reported lower energy than before the operation was fairly high (e.g., 40% at five months).

#### Weight gain

Weight gain (here defined as an increase of at least 2 kg) was frequent in both groups but clearly more frequent in the chemotherapy group where 55% of patients reported weight gain at 5 months (compared to 30% in the no treatment group). Again, this result shows the importance of a control group to avoid attributing all changes to chemotherapy. Many other studies have reported weight gain based on patient self-report or measured objectively. In 1990, a study describing weight gain was published

along with a review of 12 similar published studies. In their study, the authors found that weight gain was associated with increased risk of recurrence and death [145]. Subsequent research has not clarified this issue: while obesity is associated with poorer prognosis, it is still unclear whether weight gain also worsens prognosis [191]. Weight gain seems to be related to the duration of chemotherapy, to onset of menopause (weight gain being more frequent in patients becoming postmenopausal from chemotherapy), to reduced energy expenditure [192], and to reduced physical activity [191, 193, 194]. A recent study proposed that reduced thyroid function following chemotherapy might also contribute [195].

#### Hair loss/wearing a wig

Hair loss and alopecia following chemotherapy are well documented, and included in most studies of HRQL (Table 4). Hair loss has been named the most troublesome [149] or the second most significant/problematic side effect of chemotherapy [138]. Complete alopecia is common following anthracycline-containing chemotherapy (i.e., epirubicin in CEF, adriamycin in CAF, CA) whereas hair loss after e.g. CMF is less severe [172, 174]. In an American study published in 1989, hair loss was experienced by 89% but only 44% had expected this [131]. Few HRQL studies have assessed whether patients wear a wig; in our study only 8% of patients in CMF chemotherapy reported wearing a wig.

#### Anticipatory nausea

Anticipatory nausea is a well known though not routinely assessed phenomenon experienced by 53% of the patients in chemotherapy in our study (compared to 3% of controls). It is thought to develop through a classical conditioning process, and has been carefully studied by psychologists [135, 166]. The problem tends to increase during the treatment period (Fig. 2, paper VII) [166], and is ideally prevented through effective treatment of nausea and vomiting. Psychological [196] and pharmacological [197] intervention have been shown to have some effect.

#### Sore mouth

This symptom was seen in 27% of the patients in chemotherapy compared to 5% of the control group. The symptom has been assessed in a few other HRQL studies [132, 142, 169]. One study reported mouth sores [131] and similarly, studies based on physician-rating of toxicity usually evaluate stomatitis [174]. Mouth soreness/stomatitis is a reflection of the well known consequence of cytotoxic therapy, mucositis [171, 174].

#### Hot flushes/sweats

These menopause-related symptoms were reported by 86% of the patients in chemotherapy compared to 25% of the (premenopausal) controls (Table 2, paper VII). It would have been ideal to investigate the symptoms separately as has been done in other studies, but we chose to combine them to save space. An American study reported that at the end of primary treatment 61% reported hot flushes, and this symptom as well as night sweats were more frequent (percentages not reported) in patients who had undergone chemotherapy [152]. Another American study reported hot flushes in approximately 40% of the women [131] after chemotherapy. In a British study patients reported increases in night sweats, daytime sweats, and hot flushes after standard chemotherapy [162]. In a small, Canadian interview-based study about half of the patients reported being awakened by hot

flushes [148]. Vasomotor symptoms (hot flushes, feeling warm, and sweats) were also found in a Swedish study [167]. Given that many premenopausal patients become postmenopausal as a result of chemotherapy (see below) the menopausal symptoms are certainly not surprising. Our study showed that the menopausal symptoms start shortly after initiation of chemotherapy (Figure 2, paper VII). A parallel increase was observed in sleeping difficulties (Figure 1, paper VII), which may, of course, have other causes as well.

#### Urinary incontinence

This symptom was added to the questionnaire as a possible menopausal symptom following advice from a gynaecologist (paper I). During chemotherapy up to 43% of the patients in chemotherapy (vs. 25% in the control group) reported at least 'a little' incontinence. None of the other studies has assessed this symptom but two studies have reported 'bladder problems' being much more frequent after CMF than AC [159] and that 'urinary problems' were present in 12% of patients after different kinds of chemotherapy [131]. Therapy-related bladder problems may be directly to cytostatics (e.g. a toxic effect of cyclophosphamide on the bladder is well known and necessitates the use of the preventive agent when high-dose treatments are given), whereas indirect effects via endocrine changes are longer-term issues. Given that we have not identified other studies fully confirming our observation, additional research about the occurrence of urinary incontinence is needed.

#### Vaginal dryness

In our study, 42% reported at least 'a little' vaginal dryness' compared to 15% of controls. The three articles listed in Table 4 did not report percentages. Vaginal dryness is associated with menopause and negatively affects sexual function [187].

#### Vaginal flux

We added this symptom to the questionnaire following interviews where it was reported by four of 14 patients (paper I). However, the symptom was not seen more frequently in the chemotherapy group, a finding consistent with a Swedish study [167]. The symptom has not been studied in the other papers reviewed (Table 4).

#### Amenorrhoea

It is well established that chemotherapy leads to amenorrhoea in a significant proportion of premenopausal women [153]. The likelihood of amenorrhoea increases steeply with age [153, 187, 198] as it depends on the follicular reserve [187]. Our analysis of this association in premenopausal women treated with chemotherapy (paper VIII) showed that 25% of women below 40 years, 57% of women aged 40-44 years, 90% of those aged 45-49, and 100% of women aged 50 years or above became amenorrhoeic. These results fit a graphical model of the relationship between age and amenorrhoea closely [153]. Amenorrhoea after chemotherapy has been reported to be permanent [199] but there may be exceptions where menses return after amenorrhoea [187]. In addition to amenorrhoea in direct relation to chemotherapy, young women who preserve their menses may experience premature menopause [187].

Irregular bleedings. Irregular bleedings were experienced by 86% of patients in chemotherapy compared to 30% of controls. This symptom was not assessed in the other studies (Table 4).

#### Frequency of intercourse and interest in sex

Our study suggested that chemotherapy reduced sexual interest but not frequency: 19% of patients in chemotherapy reported no interest in sex (vs. 7% in the control group). Contrary to expectation no difference was seen in frequency of intercourse (Table 2, paper VII). Several studies have reported that sexual function is affected by chemotherapy [139, 144, 152, 159, 162, 164, 165, 168]. The largest decrease in sexual function has been observed in patients who become postmenopausal after chemotherapy [144, 187, 200].

**Table 5**

**Results from studies reporting 'HRQL effects of chemotherapy' other than those included in Table 4, listed alphabetically according to first author. In all reported cases reduced wellbeing was associated with chemotherapy.**

---

#### HRQL effects

---

Physical wellbeing, satisfaction with sex life [137]

Physical wellbeing [140]

Anger, recreation [142]

Nail problems [138]

Menopausal symptoms (a 13-item subscale of the FACT-ES), detailed description of the nature and consequences of cognitive impairment during chemotherapy [148]

Thought of actually having treatment (3rd most troublesome), altered sense of taste (5th most troublesome), problems with needle injections (9th most troublesome) [149]

Physical side effects (BCQ) [202]

Systemic side effects (a subscale of the EORTC QLQ-BR23 Breast Cancer Module including items on dry mouth, food and drink tasting different than usual, painful, irritated or watery eyes, hair loss, feeling ill or unwell, hot flushes, headaches) [151]

Pain with intercourse [203]

'Health problems stand in the way of activities' scale [154]

Physical wellbeing, perceived adjustment/coping [156]

Change in bowel pattern, mobility [158]

Increased production of gas, knowing you must come to clinic to receive treatment; drowsy after chemotherapy; tearfulness; runny, dripping nose; burning, watery eyes [160]

Eye problems, restlessness, weakness, mood swings [131]

Physical wellbeing scale (sleep, energy, pain and physical discomfort, eating, sexual functioning, sensory function, capability of daily living) [161]

Headache, stomach pain, muscular pain, articular pain, skin rash, weight loss, cystitis [132]3

Sexual pleasure, sexual discomfort, aches and pains [162]

Nervousness, irritability, tearfulness, flu-like symptoms [164, 165]

Sickness, sore eyes, heartburn, taste change [169]

Spiritlessness, shivering [170]

---

<sup>3</sup> This study evaluated NCF chemotherapy, tamoxifen was also given.

### Satisfaction with appearance

Hair loss and other consequences of chemotherapy may obviously affect body image, which may already be impaired due to breast surgery. In our study 65% of the patients in chemotherapy compared to 46% in the control group reported a score of 5 or less on a 7-point scale for satisfaction with appearance (Table 2, paper VII). A previous study has reported impaired feeling of attractiveness [142], and a Swedish study reported poorer body image associated with chemotherapy [167]. The other studies reviewed did not assess this issue. Body image is included in currently used instruments such as the EORTC QLQ-BR23 Breast Cancer Module [201].

### Employment and full-time work

As discussed in the section on financial difficulties (above) we did not find any association between chemotherapy and work.

### Other findings

Obviously, parts of the findings in the literature could not be categorised according to our 'system' because HRQL aspects other than those assessed in our study had been investigated. These findings are listed in Table 5. Some findings could not be included in Table 4 because the terms used were less specific (or, at a more global level) than in our study, e.g., 'physical wellbeing'. Such results do not add to the description provided in Table 5. The same is the case when overall results from multi-item scales measuring several symptoms are presented. The findings that chemotherapy was associated with higher scores for 13 menopausal symptoms in the FACT-ES, for eight symptoms in the systemic side effects scale of the EORTC QLQ-BR23, or for seven different HRQL aspects combined in a physical wellbeing scale do not necessarily imply that chemotherapy had an effect on all individual items (unless separate analyses of the items or an overall DIF test showing no DIF was reported). In contrast, other results add potentially important information. First, some specific issues not included in Table 4 have been reported: altered sense of taste [132, 149, 169, 204], nail problems [138]; increased production of gas [160]; runny, dripping nose [160]; eye problems [131, 132, 160, 169]; flu-like symptoms [164]; headache, stomach pain, muscular pain, and articular pain [132] (the latter may also have been caused by tamoxifen); and heartburn [169]. Eye problems and altered sense of taste were each reported in four studies, and eye problems are likely manifestations of mucositis/conjunctivitis. The remaining findings must be viewed as preliminary. Second, additional aspects of some of the issues already included in Table 4 have been reported. This includes psychological aspects (e.g., anger, perceived adjustment/coping, being tearful, mood swings) and aspects of sexuality (e.g., satisfaction with sex life, pain with intercourse). Finally, the 'thought of actually having treatment' (the third most troublesome aspect of chemotherapy) and 'knowing you must come to clinic to receive treatment' are probably good expressions of the mental burden associated with repeated treatments [149]. Similar findings were made in our pilot study (paper I) where the repetitive character of coming to treatments and becoming sick was mentioned as very distressing.

As discussed in the section on emotional function above, the psychological experience and distress experienced by patients in chemotherapy seems insufficiently investigated, at least in the literature reviewed here. The contrast between our finding of no impairment of emotional function after chemotherapy and the

distress reported in various ways (e.g., the 'thought of actually having treatment') calls for conceptual clarification. 'Emotional function' is conceptually closely linked to anxiety and depression, and while it is certainly important to investigate psychological/psychiatric morbidity, it seems extremely relevant also to investigate patients' experience of chemotherapy in more detail. Such research could have important consequences. First, it would clarify whether additional issues need to be added to the measurement tools currently used in order to obtain a sufficient description of the HRQL impact of chemotherapy, which can form the basis for comprehensive information to future patients. Second, increased understanding of the patients' experiences may also be valuable for the clinical management of the problems. Table 5, which reports the findings not fitting into the categories used in our study, thus shows that:

- Relatively few 'new' topics were identified, and most of these were reported in only one or a few reports. Issues that have been reported in one study only are of uncertain importance, whereas altered taste and eye symptoms are relevant topics to include in future studies.
- Some HRQL aspects have been assessed differently in the studies reviewed (e.g., by asking in a more overall or a more specific way, or by focusing on different sub-dimensions of the HRQL aspect). This highlights a need for further research into the nature of the HRQL aspects assessed, a finding that is particularly obvious concerning the psychological aspects.
- It seems surprising that so few HRQL aspects not included in our study have been identified. Our questionnaire was constructed in 1991. One could have expected that the extensive research published subsequently had identified many previously unknown effects of chemotherapy. This has, however, not been the case. Two interpretations can be made of this. The first is that 'almost everything was known at that time'. The second is that the subsequent research has failed to identify the topics we overlooked.

The conclusion drawn in the last section above can be contrasted against the picture shown in Table 4: despite the fact that it can now be concluded that it was possible to compose a reasonably 'complete' questionnaire several years ago, all the subsequent studies have assessed subsets of the issues identified at that time only. The second most comprehensive study published was Love's study<sup>4</sup>. Another relatively complete though small study (39 patients in chemotherapy) used the EORTC QLQ-C30 and the breast cancer module QLQ-BR23 [151]. Of course, many studies may not have attempted to carry out a comprehensive assessment but still it appears surprising and problematic that most studies claiming to investigate chemotherapy effects have assessed so few. This line of thought raises the question whether it is at all relevant to assess many different aspects of HRQL? And if all are not relevant, which should be included? As always, such questions about research methodology can be answered only when the aims are taken into account. HRQL studies may have descriptive aims in order to be able to provide comprehensive information to future patients. In such studies, the question is whether there are any of the topics in Table 4 patients do not want information about. If studies of HRQL are carried out to improve the care of patients, the main issue is to identify and monitor the symptoms or problems that can be prevented or ameliorated through inter-

<sup>4</sup> Love's study was actually conducted prior to our study but was excluded from our initial review because it included lymphoma patients and many different chemotherapy regimens.

ventions, and the selection of topics may be guided by the opportunities for prevention or treatment. If studies have a comparative aim, as in trials comparing treatment alternatives, valid comparisons demand that all HRQL issues that may differ between treatments are included [33]. Based on these considerations it is clear that all studies need not assess all 23 issues found to be affected in chemotherapy patients in our study (+ potential issues reported in other studies, e.g., physical function, emotional function, pain, constipation, anxiety, and additional psychological, sexual, and work-related issues). However, unless specific reasons for restricting the focus can be given it seems difficult to choose among the many issues in Table 4.

It is therefore recommendable that future research more explicitly considers content. Researchers should present arguments and clear criteria for the composition of questionnaires. This study shows that the EORTC QLQ-C30 along with the DBCG 89 Questionnaire or the QLQ-BR23 cover the vast majority of HRQL issues associated with CMF chemotherapy.

In summary,

- In our study we tested 30 hypotheses about associations between adjuvant chemotherapy and HRQL and confirmed 23 of these; 22 of these 23 findings have been found in other studies as well, and can be viewed as well documented. The exception is urinary incontinence, which must be viewed as a preliminary finding.
- In addition, we found that patients in chemotherapy may experience dyspnoea and constipation. Dyspnoea has been found in a few other studies. Constipation may be fully or partially explained by antiemetic therapy.
- Other studies have found additional variables to be associated with chemotherapy. Changes in taste and eye problems have been found repeatedly, whereas other findings must be viewed as preliminary because they have been reported in only one study.
- It can therefore be concluded that there are about 27 symptoms/problems that are likely effects of adjuvant chemotherapy (the exact number depends on the concepts used: the number may be increased by splitting up symptoms or reduced by collapsing them).
- The psychological aspects of chemotherapy do not appear fully clarified. In particular, the methods used in our study may not provide a sufficient evaluation.
- The questionnaires used in most other studies have had inadequate content.

If these findings summarised above are contrasted with publications reviewing the HRQL effects of chemotherapy, it is clear that these reviews are missing many findings. As shown in the lower part of Table 4, each of the four authoritative texts is missing many well-documented symptoms and problems [171-174].

#### 4.9.7 Discussion – HRQL after the treatment period

Most of the literature concerning the relationship between adjuvant chemotherapy and HRQL after the treatment period consists of cross-sectional studies of ‘survivors’, typically 2-10 years after their diagnosis. As the time horizon in most such studies is longer than the present study, many of the results are not directly relevant in the discussion of our results for the period 9-24 months after start of treatment (i.e., 3-18 months after completion of chemotherapy). However, as symptoms found e.g. five years after treatment are likely to be present in the previous years as well, some findings from the long-term studies are relevant.

Overall, our results and other research show that some symptoms and problems disappear or diminish, and that others persist.

Thus, a main finding in our data was the dramatic reduction in the number and magnitude of differences between patients in chemotherapy and the control group. Most of the problems associated with chemotherapy had diminished or disappeared at the nine-month assessment (three months after chemotherapy completion). This is clinically important. If Figs. 1 and 2 (Paper VII) are visually compared, Fig. 1 shows a general tendency towards disappearance of differences between the groups from five to nine months, whereas Fig. 2 shows evidence of persistence of problems in the chemotherapy group for many areas. The reduction of impairments in Fig. 1 is pronounced in most of the upper eight panels, which are the functional scales, fatigue and nausea/vomiting. The gastrointestinal symptoms lack of appetite, constipation, and diarrhoea were also reduced to the level of the control group. In contrast, Fig.2 shows that higher levels of hot flushes/sweats, anticipatory nausea, irregular bleedings, amenorrhoea, reduced sexual interest, vaginal dryness, and weight gain persisted in the chemotherapy group throughout the two years. This difference between Figs. 1 and 2 probably reflects that the EORTC QLQ-C30 (Fig. 1) is a broad and general HRQL questionnaire and therefore mainly elucidates the acute toxicity and functional impact of chemotherapy, whereas Fig.2 is the specific DBCG 89 Questionnaire, and includes several items measuring the hormonal consequences of chemotherapy.

The pattern of relatively quick resolution of chemotherapy toxicity and functional impact is consistent with other prospective studies [140, 141, 150, 156, 159].

The other main finding concerns the symptoms/problems that persist. Many ‘breast cancer survivor studies’ have investigated patients at various time points but a large part of these studies have not investigated chemotherapy separately. They have reported cross-sectional studies of breast cancer patients in general, and do not make it possible to separate effects of diagnosis, surgery, and radiotherapy from chemotherapy. The following text will concentrate on studies that had designs that made it possible to elucidate the effect of chemotherapy.

An early, American study consisted of interviews with 27 women who were 21 months after chemotherapy [205]. This study is mentioned here because patients were directly interviewed about their perception of effects of chemotherapy. The patients reported trouble sleeping, weight gain, hair loss (still), anticipatory nausea, amenorrhoea, and loss of energy. The mean time it had taken before they ‘returned to their old selves’ was 6.4 months [205].

In an American interview study with 18 participants six months after chemotherapy patients reported hair problems (N=18), fatigue (N=15), weight gain (N=8), menopausal problems (N=7), emotional problems (N=6), and nail problems (N=6). The order of importance was weight gain (most important), emotional problems, reproductive/menopausal problems, and fatigue [138].

A Swedish randomised trial comparing adjuvant chemotherapy and radiotherapy with follow-up 2-10 years after treatment found that the only difference indicating higher symptomatology in the chemotherapy group was smell aversion (patients randomised to radiotherapy were found to have decreased stamina, more symptoms related to operation scar, and higher anxiety) [206].

A small cross-sectional, American study (N=25 in the chemotherapy group) found that patients who had undergone chemotherapy seven years earlier had more fatigue, weight gain, hot flushes, vaginal dryness, mood swings, dyspareunia, and difficulty reaching orgasm than other breast cancer patients [207].

A much larger American study (mean three years after diagnosis) found no differences in most aspects of HRQL (SF-36) when comparing breast cancer patients to healthy women but found higher levels of hot flushes, joint pains, and headaches, and sexual dysfunction [203]. The relationship to chemotherapy was investigated for sexual dysfunction only, and an association with chemotherapy was found [203, 208]. The study also reported that among women below 50 years sexual dysfunction was most frequent in those who had stopped menstruating [203]. In cross-sectional analyses from the same study, patients treated with chemotherapy had more weight gain, vaginal dryness, poorer sexual function, and more pain with intercourse [209]. The levels of energy were investigated in the same breast cancer data compared to a general population sample and to baseline data from the Breast Cancer Prevention Trial. The ambiguous result was that breast cancer survivors had slightly better energy than women in the general population sample but lower energy than participants in the prevention trial. Among breast cancer patients, fatigued patients were more likely to have been treated with chemotherapy [210].

A more recent American study (mean 10 years after treatment) found lower scores for overall quality of life and on a social subscale (interference with activities at home, financial burden, sexuality) in a subgroup of patients who had received chemotherapy [211].

A Canadian study followed patients in chemotherapy and healthy controls for two years, and found no difference in total FACT-G score after one year but persisting fatigue (FACT-F) and menopausal symptoms (FACT-ES) [212].

A prospective American study followed patients for five years and found a lasting effect of standard dose chemotherapy on menopausal symptoms (sweats, hot flushes, vaginal dryness) and on sexual pleasure and discomfort but not frequency of sexual activity [162].

A recent Swedish study found that chemotherapy was negatively associated with return to work 2-3 years after diagnosis [213] whereas no such association was found in two American studies [214, 215].

A relatively small French study found no effect of chemotherapy (compared to no chemotherapy) at a mean of 9.6 years after chemotherapy [216].

In an analysis from the Nurses' Health Study women diagnosed with breast cancer treated with chemotherapy reported a higher level of sexual dysfunction and a larger decline over time in the SF-36 scales role emotional and vitality compared to other women diagnosed with breast cancer [217].

A recent Korean study showed markedly poorer HRQL in breast cancer survivors compared to a general population sample, e.g. in most EORTC QLQ-C30 scales, but none of these was associated with chemotherapy [107].

A recent Danish study compared the SF-36 scores of breast cancer survivors (N=1,316) 5-15 years after their diagnosis against age-matched reference data from the large Danish Health and Morbidity Survey. There were almost no differences between breast cancer patients and healthy women. The breast cancer patients reported less pain, better general health, but worse mental health; however, these differences were very small. In other scales, younger breast cancer patients reported slightly poorer HRQL than healthy controls whereas the opposite was seen for older women. Remarkably, among breast cancer patients no associations were found between past chemotherapy and any of the subscales [218].

The issue of possible long-term effects of chemotherapy on cognitive function needs separate attention due to the importance of this relatively newly described consequence of treatment and because, despite much recent research, the results are still controversial [219-222]. Studies reporting short-term effects were listed in section 4.9.6.

In 1995, an American study investigated 28 breast cancer patients who had completed adjuvant chemotherapy 1-12 months previously, and made preliminary findings suggesting cognitive impairment [223]. Three years later, a Dutch study reported that 32% of breast cancer patients treated with high-dose chemotherapy had cognitive impairment compared to 17% of patients treated with standard-dose and 9% of control patients [224].

Another study of standard dose CMF chemotherapy from the same group found that 31% of the patients reported problems with concentration and 21% problems with memory compared to 6% and 3% of controls, respectively. Impaired ('objectively' measured) cognitive function was found in 28% of chemotherapy patients compared to 12% of controls [225]. As in other studies [226, 227], self-reported and measured cognitive function were uncorrelated [225]. An American study of breast cancer and lymphoma survivors 5 years after diagnosis found poorer cognitive function in those treated with chemotherapy [228]. Another American study questioned the results from the previously published cross-sectional studies. They investigated 84 breast cancer patients before chemotherapy, and found that 35% experienced cognitive impairment [229]. However, when following up a subgroup of 18 of these women, more than half of them (11) experienced further decline [230]. Two metaanalyses produced somewhat different conclusions but both found evidence of effect of chemotherapy on cognitive function [231, 232]. The most recent included 16 studies of which nine involved breast cancer patients and concluded that the evidence of effect was seen also in the subgroup of breast cancer studies [232]. The three newest, prospective studies have reported conflicting results. In an American study, seven of 28 patients declined in two or more domains from pre- to post-test [233]. A Dutch study compared high-dose-chemotherapy (N=28), standard dose chemotherapy (N=39), no chemotherapy (N=57), and healthy controls (N=60). Only patients in high-dose chemotherapy deteriorated in cognitive function [234]. A British study followed 85 women allocated to chemotherapy as well as controls. No difference in the proportions experiencing decline in cognitive function was found [235]. A parallel paper from the same study reported interview data. Four weeks after completion of chemotherapy, 77 patients (83%) reported memory problems and 73 (78%) concentration problems but one year later, there was no longer any difference when compared to a control group [227]. Finally, the most recent addition to the literature is a pilot study showing promising results for a cognitive-behavioural intervention program against chemotherapy-related cognitive dysfunction [236].

To summarise, the results of the studies of the impact of chemotherapy on HRQL after the treatment period are relatively complex. Looking mostly at the largest, newest, and most well-designed studies, there seems to be little or no effect on the general aspects of HRQL as measured by the SF-36, the FACT-G, or the EORTC QLQ-C30. In contrast, symptoms such as anticipatory nausea, weight gain, and not least endocrine effects (e.g., hot flushes/sweats, irregular bleedings/amenorrhoea, vaginal dryness), disturbed sleep, and sexual dysfunction are well documented. Concerning cognitive function, there seems to be little doubt that many patients experience problems and that in some

patients objective measures support this. However, the frequency, severity, and nature of cognitive problems remain unclear, as do their relationships with types and doses of chemotherapy.

#### 4.9.8 Discussion – staff study

The results of the staff study proved useful as a means of elucidating the knowledge of the physicians and nurses, although the study was not intended for this purpose. The somewhat surprising result was that the staff did not expect many of the ‘well-known’ side effects of chemotherapy. The proportions not expecting these problems were relatively high for role function (37%), cognitive function (53%), social function (47%), overall quality of life (25%), weight gain (31%), hair loss (25%), dissatisfaction with appearance (66%), and impaired sexual life (43%). In contrast, only one of the 36 participants did not expect fatigue and amenorrhea (paper VII, table 3).

As discussed in paper IV, the results seem to indicate that in general the staff understood the exercise correctly – for example, nobody appeared to have misunderstood the direction of the scale by answering that patients in chemotherapy had fewer symptoms than controls. And as stated above, almost all expected fatigue and amenorrhea. The staff responses can be viewed as a description of the ‘maximally possible information given to patients’: it is unlikely that the staff informed patients about side effects they did not expect to occur. Furthermore, it seems unlikely that they had passed on all information they possessed to the patients. Finally, the participants selected for the staff study was the most experienced nurses and physicians. Some patients have been in contact with less experienced staff who had less knowledge. In other words, the results suggest that the oral information given to patients has been incomplete. This assumption also seems probable given the results presented previously in this chapter: the quality of life studies – which were conducted, one must assume, by specialists in the field – have been extremely incomplete in their coverage. Many symptoms and problems shown long ago to be associated with chemotherapy were not assessed in later studies. Similarly, the reviews of side effects in prestigious journal were also remarkably incomplete as illustrated in Table 4 (bottom). Therefore, it is not surprising that the knowledge of the health care professionals and the information given to patients has been incomplete.

Furthermore, it is well known that not all information given to cancer patients is understood and remembered by the patients [237, 238]. Thus, there are two aspects of physician-patient communication that are far from optimal: the information given to patients is incomplete and only parts are remembered. The lack of knowledge of side effects by professionals was also found in a French study. Macquart-Moulin compared 50 patients’ self-report of 17 symptoms throughout chemotherapy against physician-ratings using a standardized sheet at the same time points. Symptoms were systematically and strongly underestimated [132]. For example, 73% of the patients reported nausea but this was noted for 38% only.

Many other studies have investigated the extent of agreement between patients’ assessment of their HRQL and health care professionals’ [79, 80, 89], and have generally found low to moderate agreement although in some newer studies the concordance has been better [80].

The two extremes of the communication process about HRQL effects of chemotherapy, at one end the patients’ experience and at the other end, their expectations, can be studied by comparing experiences against expectations in the same patients. This was

done in an American study. Many patients did not expect even very frequent symptoms. Only 8% expected tiredness whereas 86% experienced it. The corresponding figures were 44% vs. 89% for hair loss, 57% vs. 87% for nausea, 3% vs. 45% for weight gain, 4% vs. 44% for mouth sores, 7% vs. 37% for diarrhoea, and 3% vs. 19% for constipation [131]. The authors concluded that there was a two-way communication problem: ‘Patients not only fail to receive all the necessary information from clinicians, but they also fail to provide clinicians with a complete picture of their treatment experience.’ [131]. Similarly, a Scottish study showed that most side effects experienced by breast cancer patients were not expected by them [169, 204].

Several studies have indicated that patients want to be carefully and completely informed about consequences of treatment [239-247]. The available research data, as discussed above, indicate that this has not been the case. This is an important result. The content of oral and written information to patients today should be critically reviewed. A logical, next step in the research must be further investigation of the preferences of patients for information (i.e., which, and how much information is desired at various phases of the trajectory, and how should it be given?) (see also section 7.2.2.).

Concerning the use of the staff study to elucidate the staff members’ knowledge and information practices it should be emphasised that larger and much more detailed studies are needed in order to investigate these issues sufficiently.

#### 4.9.9 Discussion - overall comments

Today (2007), the results of our study (and many of those reviewed) are relatively old. This must be taken into consideration, and limits their generalisability to the situation today in two ways. First, improvements in supportive care mean that some of the problems, particularly nausea and vomiting, must be expected to be reduced today. Other changes in treatment or other aspects of care may also have taken place, and can of course be positive (e.g. better communication skills) or negative (e.g., less time for the individual patient). The HRQL of patients treated with CMF today may therefore be different, and apart from the likely reduction in nausea and vomiting it is unknown whether other potential differences are for the better or worse. It has not been possible to locate any studies that have replicated the sampling and methodology of previous studies, and thus have been able to investigate whether any major changes have taken place during the last decades.

The other major limitation due to time is that while our study concerned CMF chemotherapy, other treatments have become standard today. The current recommendation in Denmark is three series of CE followed by three series of the taxane docetaxel ([www.dbcg.dk](http://www.dbcg.dk), accessed June 2007). It is the impression from the literature reviewed in this section that there appears to be relatively modest differences between the ‘traditional’ regimens such as CMF, CAF, CEF, AC, etc., although there are few formal comparisons. In contrast, there may be significant differences when these regimens – which have been the standard for approximately 25 years – are compared to new chemotherapeutic standards including taxanes [248] or other new drugs.

These limitations should be taken into account when interpreting the results and it is recommendable that the present study is followed up by new studies of current treatments.

It should also be remembered that the discussion has been restricted to studies based on self-report (except for cognitive function and weight gain). Hundreds of studies could have been

added if those reporting physician-rated toxicity were also included; this could have added substantially to documentation for some side effects. However, the focus here is on patients' own assessments.

Our review, as presented above, was subsequently checked against the systematic literature review carried out by Mandelblatt et al. [249]; no missing studies were identified.

The use of results from HRQL studies is further discussed in the Perspectives section.

#### 4.9.10 Conclusions

- Based on the initial literature review and interviews (paper I) we hypothesised that 30 different HRQL issues would be impaired in patients undergoing CMF chemotherapy compared to patients not in chemotherapy; 23 of these hypotheses were confirmed (paper VII). Our study and other research suggest that additional HRQL aspects may be affected by chemotherapy. Thus, there is considerable evidence that patients in chemotherapy may experience effects on a wide spectrum of HRQL issues.
- Concerning comprehensiveness, our study is clearly the most complete; most other studies have assessed surprisingly few of the HRQL issues shown in our study to be impaired in patients receiving chemotherapy.
- Current review articles on HRQL effects of adjuvant chemotherapy mention only relatively few of these topics.
- These discrepancies may seem surprising given that our questionnaire was based on the literature available in 1990. Looking back one can say that the nature of quality of life effects of chemotherapy was almost fully clarified relatively early (or, at least few new effects have been identified), but this 'clarification' was not generally acknowledged.
- Concerning HRQL after the treatment period, our main finding was that many symptoms and problems had declined or disappeared three months after chemotherapy, but some persisted: anticipatory nausea, weight gain, endocrine effects (e.g., hot flushes/sweats, irregular bleedings/amenorrhoea, vaginal dryness), disturbed sleep, and sexual dysfunction. These findings are in agreement with the literature. Many patients seem to experience cognitive problems but despite intensive research it is still not clear whether or to what extent chemotherapy leads to lasting and objectively measurable cognitive dysfunction.
- The staff study showed that experienced physicians and nurses did not expect many of the 'scientifically well documented' consequences of chemotherapy.
- These findings indicate that knowledge is not just knowledge: there may be large differences between 'what is known in the literature' and what experts know.
- It is important for patients to be informed about consequences of a treatment before and during treatment. Both written material and oral information is important. The current study suggests that the information given to patients about chemotherapy should be more comprehensive than in current review articles and – probably – than that which has been practised in most places.
- Adjuvant therapy for breast cancer is undergoing quick and profound changes and new studies of HRQL consequences are needed in order to obtain up-to-date knowledge for future patients.

### 4.10 Chemotherapy or ovarian irradiation: impact on HRQL (paper VIII)

#### 4.10.1 Participation

The flow of patients in the DBCG 89 B protocol in Denmark was illustrated in Fig. 1 of paper VIII. The majority of patients (540 of the 762 in the protocol) were included from Denmark with the remaining patients recruited from Sweden and the Netherlands. During the inclusion period, 340 patients from Denmark were randomised in the protocol; 23 of whom were not reported from the DBCG Secretariat to the quality of life office due to administrative errors. The remaining 317 patients were invited to take part in the study. We chose to exclude the 14 patients who decided not to have the treatment they were randomised to from the analyses. Had we used the intention to treat principle for the analyses, our results for e.g., ovarian ablation would have been confounded with results from patients undergoing chemotherapy. After two years, 260 of the 317 patients were alive and recurrence-free, of whom 196 (75%) had filled in all six questionnaires, of whom 87 were in the chemotherapy group and 109 in the ovarian ablation group. We do not have an explanation of why the randomisation during this period was somewhat skewed towards ovarian ablation, but in the entire trial almost the same numbers of patients were randomised to the two arms (386 and 376, respectively) [250]. Of the 109 women in the ovarian ablation group 107 had ovarian radiation and 2 underwent surgical oophorectomy.

As described in paper VIII analyses at several levels showed no evidence of bias in the patients included in the final analysis compared to the Danish patients included in the trial.

#### 4.10.2 Results

The comparison of patients randomised to chemotherapy and ovarian ablation, respectively, is shown in Figs. 2 and 3 in paper VIII. The overall result was that patients in chemotherapy have higher levels of symptomatology in the treatment period (the first three assessments). The opposite was seen for a few variables related to ovarian ablation. There were few differences after the treatment period.

Thus, patients in chemotherapy reported more impairment of cognitive<sup>5</sup> and social function, more fatigue, nausea and vomiting, dyspnoea, sleep disturbances, loss of appetite, constipation, depression, weight gain, hair loss, anticipatory nausea, sore mouth, and urinary incontinence, less satisfaction with appearance, and poorer global quality of life. Directly after radiotherapy the patients in the ovarian ablation group had more diarrhoea. Patients in the ovarian ablation group had more hot flushes/sweats and became amenorrhoeic more quickly and more completely. However, as 77% of the patients in the chemotherapy group had stopped menstruating at two years, and only 9% had regular bleedings by two years, the difference was relatively small. The age-stratified analyses (paper VIII) showed a pronounced age-effect with only a quarter of the patients below 40 years becoming postmenopausal after chemotherapy compared to 57% in the age-group 40-44 years and 90% in those aged 45-49 years at diagnosis.

<sup>5</sup> As for the comparison of patients in chemotherapy to patients not in chemotherapy, the cognitive function scale was the only scale showing DIF: the difference observed for concentration difficulties was much larger than for memory problems (section 4.4.3 and Appendix A).

The additional analysis conducted to elucidate whether the exclusion of patients who completed less than all six questionnaire affected the results showed that this was not the case (data not shown). All conclusions would have been the same if these patients were not excluded. In contrast, the plots were slightly changed if patients with a recurrence were not excluded (data not shown): the main difference was that the levels of several symptoms increased from 15 to 24 months as a reflection of disease-related symptoms or treatment. However, all conclusions drawn from the comparison of the two groups would have still been the same.

Table 6 summarises the results of the comparison of patients in chemotherapy and controls (paper VII) as well as the randomised trial comparing chemotherapy to ovarian ablation (paper VIII). The mean ages of the groups of patients are similar (between 44 and 45 years).

The outcome of the comparison is shown with a single symbol for each variable. Table 6 thus shows whether there is a difference and in which direction – not the magnitude or duration. A comparison of the graphs (Figs. 1 and 2 in paper VII vs. Figs. 2 and 3 in paper VIII) shows that, as one would expect, the differences were generally larger when chemotherapy was compared to controls than to ovarian ablation.

With these reservations Table 6 shows that most of the differences were the same (though not necessarily of the same magnitude or duration) when chemotherapy is compared to controls and to ovarian ablation. However, no differences were seen for role function, energy, vaginal dryness, and interest in sex when comparing chemotherapy to ovarian ablation. In other words, these dimensions must have been affected by ovarian ablation as well. Three differences were reversed, i.e. higher scores were seen in the ovarian ablation group for hot flushes/sweats, irregular bleedings, and amenorrhea. Finally, the patients undergoing ovarian ablation had higher levels of diarrhoea at the one-month assessment whereas at five months the difference was opposite. From a medical/clinical point of view, the two main differences between chemotherapy and ovarian ablation are the duration of the treatment (six months' chemotherapy compared to daily radiation for a week) and their content (cytotoxic or endocrine treatment). Table 6 can be seen as an analytical attempt to separate the cytotoxic and endocrine effects of chemotherapy: the differences seen both when chemotherapy is compared to controls and to ovarian ablation are those that can be attributed mainly to cytotoxic, not endocrine effects. The differences that are reversed (hot flushes/sweats, irregular bleedings, and amenorrhea) are those where the endocrine effect of ovarian ablation is stronger than that of chemotherapy.

Thus, the overall picture is that while the HRQL impact of chemotherapy is clearly stronger and more diverse than that of ovarian ablation, it is also evident that ovarian ablation has considerable impact on important aspects of HRQL due to its endocrine effects and due to their consequences for sexuality.

#### 4.10.3 Discussion

In this trial no difference was found in the efficiency of the two treatments [250]. This is consistent with the other randomised trials [199, 251-255]. In contrast, major differences in HRQL were detected. Assuming that the treatments are truly equally efficient, ovarian ablation is thus clearly preferable to CMF chemotherapy if only 'symptomatology' is considered: many symptoms were more pronounced during the much longer cytotoxic treatment, and only hot flushes/sweats were worse with ovarian

ablation. However, the possible preservation of premenopausal status and fertility may outweigh the problems associated with chemotherapy for some women. Age is an important factor in this consideration, as the probability of becoming postmenopausal following chemotherapy is strongly age-related. Our results followed a graphical model for the relationship between age and amenorrhea following chemotherapy [153] closely, and showed that while 'only' a quarter of the women below 40 years of age at diagnosis became postmenopausal following chemotherapy, the probability of amenorrhea was 90% in women aged 45-49 years. Thus, there is relatively little probability of staying premenopausal for a relatively large part of the women (in our study 58% of the premenopausal breast cancer patients were at least 45 years old). In addition to the results of the present trial there are two additional factors that must be taken into consideration in treatment decisions concerning premenopausal, receptor-positive women. First, as discussed in the section on chemotherapy above, CMF chemotherapy is no longer the standard chemotherapy. The Early Breast Cancer Trialists' Collaborative Group's 1998 review [256] suggested that anthracycline-containing regimens are more effective than CMF chemotherapy, and such regimens are now usually the standard, sometimes in combination with taxanes [257] and/or trastuzumab [6]. This does not affect the conclusions concerning better HRQL with ovarian ablation (as the other chemotherapeutic regimens are probably similar to or worse in HRQL impact) but when other regimens are considered more effective than CMF, they are probably also more effective than ovarian ablation [258]. Although the differences may be small for the individual patient, the choice of ovarian ablation may thus be a trade-off of probability of cure against quality of life.

The second factor in the decision-making is that instead of permanent ovarian ablation (as in this study) temporary ovarian ablation can be obtained through regular (typically monthly) injections with goserelin for 2-3 years. No direct comparisons of either effectiveness or HRQL impact of permanent and temporary ovarian ablation have been made but they are usually considered of similar effectiveness [259] and probably have a roughly similar HRQL profile in the two- or three-year treatment period. After this, the balance must be expected to shift towards an advantage for temporary ablation (as discussed below).

Before proceeding with the discussion of chemotherapy or ovarian ablation it should be briefly noted that there is also a discussion of whether the treatments should be combined. Specifically, it has been argued that for women who do not become amenorrhic after chemotherapy it may be advantageous to add ovarian ablation [258, 260-263].

While there are no other published studies comparing HRQL between chemotherapy and permanent ovarian ablation, the results can be compared against those of three studies comparing CMF chemotherapy against temporary ovarian ablation by means of goserelin for two years [264].

de Haes et al.'s international, randomised trial followed patients for three years [264]. Consistent with our study the goserelin trial found generally better HRQL during the first six months, where the chemotherapy group received treatment, but higher levels of 'hormonal symptoms' (hot flushes was assessed along with other symptoms as a total scale score) were found in the goserelin group during the two years of ovarian suppression. However, at three years the latter difference was reversed and the level of hormonal symptoms was higher in the chemotherapy group (probably because the menopause induced by chemotherapy is irreversible) than in the goserelin group, where ovarian suppression had been stopped one year earlier. Whereas we found no

differences in sexual interest, the scores for sexual interest were lower in the goserelin group during treatment, but higher at three years [264], thus suggesting a benefit from cessation of ovarian suppression.

Bernhard et al. recently reported a similar trial with similar results [262]. It showed less nausea/vomiting, better coping, and better overall health in the goserelin group but initially (at three months) these patients had more pronounced hot flushes and lower mood. At three years, many more patients in the chemotherapy group had hot flushes and amenorrhea, thus confirming the reversibility of the ovarian suppression, particularly among younger women [262].

Finally, HRQL was assessed in the Swedish participants of an international study comparing various endocrine regimens including goserelin [139, 167]. Chemotherapy was given to the node-positive patients. The study thus allowed an indirect (non-randomised) comparison of patients in chemotherapy with patients treated with goserelin. Sexual function was impaired to the same extent during the first two years, whereas, as in de Haes' study, the assessment one year after goserelin cessation indicated that the effect of goserelin, but not chemotherapy, on sexuality was reversible [139]. Vasomotor effects tended to be more pronounced during goserelin treatment but declined after goserelin was stopped, whereas they persisted in the chemotherapy group [167].

Thus, all three goserelin studies found that endocrine symptoms were reversible following completion of goserelin treatment. This suggests a likely advantage of temporary ovarian suppression over permanent ovarian ablation.

A theoretical study based on quality-adjusted life-years investigated the trade-offs between treatment efficacy and HRQL when choosing between chemotherapy, surgical ovarian ablation, and medical ovarian suppression [265]. The study tested how different combinations of the relative efficacy of treatments and the relative utility of side effects would affect treatment decisions. It was found that even small differences in treatment efficacy would shift the balance towards the most effect treatment irrespective of how patients evaluated the side effects of the treatments. The magnitude of differences necessary to shift the balance were smaller than those detectable in the available trials [265]. Assuming equal efficacy the cut-point for treatment decisions would be a relative utility of side effects of 0.95. Perfect health has the utility of 1.0. The relative utility value of 0.95 corresponds to a utility of chemotherapy side effects of 0.86 (during a six-month period) and a utility of ovarian ablation side effects of 0.90 (assumed to last for two years). Thus, if a potential patient would say that the utility during chemotherapy was lower than 0.86 (and that of ovarian ablation was still 0.90) the model would then favour ovarian ablation. In contrast, if the difference in utility between treatments were viewed as smaller, the model would then favour chemotherapy [265]. Even though relative utility values may be difficult (if at all feasible) to use in clinical practice, the study nicely illustrates the interplay between effectiveness and HRQL of treatments.

A British study investigated the preferences of 200 healthy, premenopausal women for goserelin and CMF chemotherapy, respectively [266]. The participants were asked to imagine they had breast cancer and were given detailed descriptions of the treatments and their side effects. They were informed that the treatment had the same efficacy. Most women (78%) preferred goserelin, 11% preferred chemotherapy, and 11% remained undecided [266]. This finding is consistent with the decision-analytic study mentioned in the previous section provided that

the difference in utility between treatments is seen as significant (e.g., more than 5%) [265].

These results suggest that there are two main factors to consider when informing future patients. First, whether more than one treatment option should be presented. Second, if more than one option is considered then the alternatives must be determined. It is meaningful to present treatment alternatives only if there are differences that may be of relevance to the patients. The HRQL differences between chemotherapy and ovarian ablation represent such differences [258, 266, 267] and there may be other alternatives. Therefore, there are arguments for presenting ovarian ablation/suppression as an alternative to chemotherapy. The evidence presented here thus indicates that adjuvant therapy is one of the still relatively rare cases [268], where findings from HRQL research could potentially affect treatment decisions. However, it is controversial whether this will take place. The question is: will physicians and/or guidelines present a dilemma (i.e., a detailed description of the advantages and disadvantages of chemotherapy vs. ovarian ablation/suppression) to patients, or will treatment recommendations continue to be unidimensional ('the best treatment is the most effective treatment as measured by probability of survival')?

While it is easy to argue that patients should have a choice, it is certainly not simple to go from unidimensional treatment recommendations ('the most effective treatment we can offer you is...') to multidimensional recommendations [267]. Obviously, physicians can present only parts of the available scientific evidence concerning the treatment of a condition – otherwise one could imagine absurd situations where endless explanations and details would only confuse the patient. On the other hand, if different valuations of the scientific evidence may lead to different treatment decisions, it is more problematic to deny the patient insight into complexity [267]. For example, for a 47-year-old premenopausal women with receptor-positive breast cancer, is the added toxicity associated with 'optimal chemotherapy' compared to ovarian ablation justified by the increased probability of survival? If explained all details most patients will probably think so but some might not. In other words, the consequence of presenting one treatment option only is that some patients will receive a treatment they would not have chosen had they been given a choice. Is it ethically acceptable to deny patients insight into a dilemma like this? And conversely, given that most patients will probably choose chemotherapy if it is presented as more effective, is it ethically acceptable to expose the patients to all the statistical information and medical complexity needed to make a truly informed decision?

Obviously, the answers to these questions involve more than just the scientific knowledge about treatments – they also depend on how the relationship between physicians and patients is viewed. Patients' expectations to the health care system is changing rapidly in these years, and more research into the patients' views and expectations as well as public discussion of the medical decision-making process is needed.

#### 4.10.4 Conclusions

- Chemotherapy was associated with more impact on HRQL during the six-month treatment period; only hot flushes/sweats were more pronounced in the ovarian ablation group. Thus, from an overall, 'HRQL perspective' ovarian ablation or suppression may be the preferable treatment for many patients. This conclusion is in agreement with a study eliciting the views of healthy women.

Table 6

Overview of results from the comparison of patients in chemotherapy and controls (paper VII, upper row) and from the randomised comparison of chemotherapy and ovarian ablation (paper VIII, lower row).

- More symptoms/problems in chemotherapy group.
- Δ More symptoms/problems in ovarian ablation group.

	EORTC QLQ-C30														HA D	DBCQ 89 Questionnaire																				
	Physical function	Role function	Emotional function	Cognitive function	Social function	Global health/QL	Fatigue	Nausea/vomiting	Pain	Dyspnoea	Trouble sleeping	Loss of appetite	Constipation	Diarrhoea	Financial difficulties	Anxiety	Depression	Energy	Weight gain	Hair loss	Wig	Anticipatory nausea	Sore mouth	Hot flushes/sweats	Urinary incontinence	Vaginal dryness	Vaginal flux	Amenorrhoea	Irregular bleedings	Frequency of sex	Interest in sex	Satisfaction with app.	Employment	Full-time job		
Chemotherapy vs. control	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Chemotherapy vs. ovarian ablation		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	Δ	●	●	Δ	Δ						●		

- Among younger, premenopausal women, who may preserve their premenopausal status (including fertility) by having chemotherapy, this concern may be an argument for chemotherapy or for temporary ovarian ablation via goserelin, rather than permanent ovarian ablation.
- No studies have directly compared the impact of permanent and temporary ovarian ablation. The studies of goserelin suggest that vasomotor symptoms and sexual function improve after cessation of goserelin therapy. This is not the case in patients who have undergone chemotherapy. Therefore, the balance between the HRQL impact of chemotherapy and ovarian ablation may be even more favourable for temporary than for permanent ovarian ablation.
- In addition to HRQL-related concerns, treatment decisions clearly involve judgements of the relative efficiency of treatments. While the available studies suggest that ovarian ablation and CMF chemotherapy are equally effective, other studies have provided evidence that alternative chemotherapeutic regimens are more effective than CMF.
- These results suggest that while ovarian ablation/suppression may be preferable for many women because of less impairment of HRQL, contemporary chemotherapeutic regimens may be more effective. The simple solution to this situation is to say that efficiency is always more important than HRQL, and that patients should therefore have one treatment option only. A more difficult solution is to determine that there is a dilemma and to involve patients in the decision.
- More research into patients' views and expectations to the health care system in cases where medical decision-making involves complex trade-offs between efficacy and HRQL issues is needed.

**4.11 Prognostic significance of quality of life data (paper IX)**

**4.11.1 Patients**

The analyses were based on 1,588 patients who had complete data for all the biological variables and who had completed the

first questionnaire. The analysis of RFS and OS included 761 and 698 'events', respectively, i.e., about half the patients had a recurrence or a second malignancy, or died.

**4.11.2 Results**

The final biological model is shown in Table 2 of paper IX. When adding one variable at a time to the biological model the EORTC emotional function and fatigue scales predicted both RFS and OS. The HADS anxiety subscale categorised 8-21 (probable or definite case) vs. 0-7 (non-case) predicted RFS but not OS. In all cases, low psychological distress or low fatigue were associated with improved probability of survival. When categorised at the median both HADS subscale were just above the level of significance in prediction of RFS. The other variables did not significantly predict RFS or OS (Table 3, paper IX).

When added to the biological model in combination fatigue was the only variable remaining as a significant predictor of RFS, and emotional function was the only variable predicting OS (Table 3, paper IX). Addition of social class to the final models did not change the results.

When the same analyses were carried out in the subset of 432 low-risk patients, the results were the same except that fatigue and emotional function had slightly stronger associations with RFS and OS. Reflecting the smaller sample, the p-values were 'less significant'.

**4.11.3 Discussion**

Contrary to our expectations prior to the study we found that psychological distress and fatigue predicted the risk of recurrence and death even when controlling for the relevant clinical and biological variables such as tumour size, malignancy grade, lymph nodes, etc.

When interpreting the results it should be noted that the mean age of the patients at diagnosis was 52 years, i.e., many patients were in an age group, which apart from breast cancer would be expected to have a low mortality. Almost half the patients died during the almost 13 years of follow-up. The patterns of relationships between the different self-rated variables from two stan-

standard questionnaires were all in the same direction (fewer problems, better survival), and the relationships with recurrence-free survival and overall survival were relatively similar.

One might speculate that although, in addition to biological variables, all treatment variables were included in the analyses there might still be some complex mechanism of confounding resulting from the fact that treatment was associated with risk of recurrence, and treatment was also (weakly) associated with psychological distress. Therefore, it is notable that when analyses were repeated in the subgroup of 432 low-risk patients exactly the same results (in fact slightly stronger associations) were found. This mechanism of confounding is therefore an unlikely explanation of the findings.

Many studies have reported associations between HRQL and survival in metastatic cancers other than breast cancer [269-278] and metastatic breast cancer [279-283]. However, the finding that in the presence of symptomatic disease, patients' perception of their own health is related to their survival, is not surprising. The relationship between self-rated health and survival is much more scientifically interesting in disease-free patients who have no symptoms of cancer (e.g., in primary breast cancer), where (optimally) all prognostic variables that can be known to patients can be accounted for in the analyses. The only other study of primary (breast) cancer patients that has found patients' self-ratings of psychological distress to be associated with survival was Watson's study, which aimed at investigating whether the patients' coping style predicted survival. A group of 10 patients having HADS scores above 10 was found to have shorter survival than the remaining patients [284]. This finding was not confirmed when the analyses were repeated after longer follow-up [285]. It is remarkable that our findings are in contrast to five other studies carried out in primary breast cancer patients [279, 286-289]. The Discussion section in paper IX includes a relatively detailed comparison of our study against each of the five other studies. Despite many similarities there are also important differences in relation to the five studies. Several weaknesses in the other studies, compared to our study, were identified. Such differences may have contributed to the fact that the other studies found no predictive effect but still it appears strange that the relatively strong and consistent findings from our study were not seen in five other studies.

Could our results be due to some kind of error or bias, e.g., confounding? As mentioned above, it was reassuring to see that the findings were the same in the subgroup of low-risk patients as in the entire study. Could the patients have had some insight in their prognosis that was not accounted for in the analyses? Such knowledge might have affected their level of psychological distress and fatigue. All the information about the prognosis of the disease that the patients may have had (e.g., tumour size, number of lymph nodes, malignancy of the tumour, receptor status) as well as the treatment was included in the biological models. Therefore, it is unlikely that the predictive ability of emotional function, anxiety, and fatigue is mediated via the patients' knowledge of their prognosis.

More research is needed to clarify whether our findings or those of other studies best describe the reality.

In paper IX we propose that two different theoretical models can be used to interpret our results (paper IX, Fig. 2). The first is the traditional 'mind-body model' suggesting a causal effect of psychological distress on the disease trajectory. The second model, which to my knowledge is new, is called the 'robustness model'. According to this model, the causal relations are different. The concept 'robustness' is proposed as a common explanation of lack

of psychological distress and fatigue (despite the exposure to significant stressors) and increased resistance to breast cancer. Clearly, the interpretation of results is markedly different depending on the choice of models. The first provides support to the psychoneuroimmunological line of thought (e.g., [290]). Following this model, the disease course might be modified if psychological distress or fatigue were reduced.

If results are interpreted according to the robustness model, interventions that reduce psychological distress or fatigue will not affect the cancer but – of course – will still be beneficial. Irrespective of the choice of model, activities that help patients tackle their situation are therefore highly recommendable, but only if one believes in the first model should such activities be motivated with an effect on survival.

While not fully clarified [291], the psychoneuroimmunological theories related to breast cancer have lost some terrain the last years. For example, there was great enthusiasm following Spiegel's study suggesting that psychosocial intervention could prolong survival in advanced breast cancer [17] but subsequent studies could not replicate the findings [292-294].

#### 4.11.4 Conclusions

- The levels of fatigue and emotional function about two months after diagnosis predicted the risk of recurrence and death for the next many years independently of biological and clinical prognostic variables.
- Our results are consistent with a small study of primary breast cancer patients, which found an association between depression and survival, but are inconsistent with five other studies that found no associations between HRQL and survival in primary breast cancer.
- It is currently unknown whether the disparate results may be explained by differences in design or analysis between our study and the other studies.
- If our results are correct, they can then be interpreted according to two different causal models described in this chapter, the traditional 'mind-body model' or a 'robustness model' proposed in paper IX.

## 5 DISCUSSION OF MATERIALS AND METHODS

This thesis investigated the HRQL of primary breast cancer patients. A number of methodological sub-studies were incorporated to try to achieve the best possible scientific basis for the evaluation of HRQL: the questionnaire was composed after a literature review as well as a small interview study, and was pilot tested before use (paper I), the multi-item scales were evaluated for differential item functioning (DIF) (paper II), the validity of patients' self-assessment was evaluated through a new method developed for the purpose (paper III), and to consolidate the basis for hypothesis testing a framework for incorporation of staff-expectations in the analyses was investigated (paper IV). These methodological sub-studies have been discussed in the previous chapters and will not be discussed here. This chapter will take a look at the strengths and weaknesses of the materials and methods used to evaluate the impact of early breast cancer and adjuvant therapy on HRQL, and to assess whether psychological distress has prognostic significance.

### 5.1 A longitudinal, not cross-sectional design

It would have been much simpler and less resource-demanding to carry out a cross-sectional study, e.g. based on a single assessment of a random sample of patients 0-2 years after diagnosis.

Such a study could have allowed a detection of many of the common problems experienced by the patients and could have identified major differences between groups. For example, many of the symptoms and problems associated with chemotherapy would probably have been correctly identified if assessments during treatment had been obtained. Furthermore, due to the simplicity of the design, data could have been collected more quickly and more cheaply and with less effort from patients. Nevertheless, it would have taken considerable time to recruit the sufficient numbers of patients – the current study included a large proportion of the eligible patients in Denmark. Analytically such a design would have been more complicated and multivariate regression analysis or other techniques would have been needed. Thus, as the analysis would have been more demanding, some of the savings (time, resources) from the reduced data collection would have been lost. However, one could correctly argue that some of the findings of the present study could have been found through a cheaper and faster study design. The most important disadvantages of the cross-sectional design are the reduced ability to describe longitudinal patterns, a reduced power to detect differences between groups, and an increased vulnerability to bias (furthermore, as described later in this chapter, a cross-sectional design would not be suitable to utilise the advantages of the randomised design). The analysis presented in papers VII and VIII showed pronounced changes in some variables over time, and a cross-sectional design would have less power to detect such patterns. And even if patterns were reasonably well captured in the data it is much more difficult to communicate the results from multivariate models (one for each of the more than 30 variables) than to show simple graphs of mean scores over time.

Furthermore, with a given sample size, a cross-sectional design would have less ability to detect differences between groups due to the increased noise resulting from differences in the time of assessment. The graphs in papers VII and VIII show that the patterns are different for different variables and are usually non-linear. Such patterns would be extremely difficult to capture adequately in multivariate models.

As an example, a large, recent, cross-sectional study of 2,236 Chinese breast cancer patients found ‘only a marginal association of current use of chemotherapy with poorer QOL in the physical wellbeing domain, suggesting that while these symptoms may be bothersome, they are transient and may not be substantial enough to affect the major dimensions of HRQL in our population.’ [161]. The difference in the ability to describe the impact of chemotherapy of this cross-sectional study compared to our longitudinal study is large.

Finally, a cross-sectional design is subject to an increased risk of bias resulting from the reduced ability to separate the effects of the individual variables, even when multivariate models are used. It is therefore clear that a longitudinal design is much better suited to describe patterns of HRQL over time and to detect differences between groups.

### **5.2 Patient-assessed HRQL rather than physician-assessed toxicity**

Numerous studies have shown that there is poor to moderate agreement between patients’ own assessments of their HRQL and assessments done by ‘proxies’ such as health care professionals or family members [79, 80, 89, 295, 296]. In general, patients’ own assessments must be viewed as more valid [22, 79]. The difference may be even larger when patients’ assessments in

HRQL questionnaires are compared against physician-rated ‘toxicity’: the topics covered are only partially overlapping. Toxicity ratings such as WHO Common Toxicity Criteria are focused on specific, mainly physical symptoms, whereas HRQL instruments also include other aspects, e.g., psychosocial aspects. Toxicity ratings have a clear and well-established role in clinical trials but do not replace HRQL assessments.

### **5.3 Questionnaires to patients rather than interviews**

As the present study had the aim to quantify and compare the prevalence of a wide range of HRQL aspects (symptoms, problems, etc) between groups and over time, a quantitative, standardised methodology was needed. This also allowed direct comparisons with published studies.

It is important to acknowledge that the standardised, quantitative methodology used here does not give the possible insights one could have obtained from interviews. New knowledge about, for example, how patients think about, perceive, and react to treatment and disease could be obtained from interviews, whereas such information is almost ignored from a study like this. The two phases where a qualitative methodology was applied in this study, the initial interviews and in the analysis of data from the validation study (paper III), brought forward useful new information.

### **5.4 Patient participation**

The participation of 90.3% (first assessment) of the patients in the clinical study (and thus the basis for papers II, VI, VII, VIII, and IX) (Table 3, section 4.3.2) was extremely high for a study of this kind. The attrition in the longitudinal analyses reported in papers VII and VIII was modest, and as described in the papers it did not seem to affect the results. Levels of participation close to 100% have been achieved in randomised trials where participation in the HRQL was an inclusion criterion, but are rare in studies where participation is voluntary. Thus, compared to other studies it is a strength that the participation in our study was very high. The fact that patients took the time to complete a relatively extensive questionnaire at a point in time where they had many other things to do probably reflects that they found the study relevant. This interest in the study may not only have reduced the risk of bias due to non-participation; it may also have contributed to a high level of validity of results because patients took the task seriously. This assumption is coherent with the impression I got from large numbers of comments written in the questionnaires and from many telephone calls from patients during the data collection: the patients generally saw the study as very important and often made additional comments aimed at elaborating their responses.

### **5.5 Comparisons within randomised trials and between non-randomised groups**

The advantages of the randomised trial – compared to non-randomised designs – are well known.

While the internal validity of randomised trials is usually higher than non-randomised comparisons, the external validity may be limited if the experimental design leads to selection of a subgroup of patients that is not representative of the population of interest. In the current study it was clearly a strength that the comparison of chemotherapy and ovarian ablation took place in a randomised trial because this reduced the risk of confounding. The disadvantage was that patients who were strongly against one of the two treatments (e.g., a young woman wanting to

preserve her fertility), probably refused randomisation, and our results may therefore not be generalised to such patients. The comparisons in papers VI and VII were not randomised, as this was for obvious reasons not possible. With respect to internal validity, these studies are clearly weaker than the randomised trial in paper VIII. This is evidenced in the unclear results of paper VI: the study did not clarify to what extent a recent breast cancer diagnosis leads to anxiety and depression (however, as discussed previously, multiple methodological issues related to the comparison of 'patients' to persons from a general population sample were identified). The weakness is also seen in paper VII, where it was not possible to distinguish the effect of chemotherapy from that of the difference in prognosis between groups.

The control group in paper VII was probably highly representative of low-risk patients, whereas the patients in chemotherapy were those included into two randomised trials. Thus, one may argue that the representativity of the patients in the chemotherapy group is less optimal than that of the control group. There may be selection bias in randomised trials because patients who accept randomisation may differ from those refusing randomisation (e.g., in the level of trust in the health care system). However, our main interest was to elucidate HRQL differences, and most of these dimensions are probably not substantially affected by such selection: it seems unlikely that the magnitude or course of the various symptoms is markedly different in patients accepting randomisation compared to those not accepting randomisation. However, we cannot know this.

For these reasons, when designing the study we discussed carefully whether to include those patients refusing randomisation. The main argument in favour of this was that it would allow us to investigate the entire population of patients. In addition, we could have found out whether there were differences in the HRQL associated with different treatments between those randomised and those not randomised. We chose not to include patients refusing randomisation for mainly two reasons. First, there was a possible ethical problem in approaching patients who had just refused participation in a scientific study and once again ask them to participate in a different but closely related study. Second, we considered it more important to the aims of the study to use the available resources to get as large groups as possible within the randomised trials.

However, again, randomised trials are feasible under certain circumstances only. The trial reported in paper VIII might be the only randomised trial ever conducted comparing chemotherapy to permanent ovarian ablation, and therefore it was valuable that the opportunity to include an HRQL study was utilised.

### **5.6 Timing of assessments**

As discussed above, the longitudinal design with six measurements over two years is superior compared to a cross-sectional design. Clearly, one could have included more assessments or have selected other points in time but each additional assessment costs time for participants, is expensive for the research budget, and may increase drop-out. The six points of assessment seem to cover the period of acute toxicity and a subsequent 'normalisation' period leading to absence of differences between groups as well. Thus, additional assessments seem warranted mainly if one is interested in short-term fluctuations as in a recent study of fatigue [297]. On the other hand, the graphs of papers VII and VIII show that omission of one or more of the assessments would have led to loss of information. Further follow-up of the study population beyond two years might lead to additional findings,

e.g., on the duration of persisting symptoms, but the most important results seem to be those obtained during the first two years. One can argue that a major weakness of the timing of assessments in the present study was that it did not include a 'baseline' questionnaire completed before randomisation and initiation of adjuvant therapy. Articles and textbooks on the methodology of HRQL research routinely recommend baseline assessments [25, 298]. A 'baseline' assessment before randomisation can be used to investigate whether there are differences in HRQL before treatment. Such differences can be accounted for in the analysis. Furthermore, a baseline measurement would give additional possibilities in the choice of analytic strategies because 'change scores' rather than absolute scores could be used as outcomes [25](p. 236). It may, however, be difficult to ensure completion of HRQL forms before randomisation, and 'baseline' assessments after randomisation are less useful because patients may be affected by the outcome of randomisation [25, 156].

There are, however, some problems associated even with 'baseline scores' carried out before randomisation. First, of course, while a 'pre-randomisation assessment' can in principle be obtained for patients entering a randomised trial, a comparable assessment in patients not randomised may not be obtainable: patients awaiting information about their adjuvant therapy are in a stressful situation, and this will affect their HRQL scores. In the current study it would have been difficult to interpret a comparison of 'pre-randomisation' scores of patients randomised to chemotherapy compared to scores from the control group, who, obviously, were not randomised. In contrast, within the analysis of randomised trials a pre-randomisation may be useful to test for possible differences between randomised groups.

Pre-randomisation assessments are clearly not 'pre-disease' assessments: the patient is aware that she is ill, is awaiting a potentially stressful treatment, and thus is certainly not in anything similar to her normal state. In fact, the post-operative period until initiation of adjuvant therapy, where the patient is still not fully informed about her disease, treatment, and prognosis, is extremely stressful to most patients. Therefore, a 'baseline' assessment carried out at this point in time is a measure of the fluctuating problems and distress the patient is experiencing. It was decided not to include a pre-randomisation assessment in the current study mainly for two reasons. First, it was not considered practically possible to arrange a pre-randomisation assessment in all potential patients in a way that was felt to be appropriate towards the participating patients, and would result in reasonably complete data. Second, it was not considered vital for the validity of the study to have such an assessment. The study had its focus on the period during and after initiation of adjuvant therapy, not on the period preceding it. The most important, planned comparisons were to take place within the randomised trials, and the risk of imbalanced randomisation was considered small.

Thus, if feasible, pre-randomisation baseline assessments may be useful in randomised trials but would probably not have added very much to the validity of the present study.

Another aspect of the timing of assessments concerns their relationship to the fluctuations caused by particularly chemotherapy. It is well known that side effects of chemotherapy, e.g. nausea/vomiting and fatigue, have cyclic patterns. Nausea and vomiting is typically most pronounced on the day of infusions and possibly the following days (with different drugs having different temporary patterns). Fatigue is typically a problem for a longer period but also tends to improve with time from last infusion. Thus, both nausea and vomiting, and fatigue tend to be minimal

when the patients come to the hospital for treatment, whereas anxiety may be higher at this point in time than for example one week earlier.

Although investigated in a few studies [297], such temporary patterns have to a large extent been ignored in HRQL research. The reason for this is probably mainly practical: if the selected mode of questionnaire administration is to give patients the questionnaire in the hospital (and patients come every three to four weeks for treatment) then the typical one-week time frame employed in the questionnaire elucidates the week before treatment, not the week after treatment (some questionnaires have a longer time frame but also a time frame of, e.g., four weeks assesses the week after treatment poorly). It is impractical to arrange an extra visit to the hospital or to allow a research assistant travel to the patient's house. And the compliance may be higher when patients are asked to complete the questionnaire at once than if the patient is to take the questionnaire back home to complete. Questionnaires are therefore often handed out when the patients come for treatment and are completed at that point in time. The logical consequence is that treatment-related problems are under-estimated.

It is a strength of our study that it employed a post-based administration system aimed at obtaining questionnaire completion one week after chemotherapy (and at the corresponding point in time for patients not in chemotherapy). We have not investigated the extent to which this actually took place (and this can be criticized), but although some patients may have delayed the completion of the questionnaire, our system must have had a better ability to capture the treatment-related symptoms than procedures where questionnaires were handed out and completed at the hospital.

### **5.7 The location of questionnaire completion**

It follows from the discussion above that questionnaires were completed at home in our study. The main alternative is completion at the hospital. Each location has advantages and disadvantages. Advantages related to completion at home include the patient having sufficient time, can plan to complete the questionnaire when and how it is most suitable, and, not least, that the patient is in her 'normal state', not in the often stressful situation at hospitals when awaiting treatment or consultation with a doctor. The main disadvantage is that the staff is not there to give help or to supervise that the assessment takes place as intended. Many patients utilized the possibility of calling me by telephone while completing the questionnaires, and asked for advice as to how to do; typically the questions concerned relatively unimportant issues but the carefulness exhibited was impressive and encouraging. Thus, when completion at home is coupled with a 'hotline', advice can be given. In my view the advantages of completion at home outweigh its disadvantages in most cases. The main problem is that it is logistically demanding to arrange the posting of questionnaires and reminders in large, longitudinal studies.

### **5.8 This approach compared to other approaches to HRQL assessment**

Specific aspects of the research strategy used in this study are discussed in other parts of this chapter but it may also be of interest to take a look at the profile of this study compared to other studies in the HRQL research tradition. First, one can discuss whether there is a single 'HRQL research paradigm/tradition' or whether there are actually several competing para-

digms/traditions. Many researchers in the field may take the latter view arguing that there are markedly different approaches being applied. On the other hand, one can argue that there is a field of research using the terms 'quality of life research', HRQL research, etc., which has a number of general characteristics. Although there is no generally accepted, single definition of HRQL, an excellent review of definitions is given by Ferrans [24], and several textbooks describe and give recommendations for a wide range of conceptual and methodological issues [25, 299-307]. Furthermore, guidelines, which to some extent can be viewed as expressions of consensus, are being published regularly [306, 308, 309] and there exists a scientific society called International Society for Quality of Life Research holding yearly, well-attended meetings. Finally, the US Food and Drug Administration recently issued guidance for the pharmaceutical industry providing extremely specific and detailed recommendations on the methodology of HRQL research [310]. This latter document used the term 'patient-rated outcomes' (PRO) instead of HRQL, but while the PRO concept is more inclusive, most of the content and methodology is the same.

Thus, while one may argue that 'HRQL research' represents a reasonably well-established research paradigm, there are a number of 'internal' differences where the present study has 'chosen side'. Such decisions can of course be discussed. One important line can be drawn between multidimensional research describing several aspects of HRQL (as in this thesis), and unidimensional assessments. The latter aims at describing quality of life on a single axis, e.g., from 0 to 1, and this is a prerequisite for health economic analyses such as estimation of quality of life-adjusted life-years (QALYs). Such methodology allows for comparison of different interventions with regard to the costs per QALY gained. Although there have been researchers working to establish links between multidimensional and unidimensional measures [311], the general view is that results from these lines of research are incompatible. In this thesis, this incompatibility can exemplified with the comparison of chemotherapy and ovarian ablation: the results cannot be translated into figures describing the HRQL of the treatments on a 0-1 scale – in fact, as described in the discussion, the relative merits of the treatments depend on the patients' values, and thus the outcome of the comparison cannot be described in a single figure.

Thus, it must be acknowledged that results from the present work are not relevant for health economic analyses needing quality of life data on a single axis.

Another division within HRQL research is between methods measuring pre-selected dimensions only (as in this thesis) and those focused on 'individual quality of life'. In the latter, newer methodology, the dimensions to be measured vary between participants [312]. Thus, the first step in the assessment is to identify the dimensions to be investigated. This has the clear advantage that the individual persons' values and situation are taken into consideration but severely limit the possibilities of comparison across individuals and between groups. For this study, a standardised assessment of the same HRQL dimensions in all participants was considered mandatory to reach the goals but, obviously, addition of individualised information would have been useful. It is also important to be aware that there are aspects not usually covered by typical HRQL assessments that are viewed as extremely relevant by patients. A recent example of this came from the Danish Cancer Society project 'Kræftpatientens Verden' ('The Cancer Patient's World') [313]. The qualitative part of that study showed that, in general, patients were more interested in discussing problems and frustrations related to the encounter with the

health care system (e.g., 'service issues' such as problems related to waiting time, information, communication, lack of continuity, and psychosocial care) than they were in discussing symptoms and disease-related problems [314]. The main reason for the patients' focus on the problems related to the health care system was probably that such problems might potentially have been avoided, whereas patients saw symptoms and disease-related problems as unavoidable consequences. This example illustrates that there are important issues that are not covered by traditional HRQL assessments, which have profound influence on the patients' quality of life, and which certainly need investigation. In summary, there is not a single delineation of HRQL research that can be termed as most correct; HRQL research represents activities aimed at elucidating the patients' experiences and perspectives but the way such research is best carried out depends entirely on the aims. This thesis aims to provide answers to certain questions but other questions could have been asked; thus different answers about the quality of life of breast cancer patients would have been obtained.

### **5.9 The statistical analysis strategy**

Several different statistical analysis strategies could have been applied in the two descriptive and comparative, longitudinal studies (papers VII and VIII). Our study included many variables assessed six times over two years. Some (surprisingly few) patients dropped out during the study or provided incomplete questionnaires. This situation left us with several decisions concerning the level of aggregation at each point in time as well as across time. We chose to aggregate very little. As a result we were left with many variables (as shown in the figures of papers VII and VIII.), and we could compare the variables six times. The 'rules' we applied in paper VII limited this a little; nevertheless, one can argue that we carried out too many significance tests, thus increasing the risk of false positive findings.

Another important drawback associated with our lack of aggregation is the complexity of results (as evidenced in excessive numbers of graphs), which is clearly a limitation because it makes communication of our results time-consuming and complicated. However, we can argue that the reality is complex and that it is inappropriate to limit this complexity as long as it is not well understood. Furthermore, our studies have given insights that may be utilised to plan simpler studies in the future, e.g., studies with shorter questionnaires or fewer assessment points. We could thus have chosen to aggregate variables more than we did (e.g., by constructing multi-item scales from the DBCG 89 Questionnaire) but at the point in time when analyses were planned and conducted this would, in my view, not have been appropriate. Papers VII and VIII documented that almost all variables analysed contributed new information.

We could have combined information from several assessments, e.g. by making 'area under the curve' estimations or by modelling the data in various ways. Such analyses generally increase the power to detect differences between groups and over time and may help in simplifying and/or interpreting the results. Such approaches are described in good textbooks [25, 315, 316] but after careful consideration our decision was to keep analyses as simple as possible to preserve the information obtained.

The sample sizes of our studies (papers VII and VIII) allowed detection of differences in mean scores between groups of about 6-8 points and up. A clinically important difference in the EORTC QLQ-C30 and other similar questionnaires is typically viewed as being around 10 on a 0-100 scale, although smaller differences

may also be relevant [25, 45, 317]. Thus, our approach does not seem to have severely limited the ability to detect clinically important differences.

Another advantage of simple approaches is that it is easy for readers to judge the results and to extrapolate them to comparisons with their own data.

Our preference for minimal aggregation of data thus mainly stems from a view of our study as having its main strength in description and simple, clinically anchored comparisons (papers VII and VIII). In contrast, in studies where HRQL data are primary end-points it is of great importance to maximise power and to limit the number of end-points. Therefore, in such studies one could choose to combine data from more than one assessment, to limit analyses to one or a few end-points, or to use more complex statistical models.

### **5.10 The data and analysis for prognostic factors**

In addition to the strengths and weaknesses of the data and analyses discussed in paper IX a few comments will be made here. In general, the quality of the available data must be viewed as close to optimal with regard to its completeness and quality: very high patient participation, high completeness and quality of clinical and biological background variables, and high quality of outcome variables (survival was obtained by linking with the Central Personal Register, the others had been subject to extensive quality control because they formed the basis for randomised trials). Furthermore, the size of the data set was large.

As also discussed in the paper, it is controversial whether the timing of assessments in our study was optimal. We cannot exclude the possibility that other results could have come from assessments at other points in the trajectory but the mechanism of such a difference remains unclear. And even if this were the case, it would not affect the finding that patients' self-assessments about two months after their diagnosis predict survival for about 13 years.

We could have included fewer or more than the six variables (plus social class) in the analyses. If we had had to focus on one variable only we would have selected the EORTC emotional function scale or the HAD depression subscale, and this would have led to close to the same results. Of course, we cannot know what the results from inclusion of additional variables would have been but that could be investigated in the future. The finding of a similar effect across a number of variables reduces the risk that findings are due to chance.

The variables can be handled in many different ways in the analyses, i.e., as continuous, categorised or otherwise transformed. Our choice of a uniform categorisation (dichotomization at the median) across all the six variables reduces the risk of over-estimation of effect. In contrast, the testing of a variable in more than one categorisation (as done for the two HAD subscales in order to test them at the clinical cut points as well) can be criticised, but the findings for the HAD scale were not important for the paper's results.

One can discuss whether our choice of categorisation was the right one but there exists no single commonly agreed 'most appropriate' way of categorising a variable [97] (p. 205-7). Dichotomization at the median of the distribution of a variable is one of a number of common approaches.

One might argue that we were missing important prognostic variables from the biological/clinical model. However, we included the commonly used variables and more variables than most of the other studies. Furthermore, we included more infor-

mation about prognosis than patients had been given (i.e., as we included the data available, the doctor could not have told the patient more, and the vast majority of patients have not been given all details). Finally, when we compared the prognostic ability of a self-rated variable with and without control for the biological model the results were about the same – in fact, there tended to be an increase in predictive power when we adjusted for the biological model. This may appear surprising but is not uncommon [318].

Based on these considerations, our findings – which contradict most of the literature – do not appear to be artefacts resulting from obvious methodological weaknesses.

## 6 CONCLUSIONS

This thesis has shown that it is possible to carry out a large, prospective study of primary breast cancer patients' health-related quality of life (HRQL) with a very high level of patient participation, to elucidate the patients' HRQL in detail, and to produce clinically relevant information that can be used in several ways. The methodological parts of the study generally supported the validity of the questionnaire-based methodology but also identified limitations, particularly if questionnaires are used to compare groups of persons who are in very different circumstances. Finally, an unexpected result was the predictive ability of patients' self-assessments of their HRQL.

Referring to the aims, the overall conclusions are:

1. The questionnaire composed for this study was feasible for use in a longitudinal study and was shown to have excellent content validity. It made it possible to assess the impact of early breast cancer and adjuvant therapy on HRQL in more detail than in previous studies.
2. In general, the multi-item scales included in the questionnaire were adequate representations of the information collected through their items. Thus, despite some findings of differential item functioning (DIF), the frequency, magnitude and practical importance of DIF in the EORTC QLQ-C30 and the HAD Scale were very limited.
3. To a large extent patients understood and responded to the items of the questionnaire in the same way as did the researchers. This indicates that, in general, patients answered the questionnaire without frequent misunderstandings or other errors in their responses. However, a mechanism that in some cases may lead to under-estimation of the levels of symptoms and problems (and was termed 'selective reporting') was identified.
4. The views and experiences of health care professionals involved in the treatment of breast cancer patients did not contribute substantially to the handling of analytical problems related to multiple hypothesis testing. Health care professionals' insight into patients' HRQL was found to be limited.
5. A general population study involving a large group of Danish women was conducted, but previously unacknowledged problems with the use of such data were identified. These problems, which may be at least partially caused by the 'selective reporting' mechanism identified in the validation study, may invalidate comparisons of data from patients against general population data.
6. Due to the problems identified in relation to comparisons between general population data and (breast cancer) patients, the prevalence of anxiety and depression in newly di-

agnosed breast cancer patients relative to women of the general population could not be reliably evaluated.

7. Many important differences in HRQL were found between premenopausal low-risk patients not offered any systemic therapy and patients on chemotherapy.
8. Several important differences in HRQL were found between premenopausal patients with receptor-positive tumours randomised to chemotherapy or ovarian ablation.
9. Psychological distress and fatigue were found to carry prognostic information independent of biological variables.

## 7 PERSPECTIVES FOR FUTURE RESEARCH

This thesis has presented an early example of HRQL research. Since the conception of the study, extensive experience with HRQL research has been achieved. For example, the number of PubMed publications indexed with 'quality of life' as MeSH term in combination with 'breast neoplasms' increased more than tenfold from 281 in 1990 to 3,205 in June 2007 ([www.pubmed.gov](http://www.pubmed.gov), accessed 6 June 2007).

This development, where traditional biomedical outcomes are increasingly being supplemented by patient-rated outcomes, is likely to continue. In addition to the medical background presented earlier it also builds on political evolutions including increased focus on the patient as a participant in the decision-making process who needs sufficient information (e.g., patient empowerment), and increasing recognition of the importance of the consumer-perspective in the development of the health care system.

Breast cancer is probably the disease that has fostered the most HRQL research. Of 59,205 publications on 'quality of life' (MeSH term) in PubMed, more than 5% were related to breast cancer ([www.pubmed.gov](http://www.pubmed.gov), accessed 6 June 2007).

Despite this extensive research there are still many methodological challenges and considerable controversy and uncertainty as to what role HRQL research is to have in clinical trials, medical decision-making, and clinical practice.

### 7.1 Methodological aspects

#### 7.1.1 The content of questionnaires

One of the most basic methodological issues in HRQL is the content of the questionnaires used. As shown in this thesis, most of the published research related to breast cancer adjuvant therapy has employed questionnaires that did not comprehensively elucidate the HRQL associated with even the most commonly used types of chemotherapy. The development of breast-cancer focused supplements to the most widely used standard questionnaires for cancer patients (EORTC QLQ-B23 [201] and FACT-B [319]) was an important advance, but as the treatments used for adjuvant therapy of breast cancer are rapidly changing, additional research is needed to investigate which HRQL aspects are affected by the new treatments. It will be necessary either to revise the existing questionnaires or to develop new, supplementary questionnaires [268]: like any other aspect of validation, the content validation of questionnaires is an ongoing process [320].

#### 7.1.2 Validation of questionnaires

In addition to the work needed to secure that the right questions are asked, the ongoing work with validation of questionnaires for breast cancer research should follow multiple tracks.

An important line of future research is the application of cognitive interviewing techniques to explore respondents' perception of

questionnaire items. These techniques elucidate the cognitive processes involved in the completion of questionnaires, and include verbal probing by an interviewer, 'think-aloud' techniques, etc. [34, 84, 85]. Our validation study (paper III) comparing patients' and observers' ratings was one of the first applications of such techniques in oncology [34]. Cognitive interviewing are useful both in the development (i.e., as an extension of the traditional pilot testing) and in the validation of existing methods [34].

Whereas DIF analyses were virtually unknown in the HRQL field when paper II was published, the relevance of DIF analyses is now widely acknowledged. For example, in 'The ten Ds of health outcomes measurement for the twenty-first century', McHorney and Cook emphasise the importance of DIF analyses as being one of the 'ten Ds' [321], and this methodology also forms part of most work with item response and computer-adaptive testing (see below). We recently presented recommendations for the future use of DIF analyses in clinical trials [67].

There are several important aspects of the validation and evaluation of questionnaires that have not been touched upon in this thesis, e.g., investigation of sensitivity and responsiveness, floor/ceiling-effects, and the use of various psychometric methods such as multitrait scaling and factor-analysis-based methods. These issues will not be further discussed but clearly have important roles also in future work.

Investigation of the equivalence of translations of standard instruments and of potential cross-cultural differences also remains an important area. Our recent applications of DIF analyses are relevant examples [65, 66, 74] but other approaches are also needed in this emerging area.

### 7.1.3 Clinical significance

It is a paradox that the usual approaches to analysis of standard questionnaires involve construction of multi-item scales whereby the responses to items that were originally simple and easily interpretable (e.g., 34% had 'Very much pain'), become abstract scores. One can rightly ask how a difference in role function of 7 points on a 0-100 scale should be interpreted. It may be statistically significant but is it clinically significant and how? Furthermore, any researcher planning an intervention study (e.g., a randomised trial) must specify the magnitude of anticipated difference between groups or over time, and to do this one must make assumptions about the minimal, clinically relevant difference [20]. These problems have led to extensive research into interpretation of scores on HRQL instruments [25, 130, 322-327]. While some clarity has been achieved [326, 327] there is still a pronounced need for more research. For example, when there are a number of different treatments, how should the magnitudes of differences on various dimensions be explained to patients? This applies to the scenarios discussed in this thesis in relation to premenopausal, node-positive and receptor-positive breast cancer patients and to many other situations.

### 7.1.4 The use of general population studies

The conclusion in this thesis that direct comparison of HRQL scores from breast cancer patients to a general population sample may be invalid is not generally accepted. This important issue needs further investigation.

### 7.1.5 Item response theory, item banking, and computer-adaptive testing

The last decade has seen a strong increase in the interest in application of modern psychometric methods to HRQL assessment, and their importance is now widely acknowledged [20, 328-332]. Item response theory (IRT) methods have proven useful in shortening HRQL scales with no or little loss of information [333-335] and in developing item banks that can be administered using so-called computer-adaptive testing (CAT) [329-331, 336-338]. With IRT and CAT higher measurement precision may be obtained with the same number of items administered, the number and nature of items may be varied between respondents, and results may be compared across studies using different items. Much of the future research in HRQL assessment will focus on such methodology as witnessed in the large American Patient-Reported Outcomes Measurement Information System (PROMIS) [339] and, on a smaller scale, in an ongoing, Danish led development of a CAT-based, interactive version of the EORTC QLQ-C30.

## 7.2 Clinical aspects

### 7.2.1 HRQL assessment in breast cancer clinical trials

A systematic review of randomised breast cancer trials evaluating HRQL was published in 2003 [268] and was further discussed in a subsequent publication [340]. It included literature searches until June 2001. The review identified 66 trials of which 46 evaluated biomedical interventions and 20 evaluated psychosocial interventions; only 7 trials concerned adjuvant therapy. The authors concluded that the HRQL results did not affect clinical decision-making in any of the seven trials but also acknowledged that other results might come from ongoing studies. They argued that in the future HRQL should mainly be assessed in adjuvant therapy trials when treatments are expected to have equivalent effect on recurrence and survival or when long-term effects (e.g., on cognitive function, menopause) are expected [268].

In my view this recommendation implies that HRQL is relevant in the majority of trials because the relative effectiveness cannot be known a priori and because long-term consequences are still unknown for most of the new regimens. Furthermore, even when survival differences are found, there may be short-term differences in HRQL, which may influence the overall conclusions from trials.

However, given the importance of appropriate methodology (and the frustrations resulting from inconclusive data), HRQL assessment should be done only if there are sufficient resources available to make the efforts successful. This concerns the entire process from planning to publication. It is more the rule than the exception that the discussion of possible HRQL assessment is started after most other aspects of clinical trials have been settled, and, of course, this severely limits the possibility of truly integrating the HRQL assessment in the scientific thinking behind the trial. The result is that the potential of HRQL research is under-utilized and this, again, impairs its reputation.

Therefore, rather than incorporating HRQL assessment into all trials, it can be recommended to select those trials where the methodology can be sufficient. For such studies it is worthwhile to spend the necessary time on questionnaire composition/development to secure that the relevant HRQL dimensions are actually covered. Thoughtful preparation may also optimize the utility of HRQL data by using the opportunity to answer additional, clinically or methodologically relevant questions.

### 7.2.2 *The use of HRQL data in clinical decision-making*

In an editorial on the status of HRQL research Levine and Ganz noted that 'it is disappointing that there are relatively few examples of formal quality-of-life measurement that have influenced individual patient decision-making or treatment policies.' [341]. They encouraged 'the translation of quality-of-life measurement into clinical practice to improve patient care' [341]. The use of HRQL data in decision-making is insufficiently investigated. HRQL data are often complex (as seen in this thesis) and even without HRQL data it is often complicated to inform patients about several treatment modalities and options. Various decision-aids have been developed and may be helpful [342-344], and information needs have been explored [244-247, 345]. However, it would be relevant to further investigate patients' priorities concerning HRQL information: how detailed and in which form do the patients want the information to be given? And to what extent is it desirable and feasible to present and compare HRQL profiles of more than one treatment?

### 7.2.3 *The use of HRQL assessment in clinical practice*

A number of studies have investigated whether HRQL questionnaires completed by the patient before meeting the doctor may improve communication or treatment. The aim of such research, where the idea is to use the individual patient's own data, is thus different from the use of published HRQL data (from other patients) in decision-making as discussed in the previous section. Part of the rationale is the discrepancy observed between patients' and physicians' evaluation of HRQL [79, 80, 86], which implies that physicians' insight in the patients' situation might be improved if they had access to the HRQL questionnaires. Ideally, such HRQL data might lead to better treatment of symptoms and other problems, to better communication, and to higher satisfaction. However, despite relatively extensive research and encouraging results it is still not clear how and to what extent HRQL questionnaires may best be used in clinical practice [346-352]. This is an important, future field of research.

## 8 ENGLISH SUMMARY

The treatment of primary breast cancer usually consists of surgery often followed by adjuvant therapy (radiotherapy, chemotherapy, hormonal treatment, etc.) to reduce the risk of recurrence. The cancer diagnosis and the treatments may have significant impact on the patients' quality of life.

This thesis deals with scientific aspects and clinical results of a study aimed at assessing the impact of breast cancer (and its treatment) on the patients' quality of life. Studies such as this assessing the problems and symptoms experienced by the patients are often referred to as health-related quality of life (HRQL) research. HRQL research deals with subjective experiences and raises challenging, scientific questions. Therefore, much attention was directed towards methodological issues in this clinically motivated project.

The study was a prospective, longitudinal, questionnaire-based investigation of women with newly diagnosed breast cancer registered in the Danish Breast Cancer Co-operative Group's DBCG 89 Program. The patients were sub-divided into low-risk and high-risk patients. High-risk patients were offered randomisation in one of three randomised adjuvant therapy trials involving chemotherapy, ovarian ablation, and endocrine therapy.

After a literature study and interviews with breast cancer patients, a questionnaire was composed that included two widely used standard questionnaires (EORTC QLQ-C30 and Hospital

Anxiety and Depression (HAD) Scale) and a DBCG 89 Questionnaire developed for this study.

A total of 1,898 eligible patients were invited by post to participate in the study involving six assessments over a 2-year period, and 1,713 patients (90%) completed the first questionnaire. Furthermore, a questionnaire was sent to 872 women selected at random from the general population; 608 (70%) responded. The multi-item scales of the two standard questionnaires were analysed for so-called differential item functioning (DIF) in order to investigate whether the (summary) scale scores were adequate representations of the information obtained by the individual items. The DIF analyses identified a number of cases of DIF, which, among other things, contributed to detection of possible problems in the HAD Scale. It was concluded that DIF analyses are relevant when important analyses based on multi-item scales are made.

A new way to evaluate the validity of questionnaires was developed. The results from questionnaires completed by patients were compared against results from open-ended interviews with the same patients rated by observers. The idea was that if results were similar, the patients had then probably understood and completed the questionnaire items as intended. On the other hand, if results from self-assessment and interviews deviated, misunderstandings or other errors might have taken place, and the study would give insight into possible problems. Of 57 breast cancer patients, 46 (81%) were successfully interviewed. In general, the agreement between patient-completed questionnaires and interviews was excellent, indicating very good validity. The median weighted kappa for the EORTC QLQ-C30 was 0.85 (range 0.49-1.00); it was 0.79 (range 0.65-0.95) for the HAD Scale, and 0.92 (range 0.51-1.00) for the DBCG 89 Questionnaire. However, the study identified a mechanism called selective reporting, which may affect results from most HRQL questionnaires: in order to provide correct and useful answers some patients do not report symptoms they believe are irrelevant to the study, e.g., symptoms unrelated to cancer. This mechanism may lead to bias if results from patients are compared to results from populations reporting their symptoms more completely, e.g., general population samples. In contrast, this mechanism has little importance when results from different sub-groups of cancer patients are compared.

In this study multiple variables were assessed at multiple points in time and we did not have a priori hypotheses for all these potential comparisons. Therefore, a staff survey involving experienced doctors and nurses was conducted in order to generate hypotheses that could be tested in the data from patients. We contacted 46 health care professionals and 36 (78%) responded. Overall, the staff survey did not prove very useful for the intended purpose. The main reason for this was probably that the health care professionals had limited insight into the patients' HRQL. A different approach to the problem of multiple hypothesis testing proved more useful. Hypotheses generated from the initial literature review were tested in the comparison of patients in chemotherapy against patients not in chemotherapy.

The study of women selected at random from the general population showed that these women experienced a considerable degree of 'morbidity' according to all three questionnaires. This shows that symptoms and problems reported by cancer patients may have causes other than cancer, and thus constitutes a good justification for the use of data from general population studies when interpreting data from cancer patients.

The levels of anxiety and depression of low-risk breast cancer patients were found to be lower than those from the general

population sample. After careful consideration we concluded that this finding was probably incorrect. The most important explanations were thought to be the wording of some HAD Scale items as well as two mechanisms that are not specific to the HAD Scale, the 'selective reporting mechanism' found in the validation study, and the response-shift problem. These findings indicate – in contrast to the conclusion above – that the comparability of HRQL data from cancer patients and general population data must be questioned. However, as this is the first study to raise the problem, this issue needs further investigation.

Based on the initial literature review and interviews we hypothesised that 30 different HRQL issues would be impaired in patients undergoing CMF chemotherapy compared to patients not in chemotherapy; 23 of these hypotheses were confirmed. In addition, our study and other research suggest that other HRQL aspects may also be affected by chemotherapy. Thus, there is considerable evidence that patients in chemotherapy may experience effects on a wide spectrum of HRQL issues.

Most other studies have assessed surprisingly few of the HRQL issues shown in our study to be impaired in patients receiving chemotherapy. Similarly, current review articles on HRQL effects of adjuvant chemotherapy mention only relatively few of these topics.

Concerning HRQL after the treatment period, our main finding was that many symptoms and problems had declined or disappeared, but some persisted: anticipatory nausea, weight gain, endocrine effects (e.g., hot flushes/sweats, irregular bleedings/amenorrhoea, vaginal dryness), disturbed sleep, and sexual dysfunction. These findings are in agreement with the literature. The staff study showed that experienced physicians and nurses did not expect many of the 'scientifically well documented' consequences of chemotherapy.

Taken together, our findings suggest that information to patients about chemotherapy should be more comprehensive than that which has been practised in most places.

When compared against ovarian ablation, chemotherapy was associated with more impact on HRQL during the treatment period; only hot flushes/sweats were more pronounced in the ovarian ablation group. Thus, from an overall 'HRQL perspective' ovarian ablation or suppression may be preferable. However, younger women may preserve their premenopausal status (including fertility) by having chemotherapy, and this may be an argument for chemotherapy or for temporary ovarian ablation via goserelin, rather than permanent ovarian ablation. Furthermore, while ovarian ablation/suppression may be preferable because of less impairment of HRQL, contemporary chemotherapeutic regimens may be more effective. These results indicate that for some patients, the HRQL data and results on treatment efficiency may be in conflict. There is no simple, universally correct solution to this dilemma. More research into patients' views and expectations to the health-care system in cases where medical decision-making involves complex trade-offs between treatment efficiency and HRQL issues is needed.

Contrary to expectations, the analyses showed that fatigue and emotional function predicted the risk of recurrence and death independently of biological and clinical prognostic variables. In multivariate Cox regression analyses patients who were more fatigued or had poorer emotional function had a worse prognosis. These results are consistent with one small study, but are inconsistent with five similar studies in patients with primary breast cancer, which found no such associations. The reasons for these important differences are currently unknown.

In conclusion, this study consisted of methodological and clinical investigations of HRQL in primary breast cancer patients. The initial questionnaire development resulted in a combination of questionnaires that was more comprehensive than in other similar studies. The results of the methodological studies generally supported the validity of the questionnaires but also gave important insights into potential scientific problems that are probably not restricted to the present study. These insights helped to prevent misinterpretations of the clinical data. The study provided the most detailed description of HRQL during and after breast cancer adjuvant chemotherapy to date, and compared results of chemotherapy against ovarian ablation. It also provided controversial results concerning the prognostic value of HRQL data. The combination of a large empirical study and several methodological sub-studies thus proved useful and gave new results.

## 9 REFERENCES

1. Sundhedsstyrelsen. Cancerregisteret 2003 (foreløbig opgørelse). København: Sundhedsstyrelsen; 2005.
2. Sundhedsstyrelsen. Dødsårsagsregisteret 2005 (foreløbige tal). København: Sundhedsstyrelsen; 2006.
3. Clemmensen IH, Nedergaard KH, Storm HH. Kræft i Danmark. København: FADL's Forlag; 2006.
4. Overgaard M, Hansen PS, Overgaard J, Rose C, Andersson M, Bach F, et al. Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy. *N Engl J Med* 337: 949-955, 1997.
5. Danish Breast Cancer Cooperative Group. DBCG's retningslinjer 2007. København: DBCG; 2007.
6. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 353: 1659-1672, 2005.
7. Jonat W, Pritchard KI, Sainsbury R, Klijn JG. Trends in endocrine therapy and chemotherapy for early breast cancer: a focus on the premenopausal patient. *J Cancer Res Clin Oncol* 132: 275-286, 2006.
8. Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. *N Engl J Med* 319: 1681-1692, 1988.
9. Danish Breast Cancer Cooperative Group. [DBCG-89. Program for Treatment and Follow-Up of Patients With Primary, Operable Breast Cancer] [In Danish]. Copenhagen: DBCG-Sekretariatet; 1989.
10. Danish Breast Cancer Cooperative Group. [Danish Breast Cancer Cooperative Group 1977-1997] [In Danish]. Copenhagen: DBCG-Sekretariatet; 1998.
11. Rose DP, Davis TE. Effects of adjuvant chemohormonal therapy on the ovarian and adrenal function of breast cancer patients. *Cancer Res* 40: 4043-4047, 1980.
12. Brincker H, Rose C, Rank F, Mouridsen HT, Jakobsen A, Døbernowsky P, et al. Evidence of a castration-mediated effect of adjuvant cytotoxic chemotherapy in premenopausal breast cancer. *J Clin Oncol* 5: 1771-1778, 1987.
13. Buckley MMT, Goa KL. Tamoxifen. A reappraisal of its pharmacodynamic and pharmacokinetic properties, and therapeutic use. *Drugs* 37: 451-490, 1989.
14. Shariff S, Cumming CE, Lees A, Handman M, Cumming DC. Mood disorder in women with early breast cancer taking tamoxifen, an estradiol receptor antagonist. An expected or unexpected effect? *Ann N Y Acad Sci* 761: 365-368, 1995.

15. Idler EL, Kasl SV, Lemke JH. Self-evaluated health and mortality among the elderly in New Haven, Connecticut, and Iowa and Washington counties, Iowa, 1982-1986. *Am J Epidemiol* 131: 91-103, 1990.
16. Idler EL, Angel RJ. Self-rated health and mortality in the NHANES-I Epidemiologic Follow-up Study. *Am J Public Health* 80: 446-452, 1990.
17. Spiegel D, Bloom JR, Kraemer HC, Gottheil E. Effects of psychosocial treatment on survival of patients with metastatic breast cancer. *Lancet* 2: 888-891, 1989.
18. Hewitt M, Herdman R, Holland J, and National Cancer Policy Board (U.S.). Meeting Psychosocial Needs of women with Breast Cancer. Washington, D.C.: National Academies Press; 2004.
19. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 118: 622-629, 1993.
20. Wood-Dauphinee S. Assessing quality of life in clinical research: from where have we come and where are we going? *J Clin Epidemiol* 52: 355-363, 1999.
21. Muldoon MF, Barger SD, Flory JD, Manuck SB. What are quality of life measurements measuring? *Br Med J* 316: 542-545, 1998.
22. Slevin ML, Plant H, Lynch D, Drinkwater J, Gregory WM. Who should measure quality of life, the doctor or the patient? *Br J Cancer* 57: 109-112, 1988.
23. Grønvold M, Iversen L, Mouridsen HT. [Quality of life with adjuvant therapy of breast cancer. A study associated with the Danish Breast Cancer Cooperative Group's (DBCG) Program for treatment and follow-up of patients with primary, operable breast cancer (DBCG-89)] [In Danish]. Copenhagen: Institut for Social Medicin, Københavns Universitet og Onkologisk afd. ONK, Rigshospitalet; 1991.
24. Ferrans CE. Definitions and conceptual models of quality of life. In: Lipscomb J, Gotay CC, and Snyder C. *Outcomes Assessment in Cancer: Measures, Methods, and Applications*. Pp. 14-30. Cambridge: Cambridge University Press; 2005.
25. Fayers PM, Machin D. *Quality of Life: the Assessment, Analysis, and interpretation of patient-reported outcomes*. 2nd ed. Chichester, UK: Wiley; 2007.
26. Norredam M, Grønvold M, Petersen JH, Krasnik A. Effect of social class on tumour size at diagnosis and surgical treatment in Danish women with breast cancer. *Soc Sci Med* 47: 1659-1663, 1998.
27. Grønvold M, Nielsen SS, Petersen MAa. [Is there social inequality in the occurrence of physical and psychological symptoms among patients receiving chemotherapy for breast cancer?] [In Danish]. In: Vallgård S and Koch L. *Forskel og lighed i sundhed og sygdom*. Pp. 109-122. Copenhagen: Munksgaard Danmark; 2006.
28. Fayers PM, Hand DJ, Bjordal K, Grønvold M. Causal indicators in quality of life research. *Qual Life Res* 6: 393-406, 1997.
29. Petersen MAa. [Quality of life in breast cancer patients - the consequences of adjuvant therapy for the patients' quality of life] [In Danish]. Copenhagen: University of Copenhagen; 1999.
30. Larsen K. Analysis of questionnaire data from longitudinal studies (PhD thesis). Copenhagen: University of Copenhagen; 2000.
31. Grønvold M. Validation of a quality of life questionnaire for breast cancer patients (PhD thesis). Copenhagen: University of Copenhagen; 1996.
32. Fayers P, Machin D. *Quality of Life - Assessment, Analysis and Interpretation*. Chichester: John Wiley and Sons Ltd.; 2000.
33. Grønvold M. Methodological issues in the assessment of health-related quality of life in palliative care trials. *Acta Anaesthesiol Scand* 43: 948-953, 1999.
34. Willis G, Reeve BB, Barofsky I. The use of cognitive interviewing techniques in quality-of-life and patient-reported outcomes assessment. In: Lipscomb J, Gotay CC, and Snyder C. *Outcomes Assessment in Cancer: Measures, Methods, and Applications*. Pp. 610-622. Cambridge: Cambridge University Press; 2005.
35. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 85: 365-376, 1993.
36. Hansen EJ. [Social classes in Denmark] [In Danish]. Copenhagen: Danish National Institute of Social Research; 1984.
37. Hansen EJ. [The Distribution of Living Conditions] [In Danish]. Copenhagen: Socialforskningsinstituttet; 1978.
38. Davis JA. A partial coefficient for Goodman and Kruskal's gamma. *J Am Stat Assoc* 69: 174-180, 1967.
39. Kreiner S. Rasch 1 & 2. Programs for item analysis. Technical report. Copenhagen: Danish Institute for Educational Research; 1985.
40. Klee M, Thranov I, Machin D. The patients' perspective on physical symptoms after radiotherapy for cervical cancer. *Gynecol Oncol* 76: 14-23, 2000.
41. Klee M, Thranov I, Machin D. Life after radiotherapy: the psychological and social effects experienced by women treated for advanced stages of cervical cancer. *Gynecol Oncol* 76: 5-13, 2000.
42. Klee M, Machin D. Health-related quality of life of patients with endometrial cancer who are disease-free following external irradiation. *Acta Oncol* 40: 816-824, 2001.
43. Klee M, Grønvold M, Machin D. Using data from studies of health-related quality of life to describe clinical issues examples from a longitudinal study of patients with advanced stages of cervical cancer. *Qual Life Res* 8: 733-742, 1999.
44. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 33: 159-174, 1977.
45. Fayers PM, Aaronson N, Bjordal K, et al. *The EORTC QLQ-C30 Scoring Manual*. 3rd edition. Brussels: European Organization for Research and Treatment of Cancer; 2001.
46. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 67: 361-370, 1983.
47. Siegel S, Castellan NJ, Jr. *Nonparametric Statistics for the Behavioral Sciences*. Singapore: McGraw-Hill; 1988.
48. Agresti A. *An Introduction to Categorical Data Analysis*. New York: Wiley Interscience; 1996.
49. SAS Institute Inc. *SAS/STAT User's Guide, Version 6, Fourth Edition, Vol 2*. Cary, NC: SAS Institute Inc.; 1989.
50. Andersen BL. Sexual difficulties for women following cancer treatment. In: Andersen BL. *Women with Cancer. Psychological Perspectives*. Pp. 257-288. New York: Springer-Verlag; 1986.
51. Aaronson NK, Ahmedzai S, Bullinger M, Grabeels D, Estape J, Filiberti A, et al. The EORTC Core Quality-of-Life Questionnaire: interim results of an international field study. In: Osoba D. *Effect of Cancer on Quality of Life*. Pp. 185-203. Boca Raton: CRC Press Inc.; 1991.

52. Fayers P, Bottomley A. Quality of life research within the EORTC - the EORTC QLQ-C30. *Eur J Cancer* 38 (Suppl 4): 125-133, 2002.
53. Fallowfield LJ, Baum M, Maguire GP. Effects of breast conservation on psychological morbidity associated with diagnosis and treatment of early breast cancer. *Br Med J* 293: 1331-1334, 1986.
54. Hopwood P, Thatcher N. Preliminary experience with quality of life evaluation in patients with lung cancer. *Oncology* 4: 158-171, 1990.
55. Ford MF, Jones M, Scannell T, Powell A, Coombes RC, Evans C. Is group psychotherapy feasible for oncology outpatients attenders selected on the basis of psychological morbidity? *Br J Cancer* 62: 624-626, 1990.
56. Snaith RP. The concepts and assessment of depression in oncology. *J Psychosoc Oncol* 5: 133-139, 1987.
57. Maguire P, Selby P. Assessing quality of life in cancer patients. *Br J Cancer* 60: 437-440, 1989.
58. Clark A, Fallowfield LJ. Quality of life measurements in patients with malignant disease: a review. *J R Soc Med* 79: 165-169, 1986.
59. Hagerup L. [Disease and Health. The Glostrup population Studies] [In Danish]. Aarhus: UNI-C; 1987.
60. Aaronson NK. Cross-cultural use of health-related quality of life assessments in clinical oncology. In: Lipscomb J, Gotay CC, and Snyder C. *Outcomes Assessment in Cancer: Measures, Methods, and Applications*. Pp. 406-424. Cambridge: Cambridge University Press; 2005.
61. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale--a review of validation data and clinical results. *J Psychosom Res* 42: 17-41, 1997.
62. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 52: 69-77, 2002.
63. Osborne RH, Elsworth GR, Sprangers MA, Oort FJ, Hopper JL. The value of the Hospital Anxiety and Depression Scale (HADS) for comparing women with early onset breast cancer with population-based reference women. *Qual Life Res* 13: 191-206, 2004.
64. Bjorner JB, Kreiner S, Ware JE, Damsgaard MT, Bech P. Differential item functioning in the Danish translation of the SF-36. *J Clin Epidemiol* 51: 1189-1202, 1998.
65. Petersen MA, Groenvold M, Bjorner JB, Aaronson N, Conroy T, Cull A, et al. Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Qual Life Res* 12: 373-385, 2003.
66. Scott NW, Fayers PM, Bottomley A, Aaronson NK, De Graeff A, Groenvold M, et al. Comparing translations of the EORTC QLQ-C30 using differential item functioning analyses. *Qual Life Res* 15: 1103-1115, 2006.
67. Groenvold M, Petersen MA. The role and use of differential item functioning (DIF) analysis of quality of life data from clinical trials. In: Fayers P and Hays R. *Assessing Quality of Life in Clinical Trials*. 2nd ed. Pp. 195-208. Oxford: Oxford University Press; 2005.
68. Crane PK, Gibbons LE, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Med Care* 44: S115-S123, 2006.
69. Crane PK, van Belle G, Larson EB. Test bias in a cognitive test: differential item functioning in the CASI. *Stat Med* 23: 241-256, 2004.
70. Crane PK. Commentary on Comparing translations of the EORTC QLQ-C30 using differential item functioning analyses. *Qual Life Res* 15: 1117-1118, 2006.
71. Petersen MA, Groenvold M, Aaronson N, Brenne E, Fayers P, Nielsen JD, et al. Scoring based on item response theory did not alter the measurement ability of EORTC QLQ-C30 scales. *J Clin Epidemiol* 58: 902-908, 2005.
72. Scott NW, Fayers PM, Bottomley A, Aaronson NK, De Graeff A, Groenvold M, et al. Response to Commentary on Comparing translations of the EORTC QLQ-C30 using differential item functioning analyses. *Qual Life Res* 15: 1119-1120, 2006.
73. Holland PW, Wainer HE. *Differential Item Functioning*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.; 1993.
74. Scott NW, Fayers PM, Aaronson NK, Bottomley A, de GA, Groenvold M, et al. The use of differential item functioning analyses to identify cultural differences in responses to the EORTC QLQ-C30. *Qual Life Res* 16: 115-129, 2007.
75. Rieker PP, Fitzgerald EM, Kalish LA, Richie JP, Lederman GS, Edbril SD, et al. Psychosocial factors, curative therapies, and behavioral outcomes. A comparison of testis cancer survivors and a control group of healthy men. *Cancer* 64: 2399-2407, 1989.
76. Andrykowski MA, Brady MJ, Hunt JW. Positive psychosocial adjustment in potential bone marrow transplant recipients: Cancer as a psychosocial transition. *Psycho-Oncology* 2: 261-276, 1993.
77. Taylor SE. Adjustment to threatening events. *Am Psychol* 38: 1161-1173, 1983.
78. Andrykowski MA, Curran SL, Studts JL, Cunningham L, Carpenter JS, McGrath PC, et al. Psychosocial adjustment and quality of life in women with breast cancer and benign breast problems: a controlled comparison. *J Clin Epidemiol* 49: 827-834, 1996.
79. Sprangers MAG, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J Clin Epidemiol* 45: 743-760, 1992.
80. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *J Clin Epidemiol* 55: 1130-1143, 2002.
81. Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S. Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire. *J Clin Oncol* 13: 1249-1254, 1995.
82. Jensen PT, Klee MC, Groenvold M. Validation of a questionnaire for self-rating of urological and gynaecological morbidity after treatment of gynaecological cancer. *Radiother Oncol* 65: 29-38, 2002.
83. Jensen PT, Klee MC, Thranov I, Groenvold M. Validation of a questionnaire for self-assessment of sexual function and vaginal changes after gynaecological cancer. *Psychooncology* 13: 577-592, 2004.
84. Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge: Cambridge University Press; 2000.
85. Willis GB. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks: Sage Publications; 2005.
86. Stromgren AS, Groenvold M, Pedersen L, Olsen AK, Spile M, Sjogren P. Does the medical record cover the symptoms experienced by cancer patients receiving palliative care? A comparison of the record and patient self-rating. *Journal of Pain and Symptom management* 21: 189-196, 2001.

87. Stromgren AS, Groenvold M, Sorensen A, Andersen L. Symptom recognition in advanced cancer. A comparison of nursing records against patient self-rating. *Acta Anaesthesiol Scand* 54: 1080-1085, 2001.
88. Homsí J, Walsh D, Rivera N, Rybicki LA, Nelson KA, Legrand SB, et al. Symptom evaluation in palliative medicine: patient report vs systematic assessment. *Support Care Cancer* 14: 444-453, 2006.
89. Petersen MA, Larsen H, Pedersen L, Sonne N, Groenvold M. Assessing health-related quality of life in palliative care: comparing patient and physician assessments. *Eur J Cancer* 42: 1159-1166, 2006.
90. Ware JE, Snow KK, Kosinski M, et al. The SF-36 health survey. Manual and interpretation guide. Boston: The Health Institute, New England Medical Center; 1993.
91. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston: The Health Institute; 1994.
92. Fallowfield LJ, Bliss JM, Porter LS, Price MH, Snowdon CF, Jones SE, et al. Quality of life in the intergroup exemestane study: a randomized trial of exemestane versus continued tamoxifen after 2 to 3 years of tamoxifen in postmenopausal women with primary breast cancer. *J Clin Oncol* 24: 910-917, 2006.
93. Fairclough D. Analysing longitudinal studies of QoL. In: Fayers P and Hays R. *Assessing Quality of Life in Clinical Trials*. 2nd ed. Pp. 149-165. Oxford: Oxford University Press; 2005.
94. Diggle PJ, Liang K-Y, Zeger SL, Atkinson AC, Copas JB, Pierce DA, Schervish MJ, and Titterton DM. *Analysis of longitudinal data*. Oxford: Oxford University Press; 1994.
95. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *Br Med J* 310: 170, 1995.
96. Michels KB, Rosner BA. Data trawling: to fish or not to fish. *Lancet* 348: 1152-1153, 1996.
97. Greenland S, Rothman KJ. *Fundamentals of epidemiologic data analysis*. In: Rothman KJ and Greenland S. *Modern Epidemiology*. Pp. 210-229. Philadelphia, PA: Lippincott-Raven Publishers; 1998.
98. Poole CJ, Earl HM, Hiller L, Dunn JA, Bathers S, Grieve RJ, et al. Epirubicin and cyclophosphamide, methotrexate, and fluorouracil as adjuvant therapy for early breast cancer. *N Engl J Med* 355: 1851-1862, 2006.
99. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics* 40: 1079-1087, 1984.
100. Lehman W, Wassmer G, Reitmeir P. Procedures for two-sample comparisons with multiple endpoints controlling the experimentwise error rate. *Biometrics* 47: 511-521, 1991.
101. Kj  ller M, Rasmussen NK. [The Danish Health and Morbidity Survey 2000 & Trends Since 1987] [In Danish]. Copenhagen: Statens Institut for Folkesundhed; 2002.
102. Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Using reference data on quality of life - the importance of adjusting for age and gender, exemplified by the EORTC QLQ-C30 (+3). *Eur J Cancer* 34: 1381-1389, 1998.
103. Michelson H, Bolund C, Brandberg Y. Multiple chronic health problems are negatively associated with health related quality of life (HRQoL) irrespective of age. *Qual Life Res* 9: 1093-1104, 2000.
104. Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Health-related quality of life in the general Norwegian population assessed by the European Organization for Research and Treatment of Cancer core quality of life questionnaire: The QLQ-C30 (+3). *J Clin Oncol* 16: 1188-1196, 1998.
105. Michelson H, Bolund C, Nilsson B, Brandberg Y. Health-related quality of life measured by the EORTC QLQ-C30--reference values from a large sample of Swedish population. *Acta Oncol* 39: 477-484, 2000.
106. Schwarz R, Hinz A. Reference data for the quality of life questionnaire EORTC QLQ-C30 in the general German population. *Eur J Cancer* 37: 1345-1351, 2001.
107. Ahn SH, Park BW, Noh DY, Nam SJ, Lee ES, Lee MK, et al. Health-related quality of life in disease-free survivors of breast cancer with the general population. *Ann Oncol* 18: 173-182, 2007.
108. Watt T, Groenvold M, Bjorner JB, Noerholm V, Rasmussen NA, Bech P. Fatigue in the Danish general population. Influence of sociodemographic factors and disease. *J Epidemiol Community Health* 54: 827-833, 2000.
109. Fayers PM. Interpreting quality of life data: population-based reference data for the EORTC QLQ-C30. *Eur J Cancer* 37: 1331-1334, 2001.
110. Lisspers J, Nygren A, S  derman E. Hospital Anxiety and Depression Scale (HAD): some psychometric data for a Swedish sample. *Acta Psychiatr Scand* 96: 281-286, 1997.
111. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert, et al. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychol Med* 27: 363-370, 1997.
112. Hinz A, Schwarz R. [Anxiety and depression in the general population: normal values in the Hospital Anxiety and Depression Scale]. *Psychother Psychosom Med Psychol* 51: 193-200, 2001.
113. Crawford JR, Henry JD, Crombie C, Taylor EP. Normative data for the HADS from a large non-clinical sample. *Br J Clin Psychol* 40: 429-434, 2001.
114. Gotay CC, Lipscomb J, Snyder CF. Reflections on findings of the Cancer Outcomes Measurement Working Group: moving to the next phase. *J Natl Cancer Inst* 97: 1568-1574, 2005.
115. Maguire GP, Lee EG, Bevington DJ, K  chemann CS, Crabtree RJ, Cornell CE. Psychiatric problems in the first year after mastectomy. *Br Med J* 1: 963-965, 1978.
116. Silberfarb PM, Maurer LH, Crouthamel CS. Psychosocial aspects of neoplastic disease: I. Functional status of breast cancer patients during different treatment regimens. *Am J Psychiatry* : 450-455, 1980.
117. Grandi S, Fava GA, Cunsolo A, Ranieri M, Gozzetti G, Trombini G. Major depression associated with mastectomy. *Med Sci Res* 15: 283-284, 1987.
118. Dean C. Psychiatric morbidity following mastectomy: Preoperative predictors and types of illness. *J Psychosom Res* 31: 385-392, 1987.
119. Hughson AVM, Cooper AF, Mcardle CS, Smith DC. Psychosocial consequences of mastectomy: Levels of morbidity and associated factors. *J Psychosom Res* 32: 383-391, 1988.
120. Maunsell E, Brisson J, Deschenes L. Psychological distress after initial treatment for breast cancer: A comparison of partial and total mastectomy. *J Clin Epidemiol* 42: 765-771, 1989.
121. Holmberg L, Omne-Pont  n M, Burns T, Adami HO, Bergstr  m R. Psychosocial adjustment after mastectomy and breast-conserving treatment. *Cancer* 64: 969-974, 1989.
122. Levy SM, Herberman RB, Lee JK, Lippman ME, d'Angelo T. Breast conservation versus mastectomy: distress sequelae as a function of choice. *J Clin Oncol* 7: 367-375, 1989.

123. Goldberg JA, Scott RN, Davidson PM, Murray GD, Stallard S, George WD, et al. Psychological morbidity in the first year after breast surgery. *Eur J Surg Oncol* 18: 327-331, 1992.
124. Schag CA, Ganz PA, Polinsky ML, Fred C, Hirji K, Petersen L. Characteristics of women at risk for psychosocial distress in the year after breast cancer. *J Clin Oncol* 11: 783-793, 1993.
125. Ramirez AJ, Richards MA, Jarrett SR, Fentiman IS. Can mood disorder in women with breast cancer be identified preoperatively? *Br J Cancer* 72: 1509-1512, 1995.
126. Longman AJ, Braden CJ, Mishel MH. Side effects burden in women with breast cancer. *Cancer Pract* 4: 274-280, 1996.
127. McArdle JM, George WD, McArdle CS, Smith DC, Moodie AR, Hughson AV, et al. Psychological support for patients undergoing breast cancer surgery: a randomised study. *Br Med J* 312: 813-816, 1996.
128. Tjemsland L, Soreide JA, Malt UF. Traumatic distress symptoms in early breast cancer. II: Outcome six weeks post surgery. *Psycho-Oncology* 5: 295-303, 1996.
129. Osborne RH, Elsworth GR, Hopper JL. Age-specific norms and determinants of anxiety and depression in 731 women with breast cancer recruited through a population-based cancer registry. *Eur J Cancer* 39: 755-762, 2003.
130. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 16: 139-144, 1998.
131. Love RR, Leventhal H, Easterling DV, Nerenz DR. Side effects and emotional distress during cancer chemotherapy. *Cancer* 63: 604-612, 1989.
132. Macquart-Moulin G, Viens P, Bouscary ML, Genre D, Resbeut M, Gravis, et al. Discordance between physicians' estimations and breast cancer patients' self-assessment of side-effects of chemotherapy: an issue for quality of care. *Br J Cancer* 76: 1640-1645, 1997.
133. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 337: 867-872, 1991.
134. Petticrew M, Bell R, Hunter D. Influence of psychological coping on survival and recurrence in people with cancer: systematic review. *BMJ* 325: 1066, 2002.
135. Andrykowski MA, Jacobsen PB, Marks E, Gorfinkle K, Hakes TB, Kaufman RJ, et al. Prevalence, predictors, and course of anticipatory nausea in women receiving adjuvant chemotherapy for breast cancer. *Cancer* 62: 2607-2613, 1988.
136. Andrykowski MA, Schmidt JE, Salsman JM, Beacham AO, Jacobsen PB. Use of a case definition approach to identify cancer-related fatigue in women undergoing adjuvant therapy for breast cancer. *J Clin Oncol* 23: 6613-6622, 2005.
137. Arora NK, Gustafson DH, Hawkins RP, McTavish F, Cella DF, Pingree S, et al. Impact of surgery and chemotherapy on the quality of life of younger women with breast carcinoma: a prospective study. *Cancer* 92: 1288-1298, 2001.
138. Beisecker A, Cook MR, Ashworth J, Hayes J, Brecheisen M, Helmig L, et al. Side effects of adjuvant chemotherapy: perceptions of node-negative breast cancer patients. *Psycho-Oncology* 6: 85-93, 1997.
139. Berglund G, Nystedt M, Bolund C, Sjoden PO, Rutquist LE. Effect of endocrine treatment on sexuality in premenopausal breast cancer patients: a prospective randomized study. *J Clin Oncol* 19: 2788-2796, 2001.
140. Bernhard J, Zahrieh D, Coates AS, Gelber RD, Castiglione-Gertsch M, Murray E, et al. Quantifying trade-offs: quality of life and quality-adjusted survival in a randomised trial of chemotherapy in postmenopausal patients with lymph node-negative breast cancer. *Br J Cancer* 91: 1893-1901, 2004.
141. Bottomley A, Therasse P, Piccart M, Efficace F, Coens C, Gotay C, et al. Health-related quality of life in survivors of locally advanced breast cancer: an international randomised controlled phase III trial. *Lancet Oncol* 6: 287-294, 2005.
142. Boyd NF, Selby PJ, Sutherland H, Hogg S. Measurement of the clinical status of patients with breast cancer: Evidence for the validity of self assessment with linear analogue scales. *J Clin Epidemiol* 41: 243-250, 1988.
143. Brezden CB, Phillips KA, Abdoell M, Bunston T, Tannock IF. Cognitive function in breast cancer patients receiving adjuvant chemotherapy. *J Clin Oncol* 18: 2695-2701, 2000.
144. Burwell SR, Case LD, Kaelin C, Avis NE. Sexual problems in younger women after breast cancer surgery. *J Clin Oncol* 24: 2815-2821, 2006.
145. Camoriano JK, Loprinzi CL, Ingle JN, Therneau TM, Krook JE, Veeder MH. Weight change in women treated with adjuvant therapy or observed following mastectomy for node-positive breast cancer. *J Clin Oncol* 8: 1327-1334, 1995.
146. de Jong N, Candel MJ, Schouten HC, Abu-Saad HH, Courtens AM. Prevalence and course of fatigue in breast cancer patients receiving adjuvant chemotherapy. *Ann Oncol* 15: 896-905, 2004.
147. Demark-Wahnefried W, Winer EP, Rimer B. Why women gain weight with adjuvant chemotherapy for breast cancer. *J Clin Oncol* 11: 1418-1429, 1993.
148. Downie FP, Mar Fan HG, Houede-Tchen N, Yi Q, Tannock IF. Cognitive function, fatigue, and menopausal symptoms in breast cancer patients receiving adjuvant chemotherapy: evaluation with patient interview after formal assessment. *Psychooncology* 15: 921-930, 2006.
149. Duric VM, Stockler MR, Heritier S, Boyle F, Beith J, Sullivan A, et al. Patients' preferences for adjuvant chemotherapy in early breast cancer: what makes AC and CMF worthwhile now? *Ann Oncol* 16: 1786-1794, 2005.
150. Fairclough DL, Fetting JH, Cella D, Wonson W, Moinpour CM. Quality of life and quality adjusted survival for breast cancer patients receiving adjuvant therapy. Eastern Cooperative Oncology Group (ECOG). *Qual Life Res* 8: 723-731, 1999.
151. Galalae RM, Michel J, Siebmann JU, Kuchler T, Eilf K, Kimmig B. Significant negative impact of adjuvant chemotherapy on health-related quality of life (HR-QoL) in women with breast cancer treated by conserving surgery and postoperative 3-D radiotherapy. A prospective measurement. *Strahlenther Onkol* 181: 645-651, 2005.
152. Ganz PA, Kwan L, Stanton AL, Krupnick JL, Rowland JH, Meyerowitz BE, et al. Quality of life at the end of primary treatment of breast cancer: first results from the moving beyond cancer randomized trial. *J Natl Cancer Inst* 96: 376-387, 2004.
153. Goodwin PJ, Ennis M, Pritchard KI, Trudeau M, Hood N. Risk of menopause during the first year after breast cancer diagnosis. *J Clin Oncol* 17: 2365-2370, 1999.
154. Hoskins CN. Breast cancer treatment-related patterns in side effects, psychological distress, and perceived health status. *Oncol Nurs Forum* 24: 1575-1583, 1997.
155. Hughson AVM, Cooper AF, McArdle CS, Smith DC. Psychological impact of adjuvant chemotherapy in the first two years after mastectomy. *Br Med J* 293: 1268-1271, 1986.
156. Hurny C, Bernhard J, Coates AS, Castiglione-Gertsch M, Peterson HF, Gelber RD, et al. Impact of adjuvant therapy on quality of life in women with node-positive operable breast cancer. International Breast Cancer Study Group. *Lancet* 347: 1279-1284, 1996.

157. Jacobsen PB, Hann DM, Azzarello LM, Horton J, Balducci L, Lyman GH. Fatigue in women receiving adjuvant chemotherapy for breast cancer: characteristics, course, and correlates. *J Pain Symptom Manage* 18: 233-242, 1999.
158. Knobf MT. Physical and psychologic distress associated with adjuvant chemotherapy in women with breast cancer. *J Clin Oncol* 4: 678-684, 1986.
159. Land SR, Kopec JA, Yothers G, Anderson S, Day R, Tang G, et al. Health-related quality of life in axillary node-negative, estrogen receptor-negative breast cancer patients undergoing AC versus CMF chemotherapy: findings from the National Surgical Adjuvant Breast and Bowel Project B-23. *Breast Cancer Res Treat* 86: 153-164, 2004.
160. Levine MN, Guyatt GH, Gent M, Pauw SD, Goodyear MD, Hryniuk WM, et al. Quality of life in stage II breast cancer: an instrument for clinical trials. *J Clin Oncol* 6: 1798-1810, 1988.
161. Lu W, Cui Y, Zheng Y, Gu K, Cai H, Li Q, et al. Impact of newly diagnosed breast cancer on quality of life among Chinese women. *Breast Cancer Res Treat* 102: 201-210, 2007.
162. Malinowszky KM, Gould A, Foster E, Cameron D, Humphreys A, Crown J, et al. Quality of life and sexual function after high-dose or conventional chemotherapy for high-risk breast cancer. *Br J Cancer* 95: 1626-1631, 2006.
163. Mcardle CS, Calman KC, Cooper AF, Hughson AVM, Russell AR, Smith DC. The social, emotional and financial implications of adjuvant chemotherapy in breast cancer. *Br J Surgery* 68: 261-264, 1981.
164. Meyerowitz BE, Sparks FC, Spears IK. Adjuvant chemotherapy for breast carcinoma. Psychosocial implications. *Cancer* 43: 1613-1618, 1979.
165. Meyerowitz BE, Watkins IK, Sparks FC. Quality of life for breast cancer patients receiving adjuvant chemotherapy. *Am J Nurs* 83: 232-234, 1983.
166. Montgomery GH, Bovbjerg DH. The development of anticipatory nausea in patients receiving adjuvant chemotherapy for breast cancer. *Physiol Behav* 61: 737-741, 1997.
167. Nystedt M, Berglund G, Bolund C, Fornander T, Rutqvist LE. Side effects of adjuvant endocrine treatment in premenopausal breast cancer patients: a prospective randomized study. *J Clin Oncol* 21: 1836-1844, 2003.
168. Tchen N, Juffs HG, Downie FP, Yi QL, Hu H, Chemerynsky I, et al. Cognitive function, fatigue, and menopausal symptoms in women receiving adjuvant chemotherapy for breast cancer. *J Clin Oncol* 21: 4175-4183, 2003.
169. Tierney AJ, Leonard RCF, Taylor J, Closs SJ, Chetty U, Rodger A. Side effects expected and experienced by women receiving chemotherapy for breast cancer. *Br Med J* 302: 272, 1991.
170. van Dam FSAM, Linsen ACG, Engelsman E, van Benthem J, Hanewald GJFP. Life with cytostatic drugs. In: Mouridsen HT and Palshof T. *Breast cancer - experimental and clinical aspects*. Pp. 229-233. Oxford: Pergamon Press; 1980.
171. Eifel P, Axelson JA, Costa J, Crowley J, Curran WJ, Jr., Deshler A, et al. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst* 93: 979-989, 2001.
172. Shapiro CL, Recht A. Side effects of adjuvant treatment of breast cancer. *N Engl J Med* 344: 1997-2008, 2001.
173. Moore HC. Impact on quality of life of adjuvant therapy for breast cancer. *Curr Oncol Rep* 9: 42-46, 2007.
174. Partridge AH, Burstein HJ, Winer EP. Side effects of chemotherapy and combined chemohormonal therapy in women with early-stage breast cancer. *J Natl Cancer Inst Monogr* : 135-142, 2001.
175. Cella DF, Tulskey DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol* 11: 570-579, 1993.
176. Morrow GR, Chiarello RJ, Derogatis LR. A new scale for assessing patients' psychosocial adjustment to medical illness. *Psychol Med* 8: 605-610, 1978.
177. Cella D, Peterman A, Passik S, Jacobsen P, Breitbart W. Progress toward guidelines for the management of fatigue. *Oncology (Williston Park)* 12: 369-377, 1998.
178. Cella D, Davis K, Breitbart W, Curt G. Cancer-related fatigue: prevalence of proposed diagnostic criteria in a United States sample of cancer survivors. *J Clin Oncol* 19: 3385-3391, 2001.
179. Adamsen L, Quist M, Midtgaard J, Andersen C, Moller T, Knutsen L, et al. The effect of a multidimensional exercise intervention on physical capacity, well-being and quality of life in cancer patients undergoing chemotherapy. *Support Care Cancer* 14: 116-127, 2006.
180. Markes M, Brockow T, Resch KL. Exercise for women receiving adjuvant therapy for breast cancer. *Cochrane Database Syst Rev* : CD005001, 2006.
181. Oo TH, Hesketh PJ. Drug insight: New antiemetics in the management of chemotherapy-induced nausea and vomiting. *Nat Clin Pract Oncol* 2: 196-201, 2005.
182. Aapro M. 5-HT<sub>3</sub>-receptor antagonists in the management of nausea and vomiting in cancer and cancer treatment. *Oncology* 69: 97-109, 2005.
183. Hesketh PJ. New treatment options for chemotherapy-induced nausea and vomiting. *Support Care Cancer* 12: 550-554, 2004.
184. Warr DG, Hesketh PJ, Gralla RJ, Muss HB, Herrstedt J, Eisenberg PD, et al. Efficacy and tolerability of aprepitant for the prevention of chemotherapy-induced nausea and vomiting in patients with breast cancer after moderately emetogenic chemotherapy. *J Clin Oncol* 23: 2822-2830, 2005.
185. Herrstedt J, Aapro MS, Roila F, Kataja VV. ESMO Minimum Clinical Recommendations for prophylaxis of chemotherapy-induced nausea and vomiting (NV). *Ann Oncol* 16 Suppl 1: i77-i79, 2005.
186. Bloechl-Daum B, Deuson RR, Mavros P, Hansen M, Herrstedt J. Delayed nausea and vomiting continue to reduce patients' quality of life after highly and moderately emetogenic chemotherapy despite antiemetic treatment. *J Clin Oncol* 24: 4472-4478, 2006.
187. Knobf MT. The influence of endocrine effects of adjuvant therapy on quality of life outcomes in younger breast cancer survivors. *Oncologist* 11: 96-110, 2006.
188. Tramer MR, Reynolds DJ, Stoner NS, Moore RA, McQuay HJ. Efficacy of 5-HT<sub>3</sub> receptor antagonists in radiotherapy-induced nausea and vomiting: a quantitative systematic review. *Eur J Cancer* 34: 1836-1844, 1998.
189. Love RR. Tamoxifen therapy in primary breast cancer: biology, efficacy, and side effects. *J Clin Oncol* 7(6): 803-815, 1989.
190. Cassileth BR, Knuiman MW, Abeloff MD, Falkson G, Ezdinli EZ, Mehta CR. Anxiety levels in patients randomized to adjuvant therapy versus observation for early breast cancer. *J Clin Oncol* 4: 972-974, 1986.
191. Goodwin PJ. Weight gain in early-stage breast cancer: where do we go from here? *J Clin Oncol* 19: 2367-2369, 2001.

192. Harvie MN, Campbell IT, Baildam A, Howell A. Energy balance in early breast cancer patients receiving adjuvant chemotherapy. *Breast Cancer Res Treat* 83: 201-210, 2004.
193. Demark-Wahnefried W, Rimer BK, Winer EP. Weight gain in women diagnosed with breast cancer. *Journal of the American Dietetic Association* 97: 519-526, 1997.
194. Demark-Wahnefried W, Peterson BL, Winer EP, Marks L, Aziz N, Marcom PK, et al. Changes in weight, body composition, and factors influencing energy balance among premenopausal breast cancer patients receiving adjuvant chemotherapy. *J Clin Oncol* 19: 2381-2389, 2001.
195. Kumar N, Allen KA, Riccardi D, Bercu BB, Cantor A, Minton S, et al. Fatigue, weight gain, lethargy and amenorrhea in breast cancer patients on chemotherapy: is subclinical hypothyroidism the culprit? *Breast Cancer Res Treat* 83: 149-159, 2004.
196. Yoo HJ, Ahn SH, Kim SB, Kim WK, Han OS. Efficacy of progressive muscle relaxation training and guided imagery in reducing chemotherapy side effects in patients with breast cancer and in improving their quality of life. *Support Care Cancer* 13: 826-833, 2005.
197. Razavi D, Delvaux N, Farvacques C, de Brier F, van Heer C, Kaufman L, et al. Prevention of adjustment disorders and anticipatory nausea secondary to adjuvant chemotherapy: a double-blind, placebo-controlled study assessing the usefulness of alprazolam. *J Clin Oncol* 11: 1384-1390, 1993.
198. Bines J, Oleske DM, Cobleigh MA. Ovarian function in premenopausal women treated with adjuvant chemotherapy for breast cancer. *J Clin Oncol* 14: 1718-1729, 1996.
199. Jonat W, Kaufmann M, Sauerbrei W, Blamey R, Cuzick J, Namer M, et al. Goserelin versus cyclophosphamide, methotrexate, and fluorouracil as adjuvant therapy in premenopausal patients with node-positive breast cancer: The Zoladex Early Breast Cancer Research Association Study. *J Clin Oncol* 20: 4628-4635, 2002.
200. Lindley C, Vasa S, Sawyer WT, Winer EP. Quality of life and preferences for treatment following systemic adjuvant therapy for early-stage breast cancer. *J Clin Oncol* 16: 1380-1387, 1998.
201. Sprangers MA, Groenvold M, Arraras JI, Franklin J, te Velde A, Muller, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *J Clin Oncol* 14: 2756-2768, 1996.
202. Fetting JH, Gray R, Fairclough DL, Smith TJ, Margolin KA, Citron ML, et al. Sixteen-week multidrug regimen versus cyclophosphamide, doxorubicin, and fluorouracil as adjuvant therapy for node-positive, receptor-negative breast cancer: an Intergroup study. *J Clin Oncol* 16: 2382-2391, 1998.
203. Ganz PA, Rowland JH, Desmond K, Meyerowitz BE, Wyatt GE. Life after breast cancer: understanding women's health-related quality of life and sexual functioning. *J Clin Oncol* 16: 501-514, 1998.
204. Tierney AJ, Taylor J, Closs SJ. Knowledge, expectations and experiences of patients receiving chemotherapy for breast cancer. *Scand J Caring Sci* 6: 75-80, 1992.
205. Meyerowitz BE, Watkins IK, Sparks FC. Psychosocial implications of adjuvant chemotherapy. A two-year follow-up. *Cancer* 52: 1541-1545, 1983.
206. Berglund G, Bolund C, Fornander T, Rutqvist LE, Sjöden PO. Late effects of adjuvant chemotherapy and postoperative radiotherapy on quality of life among breast cancer patients. *Eur J Cancer* 27: 1075-1081, 1991.
207. Young-McCaughan S. Sexual functioning in women with breast cancer after treatment with adjuvant therapy. *Cancer Nurs* 19: 308-319, 1996.
208. Ganz PA, Desmond KA, Belin TR, Meyerowitz BE, Rowland JH. Predictors of sexual health in women after a breast cancer diagnosis. *J Clin Oncol* 17: 2371-2380, 1999.
209. Ganz PA, Rowland JH, Meyerowitz BE, Desmond KA. Impact of different adjuvant therapy strategies on quality of life in breast cancer survivors. *Recent Results Cancer Res* 152:396-411: 396-411, 1998.
210. Bower JE, Ganz PA, Desmond KA, Rowland JH, Meyerowitz BE, Belin TR. Fatigue in breast cancer survivors: occurrence, correlates, and impact on quality of life. *J Clin Oncol* 18: 743-753, 2000.
211. Ahles TA, Saykin AJ, Furstenberg CT, Cole B, Mott LA, Titus-Ernstoff L, et al. Quality of life of long-term survivors of breast cancer and lymphoma treated with standard-dose chemotherapy or local therapy. *J Clin Oncol* 23: 4399-4405, 2005.
212. Fan HG, Houede-Tchen N, Yi QL, Chemerynsky I, Downie FP, Sabate K, et al. Fatigue, menopausal symptoms, and cognitive function in women after adjuvant chemotherapy for breast cancer: 1- and 2-year follow-up of a prospective controlled study. *J Clin Oncol* 23: 8025-8032, 2005.
213. Johnsson A, Fornander T, Olsson M, Nystedt M, Johansson H, Rutqvist LE. Factors associated with return to work after breast cancer treatment. *Acta Oncol* 46: 90-96, 2007.
214. Bouknight RR, Bradley CJ, Luo Z. Correlates of return to work for breast cancer survivors. *J Clin Oncol* 24: 345-353, 2006.
215. Bloom JR, Stewart SL, Chang S, Banks PJ. Then and now: quality of life of young breast cancer survivors. *Psychooncology* 13: 147-160, 2004.
216. Joly F, Espie M, Marty M, Heron JF, Henry-Amar M. Long-term quality of life in premenopausal women with node-negative localized breast cancer treated with or without adjuvant chemotherapy. *Br J Cancer* 83: 577-582, 2000.
217. Kroenke CH, Rosner B, Chen WY, Kawachi I, Colditz GA, Holmes MD. Functional impact of breast cancer by age at diagnosis. *J Clin Oncol* 22: 1849-1856, 2004.
218. Peuckmann V, Ekholm O, Rasmussen NK, Moller S, Groenvold M, Christiansen P, et al. Health-related quality of life in long-term breast cancer survivors: Nationwide survey in Denmark. *Breast Cancer Res Treat* 104: 39-46, 2007.
219. Phillips KA, Bernhard J. Adjuvant Breast Cancer Treatment and Cognitive Function: Current Knowledge and Research Directions. *JNCI Cancer Spectrum* 95: 190-197, 2003.
220. Tannock IF, Ahles TA, Ganz PA, van Dam FS. Cognitive impairment associated with chemotherapy for cancer: report of a workshop. *J Clin Oncol* 22: 2233-2239, 2004.
221. Minisini A, Atalay G, Bottomley A, Puglisi F, Piccart M, Biganzoli L. What is the effect of systemic anticancer treatment on cognitive function? *Lancet Oncol* 5: 273-282, 2004.
222. Ahles TA. Do systemic cancer treatments affect cognitive function? *Lancet Oncol* 5: 270-271, 2004.
223. Wieneke MH, Dienst ER. Neuropsychological assessment of cognitive functioning following chemotherapy for breast cancer. *Psycho-Oncology* 4: 61-66, 1995.
224. van Dam FSAM, Schagen SB, Muller MJ, Boogerd W, v.d.Wall E, Fortuyn MED, et al. Impairment of cognitive function in women receiving adjuvant treatment for high-risk breast cancer: High-dose versus standard-dose chemotherapy. *J Natl Cancer Inst* 90: 210-218, 1998.

225. Schagen SB, van Dam FS, Muller MJ, Boogerd W, Lindeboom J, Bruning PF. Cognitive deficits after postoperative adjuvant chemotherapy for breast carcinoma. *Cancer* 85: 640-650, 1999.
226. Cull A, Hay C, Love SB, Mackie M, Smets E, Stewart M. What do cancer patients mean when they complain of concentration and memory problems? *Br J Cancer* 74: 1674-1679, 1996.
227. Shilling V, Jenkins V. Self-reported cognitive problems in women receiving adjuvant therapy for breast cancer. *Eur J Oncol Nurs* 11: 6-15, 2007.
228. Ahles TA, Saykin AJ, Furstenberg CT, Cole B, Mott LA, Skalla K, et al. Neuropsychologic impact of standard-dose systemic chemotherapy in long-term survivors of breast cancer and lymphoma. *J Clin Oncol* 20: 485-493, 2002.
229. Wefel JS, Lenzi R, Theriault R, Buzdar AU, Cruickshank S, Meyers CA. 'Chemobrain' in breast carcinoma?: a prologue. *Cancer* 101: 466-475, 2004.
230. Wefel JS, Lenzi R, Theriault RL, Davis RN, Meyers CA. The cognitive sequelae of standard-dose adjuvant chemotherapy in women with breast carcinoma: results of a prospective, randomized, longitudinal trial. *Cancer* 100: 2292-2299, 2004.
231. Anderson-Hanley C, Sherman ML, Riggs R, Agocha VB, Compas BE. Neuropsychological effects of treatments for adults with cancer: a meta-analysis and review of the literature. *J Int Neuropsychol Soc* 9: 967-982, 2003.
232. Jansen CE, Miaskowski C, Dodd M, Dowling G, Kramer J. A metaanalysis of studies of the effects of cancer chemotherapy on various domains of cognitive function. *Cancer* 104: 2222-2233, 2005.
233. Hurria A, Rosen C, Hudis C, Zuckerman E, Panageas KS, Lachs MS, et al. Cognitive function of older patients receiving adjuvant chemotherapy for breast cancer: a pilot prospective longitudinal study. *J Am Geriatr Soc* 54: 925-931, 2006.
234. Schagen SB, Muller MJ, Boogerd W, Mellenbergh GJ, van Dam FS. Change in cognitive function after chemotherapy: a prospective longitudinal study in breast cancer patients. *J Natl Cancer Inst* 98: 1742-1745, 2006.
235. Jenkins V, Shilling V, Deutsch G, Bloomfield D, Morris R, Allan S, et al. A 3-year prospective study of the effects of adjuvant treatments on cognition in women with early stage breast cancer. *Br J Cancer* 94: 828-834, 2006.
236. Ferguson RJ, Ahles TA, Saykin AJ, McDonald BC, Furstenberg CT, Cole BF, et al. Cognitive-behavioral management of chemotherapy-related cognitive change. *Psychooncology*, 2006.
237. Faulkner A, Maguire P. *Talking to Cancer Patients and Their Relatives*. Oxford: Oxford University Press; 1994.
238. Brennan J, Moynihan C. *Cancer in Context: A Practical Guide to Supportive Care*. Oxford: Oxford University Press; 2004.
239. Thewes B, Meiser B, Rickard J, Friedlander M. The fertility- and menopause-related information needs of younger women with a diagnosis of breast cancer: a qualitative study. *Psychooncology* 12: 500-511, 2003.
240. Bruera E, Willey JS, Palmer JL, Rosales M. Treatment decisions for breast carcinoma: patient preferences and physician perceptions. *Cancer* 94: 2076-2080, 2002.
241. Thewes B, Meiser B, Taylor A, Phillips KA, Pendlebury S, Capp A, et al. Fertility- and menopause-related information needs of younger women with a diagnosis of early breast cancer. *J Clin Oncol* 23: 5155-5165, 2005.
242. Maguire P. Improving communication with cancer patients. *Eur J Cancer* 35: 2058-2065, 1999.
243. Thomas R, Daly M, Perryman B, Stockton D. Forewarned is forearmed--benefits of preparatory information on video cassette for patients receiving chemotherapy or radiotherapy--a randomised controlled trial. *Eur J Cancer* 36: 1536-1543, 2000.
244. Degner LF, Kristjanson LJ, Bowman D, Sloan JA, Carriere KC, O'Neil J, et al. Information needs and decisional preferences in women with breast cancer. *JAMA* 277: 1485-1492, 1997.
245. Bilodeau BA, Degner LF. Information needs, sources of information, and decisional roles in women with breast cancer. *Oncol Nurs Forum* 23: 691-696, 1996.
246. Luker KA, Beaver K, Leinster SJ, Owens RG, Degner LF, Sloan JA. The information needs of women newly diagnosed with breast cancer. *J Adv Nurs* 22: 134-141, 1995.
247. Luker KA, Beaver K, Leinster SJ, Owens RG. Information needs and sources of information for women with breast cancer: a follow-up study. *J Adv Nurs* 23: 487-495, 1996.
248. Jones JA, Avritscher EB, Cooksley CD, Michelet M, Bekele BN, Elting LS. Epidemiology of treatment-associated mucosal injury after treatment with newer regimens for lymphoma, breast, lung, or colorectal cancer. *Support Care Cancer* 14: 505-515, 2006.
249. Mandelblatt J, Armetta C, Yabroff KR, Liang W, Lawrence W. Descriptive review of the literature on breast cancer outcomes: 1990 through 2000. *J Natl Cancer Inst Monogr* : 8-44, 2004.
250. Ejlertsen B, Mouridsen HT, Jensen MB, Bengtsson NO, Bergh J, Cold S, et al. Similar efficacy for ovarian ablation compared with cyclophosphamide, methotrexate, and fluorouracil: from a randomized comparison of premenopausal patients with node-positive, hormone receptor-positive breast cancer. *J Clin Oncol* 24: 4956-4962, 2006.
251. Scottish Cancer Trials Breast Group. Adjuvant ovarian ablation versus CMF chemotherapy in premenopausal women with pathological stage II breast carcinoma: the Scottish trial. *Lancet* 341: 1293-1298, 1993.
252. Boccardo F, Rubagotti A, Amoroso D, Mesiti M, Romeo D, Sismondi P, et al. Cyclophosphamide, methotrexate, and fluorouracil versus tamoxifen plus ovarian suppression as adjuvant treatment of estrogen receptor-positive pre-/perimenopausal breast cancer patients: results of the Italian Breast Cancer Adjuvant Study Group 02 randomized trial. *J Clin Oncol* 18: 2718-2727, 2000.
253. Jakesz R, Hausmaninger H, Kubista E, Gnant M, Menzel C, Bauernhofer T, et al. Randomized adjuvant trial of tamoxifen and goserelin versus cyclophosphamide, methotrexate, and fluorouracil: evidence for the superiority of treatment with endocrine blockade in premenopausal patients with hormone-responsive breast cancer - Austrian Breast and Colorectal Cancer Study Group Trial 5. *J Clin Oncol* 20: 4621-4627, 2002.
254. Castiglione-Gertsch M, O'Neill A, Price KN, Goldhirsch A, Coates AS, Colleoni M, et al. Adjuvant chemotherapy followed by goserelin versus either modality alone for premenopausal lymph node-negative breast cancer: a randomized trial. *J Natl Cancer Inst* 95: 1833-1846, 2003.
255. Schmid P, Untch M, Wallwiener D, Koss V, Bondar G, Vassiljev L, et al. Cyclophosphamide, methotrexate and fluorouracil (CMF) versus hormonal ablation with leuprorelin acetate as adjuvant treatment of node-positive, premenopausal breast cancer patients: preliminary results of the TABLE-study (Takeda Adjuvant Breast cancer study with Leuprorelin Acetate). *Anticancer Res* 22: 2325-2332, 2002.

256. Early Breast Cancer Trialists' Collaborative Group. Poly-chemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 352: 930-942, 1998.
257. Nowak AK, Wilcken NR, Stockler MR, Hamilton A, Ghersi D. Systematic review of taxane-containing versus non-taxane-containing regimens for adjuvant and neoadjuvant treatment of early breast cancer. *Lancet Oncol* 5: 372-380, 2004.
258. Pater JL, Parulekar WR. Ovarian ablation as adjuvant therapy for premenopausal women with breast cancer--another step forward. *J Natl Cancer Inst* 95: 1811-1812, 2003.
259. Wolff AC, Davidson NE. Still waiting after 110 years: the optimal use of ovarian ablation as adjuvant therapy for breast cancer. *J Clin Oncol* 24: 4949-4951, 2006.
260. Krop IE, Winer EP. Ovarian suppression for breast cancer: an effective treatment in search of a home. *J Clin Oncol* 23: 5869-5872, 2005.
261. Dellapasqua S, Colleoni M, Gelber RD, Goldhirsch A. Adjuvant endocrine therapy for premenopausal women with early breast cancer. *J Clin Oncol* 23: 1736-1750, 2005.
262. Bernhard J, Zahrieh D, Castiglione-Gertsch M, Hurny C, Gelber RD, Forbes JF, et al. Adjuvant chemotherapy followed by goserelin compared with either modality alone: the impact on amenorrhea, hot flashes, and quality of life in premenopausal patients--the International Breast Cancer Study Group Trial VIII. *J Clin Oncol* 25: 263-270, 2007.
263. Pritchard KI. Adjuvant therapy for breast cancer: we are solving the puzzle, but pieces are still missing. *J Natl Cancer Inst* 99: 494-495, 2007.
264. de Haes H, Olschewski M, Kaufmann M, Schumacher M, Jonat W, Sauerbrei W. Quality of life in goserelin-treated versus cyclophosphamide + methotrexate + fluorouracil-treated premenopausal and perimenopausal patients with node-positive, early breast cancer: the Zoladex Early Breast Cancer Research Association Trialists Group. *J Clin Oncol* 21: 4510-4516, 2003.
265. Elkin EB, Weinstein MC, Kuntz KM, Bunnell CA, Weeks JC. Adjuvant ovarian suppression versus chemotherapy for premenopausal, hormone-responsive breast cancer: quality of life and efficacy tradeoffs. *Breast Cancer Res Treat* 93: 25-34, 2005.
266. Fallowfield L, McGurk R, Dixon M. Same gain, less pain: potential patient preferences for adjuvant treatment in premenopausal women with early breast cancer. *Eur J Cancer* 40: 2403-2410, 2004.
267. Goodwin PJ. Reversible ovarian ablation or chemotherapy: are we ready for quality of life to guide adjuvant treatment decisions in breast cancer? *J Clin Oncol* 21: 4474-4475, 2003.
268. Goodwin PJ, Black JT, Bordeleau LJ, Ganz PA. Health-related quality-of-life measurement in randomized clinical trials in breast cancer--taking stock. *J Natl Cancer Inst* 95: 263-281, 2003.
269. Coates A, Thomson D, McLeod GR, Hersey P, Gill PG, Olver IN, et al. Prognostic value of quality of life scores in a trial of chemotherapy with or without interferon in patients with metastatic malignant melanoma. *Eur J Cancer* 29A: 1731-1734, 1993.
270. Kaasa S, Mastekaasa A, Lund E. Prognostic factors for patients with inoperable non-small cell lung cancer, limited disease. The importance of patients' subjective experience of disease and psychosocial well-being. *Radiother Oncol* 15: 235-242, 1989.
271. Ganz PA, Lee JJ, Siau J. Quality of life assessment. An independent prognostic variable for survival in lung cancer. *Cancer* 67: 3131-3135, 1991.
272. Efficace F, Bottomley A, Coens C, van Steen K, Conroy T, Schoffski P, et al. Does a patient's self-reported health-related quality of life predict survival beyond key biomedical data in advanced colorectal cancer? *Eur J Cancer* 42: 42-49, 2006.
273. Dharma-Wardene M, Au HJ, Hanson J, Dupere D, Hewitt J, Feeny D. Baseline FACT-G score is a predictor of survival for advanced lung cancer. *Qual Life Res* 13: 1209-1216, 2004.
274. Roychowdhury DF, Hayden A, Liepa AM. Health-related quality-of-life parameters as independent prognostic factors in advanced or metastatic bladder cancer. *J Clin Oncol* 21: 673-678, 2003.
275. McCarter H, Furlong W, Whitton AC, Feeny D, DePauw S, Willan AR, et al. Health status measurements at diagnosis as predictors of survival among adults with brain tumors. *J Clin Oncol* 24: 3636-3643, 2006.
276. Sullivan PW, Nelson JB, Mulani PM, Sleep D. Quality of life as a potential predictor for morbidity and mortality in patients with metastatic hormone-refractory prostate cancer. *Qual Life Res* 15: 1297-1306, 2006.
277. Stockler MR, Osoba D, Corey P, Goodwin PJ, Tannock IF. Convergent discriminative, and predictive validity of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) assessment and comparison with analogous scales from the EORTC QLQ-C30 and a trial-specific module. European Organisation for Research and Treatment of Cancer. *Core Quality of Life Questionnaire*. *J Clin Epidemiol* 52: 653-666, 1999.
278. Maisey NR, Norman A, Watson M, Allen MJ, Hill ME, Cunningham D. Baseline quality of life predicts survival in patients with advanced colorectal cancer. *Eur J Cancer* 38: 1351-1357, 2002.
279. Coates AS, Hurny C, Peterson HF, Bernhard J, Castiglione-Gertsch M, Gelber RD, et al. Quality-of-life scores predict outcome in metastatic but not early breast cancer. International Breast Cancer Study Group. *J Clin Oncol* 18: 3768-3774, 2000.
280. Efficace F, Biganzoli L, Piccart M, Coens C, van Steen K, Cufer T, et al. Baseline health-related quality-of-life data as prognostic factors in a phase III multicentre study of women with metastatic breast cancer. *Eur J Cancer* 40: 1021-1030, 2004.
281. Kramer JA, Curran D, Piccart M, de Haes JC, Bruning P, Klijn J, et al. Identification and interpretation of clinical and quality of life prognostic factors for survival and response to treatment in first-line chemotherapy in advanced breast cancer. *Eur J Cancer* 36: 1498-1506, 2000.
282. Luoma ML, Hakamies-Blomqvist L, Sjoström J, Pluzanska A, Ottoson S, Mouridsen H, et al. Prognostic value of quality of life scores for time to progression (TTP) and overall survival time (OS) in advanced breast cancer. *Eur J Cancer* 39: 1370-1376, 2003.
283. Coates A, GebSKI V, Signorini D, Murray P, McNeil D, Byrne M, et al. Prognostic value of quality-of-life scores during chemotherapy for advanced breast cancer. *J Clin Oncol* 10(12): 1833-1838, 1992.
284. Watson M, Haviland JS, Greer S, Davidson J, Bliss JM. Influence of psychological response on survival in breast cancer: a population-based cohort study. *Lancet* 354: 1331-1336, 1999.
285. Watson M, Homewood J, Haviland J, Bliss JM. Influence of psychological response on breast cancer survival: 10-year fol-

- low-up of a population-based cohort. *Eur J Cancer* 41: 1710-1714, 2005.
286. Tross S, Herndon II J, Korzun A, Kornblith AB, Cella DF, Holland JF, et al. Psychological symptoms and disease-free and overall survival in women with stage II breast cancer. *Cancer and Leukemia Group B. J Natl Cancer Inst* 88: 661-667, 1996.
  287. Efficace F, Therasse P, Piccart MJ, Coens C, van Steen K, Welnicka-Jaskiewicz M, et al. Health-related quality of life parameters as prognostic factors in a nonmetastatic breast cancer population: an international multicenter study. *J Clin Oncol* 22: 3381-3388, 2004.
  288. Goodwin PJ, Ennis M, Bordeleau LJ, Pritchard KI, Trudeau ME, Koo J, et al. Health-related quality of life and psychosocial status in breast cancer prognosis: analysis of multiple variables. *J Clin Oncol* 22: 4184-4192, 2004.
  289. Lehto US, Ojanen M, Dyba T, Aromaa A, Kellokumpu-Lehtinen P. Baseline psychosocial predictors of survival in localised breast cancer. *Br J Cancer* 94: 1245-1252, 2006.
  290. Lewis CE, O'Sullivan C, Barraclough J. *The psychoimmunology of cancer - mind and body in the fight for survival*. Oxford: Oxford University Press; 1994.
  291. Ross L, Boesen EH, Dalton SO, Johansen C. Mind and cancer: does psychosocial intervention improve survival and psychological well-being? *Eur J Cancer* 38: 1447-1457, 2002.
  292. Goodwin PJ, Leszcz M, Ennis M, Koopmans J, Vincent L, Guthrie H, et al. The effect of group psychosocial support on survival in metastatic breast cancer. *N Engl J Med* 345: 1719-1726, 2001.
  293. Cunningham AJ, Edmonds CV, Jenkins GP, Pollack H, Lockwood GA, Warr D. A randomized controlled trial of the effects of group psychological therapy on survival in women with metastatic breast cancer. *Psychooncology* 7: 508-517, 1998.
  294. Edelman S, Bell DR, Kidman AD. A group cognitive behaviour therapy programme with metastatic breast cancer patients. *Psychooncology* 8: 295-305, 1999.
  295. McPherson CJ, Addington-Hall JM. Judging the quality of care at the end of life: can proxies provide reliable information? *Soc Sci Med* 56: 95-109, 2003.
  296. Lobchuk MM, Kristjanson L, Degner L, Blood P, Sloan JA. Perceptions of symptom distress in lung cancer patients: I. Congruence between patients and primary family caregivers. *Journal of Pain & Symptom Management* 14: 136-146, 1997.
  297. de Jong N, Kester AD, Schouten HC, Abu-Saad HH, Courtens AM. Course of fatigue between two cycles of adjuvant chemotherapy in breast cancer patients. *Cancer Nurs* 29: 467-477, 2006.
  298. Efficace F, Bottomley A, Osoba D, Gotay C, Flechtner H, D'haese S, et al. Beyond the development of health-related quality-of-life (HRQOL) measures: a checklist for evaluating HRQOL outcomes in cancer clinical trials--does HRQOL evaluation in prostate cancer research inform clinical decision making? *J Clin Oncol* 21: 3502-3511, 2003.
  299. Spilker B. *Quality of life assessments in clinical trials*. New York: Raven Press; 1990.
  300. Spilker B. *Quality of Life and Pharmacoeconomics in Clinical Trials*. 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1996.
  301. McDowell I, Newell C. *Measuring Health. A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press; 1987.
  302. Walker SR and Rosser RM. *Quality of Life Assessment: Key Issues in the 1990s*. Dordrecht: Kluwer Academic Publishers; 1993.
  303. Osoba D. *Effect of cancer on quality of life*. Boca Raton: CRC Press, Inc.; 1991.
  304. Lenderking WR and Revicki D. *Advancing Health Outcomes Research: Methods and Clinical Applications*. McLean, VA: Degnon Associates; 2005.
  305. Carr AJ, Higginson IJ, and Robinson PG. *Quality of Life*. London: BMJ Books; 2003.
  306. Lipscomb J, Gotay CC, and Snyder C. *Outcomes Assessment in Cancer: Measures, Methods, and Applications*. Cambridge: Cambridge University Press; 2005.
  307. Fayers PM and Hays R. *Assessing Quality of Life in Clinical Trials - Methods and Practice*. 2nd ed. Oxford: Oxford University Press; 2005.
  308. Staquet M, Berzon R, Osoba D, Machin D. Guidelines for reporting results of quality of life assessments in clinical trials. *Qual Life Res* 5: 496-502, 1996.
  309. Fayers PM, Hopwood P, Harvey P, Girling DJ, Machin D, Stephens R. Quality of life assessment in clinical trials - guidelines and a checklist for protocol writers: The U.K. medical research council experience. *Eur J Cancer* 33: 20-28, 1997.
  310. U.S. Department of Health and Human Services Food and Drug Administration. *Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Draft Guidance*. (<http://www.fda.gov/cder/guidance/5460dft.pdf>). Rockville, MD: U.S. Department of Health and Human Services Food and Drug Administration; 2006.
  311. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol* 51: 1115-1128, 1998.
  312. Joyce CRB, McGee HM, O'Boyle CA. *Individual Quality of Life. Approaches to Conceptualisation and Assessment*. Amsterdam: Harwood Academic Publishers; 1999.
  313. Grønvd M, Pedersen C, Jensen CR, Johnsen AT. *Kræftpatientens verden - en undersøgelse af, hvad danske kræftpatienter har brug for*. København: Kræftens Bekæmpelse; 2006.
  314. Johnsen AT, Jensen CR, Pedersen C, Grønvd M. *Kræftpatientens verden - en undersøgelse af, hvilke problemer danske kræftpatienter oplever*. København: Kræftens Bekæmpelse; 2006.
  315. Lenderking WR and Revicki D. *Advancing Health Outcomes Research: Methods and Clinical Applications*. McLean, VA: Degnon Associates; 2005.
  316. Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials*. Boca Raton: CRC Press; 2002.
  317. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 16: 139-144, 1998.
  318. Schmoor C, Schumacher M. Effects of covariate omission and categorization when analysing randomized trials with the Cox model. *Stat Med* 16: 225-237, 1997.
  319. Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the functional assessment of cancer therapy-breast quality-of-life instrument. *J Clin Oncol* 15: 974-986, 1997.
  320. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd edition. New York: McGraw-Hill, Inc.; 1994.
  321. McHorney CA, Cook KF. The ten Ds of health outcomes measurement for the twenty-first century. In: Lipscomb J,

- Gotay CC, and Snyder C. Outcomes Assessment in Cancer: Measures, Methods, and Applications. Pp. 590-609. Cambridge: Cambridge University Press; 2005.
322. Lydick E, Epstein RS. Interpretation of quality of life changes. *Qual Life Res* 2: 221-226, 1993.
323. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 27: S178-S189, 1989.
324. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 5: 555-567, 1996.
325. Yost KJ, Cella D, Chawla A, Holmgren E, Eton DT, Ayanian JZ, et al. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches. *J Clin Epidemiol* 58: 1241-1251, 2005.
326. Sloan JA, Cella D, Hays RD. Clinical significance of patient-reported questionnaire data: another step toward consensus. *J Clin Epidemiol* 58: 1217-1219, 2005.
327. Sloan JA, Frost MH, Berzon R, Dueck A, Guyatt G, Moynour C, et al. The clinical significance of quality of life assessments in oncology: a summary for clinicians. *Support Care Cancer* 14: 988-998, 2006.
328. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res* 6: 595-600, 1997.
329. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* , 2007.
330. Hahn EA, Cella D, Bode RK, Gershon R, Lai JS. Item banks and their potential applications to health status assessment in diverse populations. *Med Care* 44: S189-S197, 2006.
331. Lai JS, Dineen K, Reeve BB, Von RJ, Shervin D, McGuire M, et al. An item response theory-based pain item bank can enhance measurement precision. *J Pain Symptom Manage* 30: 278-288, 2005.
332. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 45: S22-S31, 2007.
333. Bjorner JB, Petersen MA, Groenvold M, Aaronson N, Ahlner-Elmqvist M, Arraras JI, et al. Use of item response theory to develop a shortened version of the EORTC QLQ-C30 emotional functioning scale. *Qual Life Res* 13: 1683-1697, 2004.
334. Petersen MA, Groenvold M, Aaronson N, Blazeby J, Brandberg Y, De Graeff A, et al. Item response theory was used to shorten EORTC QLQ-C30 scales for use in palliative care. *J Clin Epidemiol* 59: 36-44, 2006.
335. Groenvold M, Petersen MA, Aaronson NK, Arraras JI, Blazeby JM, Bottomley A, et al. The development of the EORTC QLQ-C15-PAL: A shortened questionnaire for cancer patients in palliative care. *Eur J Cancer* 42: 55-64, 2006.
336. Petersen MA, Groenvold M, Aaronson N, Fayers P, Sprangers M, Bjorner JB. Multidimensional computerized adaptive testing of the EORTC QLQ-C30: basic developments and evaluations. *Qual Life Res* 15: 315-329, 2006.
337. Thissen D, Reeve BB, Bjorner JB, Chang CH. Methodological issues for building item banks and computerized adaptive scales. *Qual Life Res* 16 Suppl 1: 109-119, 2007.
338. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. *Qual Life Res* 16 Suppl 1: 95-108, 2007.
339. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 45: S22-S31, 2007.
340. Ganz PA, Goodwin PJ. Quality of life in breast cancer - what have we learned and where do we go from here? In: Lipscomb J, Gotay CC, and Snyder C. Outcomes Assessment in Cancer: Measures, Methods, and Applications. Pp. 93-125. Cambridge: Cambridge University Press; 2005.
341. Levine MN, Ganz PA. Beyond the development of quality-of-life instruments: where do we go from here? *J Clin Oncol* 20: 2215-2216, 2002.
342. Levine MN, Gafni A, Markham B, Macfarlane D. A bedside decision instrument to elicit a patient's preference concerning adjuvant chemotherapy for breast cancer. *Ann Intern Med* 117: 53-58, 1992.
343. Baum M, Ravdin PM. Decision-making in early breast cancer: guidelines and decision tools. *Eur J Cancer* 38: 745-749, 2002.
344. Whelan TJ, Loprinzi C. Physician/patient decision aids for adjuvant therapy. *J Clin Oncol* 23: 1627-1630, 2005.
345. Wallberg B, Michelson H, Nystedt M, Bolund C, Degner LF, Wilking N. Information needs and preferences for participation in treatment decisions among Swedish breast cancer patients. *Acta Oncol* 39: 467-476, 2000.
346. Velikova G, Brown JM, Smith AB, Selby PJ. Computer-based quality of life questionnaires may contribute to doctor-patient interactions in oncology. *Br J Cancer* 86: 51-59, 2002.
347. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 22: 714-724, 2004.
348. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Role of health-related quality of life in palliative chemotherapy treatment decisions. *J Clin Oncol* 20: 1056-1062, 2002.
349. Detmar SB, Aaronson NK. Quality of life assessment in daily clinical oncology practice: a feasibility study. *Eur J Cancer* 34: 1181-1186, 1998.
350. Detmar SB, Aaronson NK, Wever LD, Muller M, Schornagel JH. How are you feeling? Who wants to know? Patients' and oncologists' preferences for discussing health-related quality-of-life issues. *J Clin Oncol* 18: 3295-3301, 2000.
351. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA* 288: 3027-3034, 2002.
352. Rosenbloom SK, Victorson DE, Hahn EA, Peterman AH, Cella D. Assessment is not enough: a randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice. *Psychooncology* 16: 1069-79, 2007.

## APPENDIX A. DIF ANALYSES

This appendix contains results of DIF analyses of the multi-item scales of the EORTC QLQ-C30 and the HAD Scale that were removed from papers VI-VIII in order to shorten the papers.

### Materials and methods

The data are described in papers VI-VIII. DIF analyses are described in paper II and section 3.7.

### Results concerning anxiety and depression in breast cancer patients compared to the general population (paper VI)

The DIF analyses of the two HAD subscales in relation to group (general population or patient) and age showed no evidence of DIF in the anxiety scale. In contrast, there was DIF in relation to group in the depression scale mainly for item 10, 'I have lost interest in my appearance': at a given level of depression, there were much higher scores on this item in the general population than in the breast cancer patients (partial gamma = 0.56,  $p < 0.0001$ ). Consequently, this item influenced comparisons between the two groups: depression scores in the general population sample were biased upward. When comparing the groups by means of the remaining 6 items, the magnitude of the difference in mean scores between the two groups (described in section 4.8.2) was diminished, but still significant (ANCOVA,  $p = 0.02$ ). The depression sub-scale also showed DIF with regard to age for 'I feel as if I am slowed down' (item 60) and 'I look forward with enjoyment to things' (item 64). These two items had stronger increments in item score with increasing age than the total depression score (age relation: partial gammas 0.33 and 0.42,  $p < 0.001$ , controlling for scale score).

### Results concerning chemotherapy versus no chemotherapy and chemotherapy versus ovarian ablation (papers VII and VIII)

No significant ( $p < 0.001$ ) DIF was detected in relation to group in either study. Specifically, the three cases of DIF in relation to chemotherapy described in paper II were not significant in the

analyses addressing this comparison in paper VII. However, for one of these cases of DIF, the cognitive function scale, DIF of similar magnitude, but not meeting the  $p < 0.001$  significance criterion, was found in the data reported in papers VII (partial gamma=0.58,  $p=0.011$ ) and VIII (partial gamma=0.47,  $p=0.004$ ). Separate analyses of the two items showed that chemotherapy had a relatively strong effect on concentration, whereas it had a much weaker effect on memory.

## APPENDIX B. VALIDATION OF HAD SCALE AND DBCG 89 QUESTIONNAIRE

### Materials and methods

The data are from the same study as paper III although the participants were the breast cancer patients only. The methodology was summarised in section 3.8 (described in full in paper III).

### Results

Please see Tables B1-4.

Table B1

Agreement between patients and the observer about responses to the items of the Hospital Anxiety and Depression Scale. Item scores for patients and observer, overall agreement (overall), and weighted kappa (kappa) with standard error (SE).

Item	Mean score		Agreement		
	Patients	Observer	Overall	Kappa	(SE)
1) I feel tense or wound up	0.83	0.78	.83	.92	(.03)
2) I still enjoy the things I used to enjoy	.41	.41	.87	.88	(.07)
3) I get a sort of frightened feeling as if something awful is about to happen	1.26	1.13	.69	.77	(.07)
4) I can laugh and see the funny side of things	.28	.37	.87	.70	(.10)
5) Worrying thoughts go through my mind	.80	.93	.67	.77	(.07)
6) I feel cheerful	.41	.50	.80	.72	(.12)
7) I can sit at ease and feel relaxed	.76	.67	.61	.71	(.11)
8) I feel as if I am slowed down	.87	**1.09	.70	.82	(.05)
9) I get a sort of frightened feeling like 'butterflies' in the stomach	.76	.63	.80	.84	(.05)
10) I have lost interest in my appearance	.36	.22	.84	.81	(.09)
11) I feel restless as if I have to be on the move	.65	** .41	.72	.69	(.13)
12) I look forward with enjoyment to things	.39	.54	.80	.65	(.09)
13) I get sudden feelings of panic	.67	.60	.78	.84	(.06)
14) I can enjoy a good book or radio or TV programme	.46	.46	.91	.95	(.03)

\*) Item numbers are those of the HAD Scale, not those used in this study.

\*\*) Scores of patients and observer significantly different ( $p = .04$  and  $p = .02$ , respectively).

Table B2

Qualitative comments to the HAD scale items.  
Quotes from patients are in brackets.

**2) I still enjoy the things I used to enjoy**

Misses a response option 'more'. Was not in doubt how to answer.  
Might consider not answering.

**3) I get a sort of frightened feeling as if something awful is about to happen**

This question is generally interpreted in relation to the breast cancer. Is considered a question about fear of recurrence or death following cancer.

'Yes, I know what 'it' can imply, but no, it is quite rare that I think of it'. The rater was in doubt how to rate this; none of the options seems suitable.

'Yes, but I am not thinking of that all of the time.' It is difficult to rate this response adequately: It seems wrong to use one of the modifications (e.g. 'Yes, but not too badly') but on the other hand I find it difficult to use the most extreme category for this.

**5) Worrying thoughts go through my mind**

This question is generally interpreted in relation to the breast cancer. 'I perceive options 3 and 4 as identical.'

Rater: difficult to discriminate between 1 and 2.

**8) I feel as if I am slowed down**

'Yes, due to age'. Everything is fine, perceives it as natural for a 70-year-old to be 'slowed down'.

Rater: The answer does not sound as being correlated to depression but rather to tiredness/age.

**9) I get a sort of frightened feeling like 'butterflies' in the stomach**

The patients recounts that originally she said 'yes' to this question because she had been tense but in the interview she reaches the conclusion that this should be a 'no' because it does not take the form of 'butterflies'.

**10) I have lost interest in my appearance**

'Yes, it's not the same as it once was when you are 68 but I try to be clean and tidy. The operation has not meant anything.' Rater: in doubt how to rate this but chose 4, corresponding to no symptomatology.

**11) I feel restless as if I have to be on the move**

'Oh no, during the last week it's been Easter so it has been less than usual because we have had visitors.' It sounds as if she had originally answered more broadly.

**12) I look forward with enjoyment to things**

More than before.

**14) I can enjoy a good book or radio or TV programme**

Could, but did not have the time.

Is in doubt as to whether the question concerns the frequency of enjoying a book/TV or whether you do enjoy it (every time).

Table B4

Qualitative comments to the DBCG 89 Questionnaire items.  
Quotes from patients are in brackets.

**1) Has your mouth or your tongue been sore?**

Because the dental prosthesis is chafing.

Answers 'a little' because she has had pain in the throat only.

**4) How satisfied have you been with your appearance?**

Had previously stated that she did not like to look at the scar. Here she first answers 'satisfied' but when asked whether this is her overall judgment she answers: 'No, I thought of the face only. If the scar has to be taken into account then 'somewhat dissatisfied'.'

The patient thinks this is a stupid question. If it is meant to be answered by someone who is not dressed then it is 'at the bottom', but with clothes on it is 'good enough'. Did not know which of these two options she should use when replying. At the end of the discussion she agrees with the interviewer to make an overall judgment.

**10) At present, do you have less hair than usual? (NB. Referring to the thickness, not the length)**

'Just as thick but only 1-1½ cm. It's growing out now.' Wears a wig.

During the interview the patient realises that it is in the past week vs. 'normally'. Answered 'yes' originally because the hair has become thinner with age.

**12) Have you had desire for sexual intercourse?**

Is unchanged compared to before. Thinks it is 'normal'. Will not call this 'Very much'.

Answers 'A little' in a way that it sounds very positive, as something totally OK. This is confirmed by the replies to the next questions where the answer in 13 is '5-6 times' and in (the deleted item on satisfaction) 'very satisfied'. This pattern has been heard a number of times: 'Very satisfied' with 'a little' desire. In other words, in these cases 'a little' cannot be interpreted negatively.

**13) How frequently have you had sexual intercourse?**

'What is sexual intercourse?' The patient chose a very strict definition when originally answering but was in doubt.

**17) How many hours a week have you worked outside your home during the last month?**

The patient takes care of other people's children at home (child minder): what is she to answer? Has 49 regular working hours per week.

Table B3

Agreement between patients and the observer about responses to the items of the DBCG 89 Questionnaire. Item scores for patients and observer, overall agreement (overall), and weighted kappa (kappa) with standard error (SE).

Item	Mean score		Agreement		
	Pa-tients	Observer	Overall	Kappa	(SE)
1. Has your mouth or your tongue been sore?	14	* 21	0.80	.77	(.06)
2. Have you had hot flushes and/or sweats?	44	49	0.78	.92	(.02)
3. Have you had nausea when thinking of your treatment?	18	20	0.98	.93	(.07)
4. How satisfied have you been with your appearance?	68	66	0.48	0.51	(.13)
5. How has your energy been compared to before your breast operation? (1)	3.3	3.2	0.71	0.83	(.06)
6. Do you have regular menstrual bleedings?	15	13	0.98	0.91	(.08)
7. Have you had any menstrual bleedings within the last 12 months?	61	61	1.00	1.00	(.00)
8. Do you have any difficulties holding your water (urine)?	15	16	0.98	0.97	(.03)
9. Has your weight changed since the breast operation? (2)	34	33	0.93	0.85	(.09)
10. At present, do you have less hair than usual? (NB. Referring to the thickness, not the length.)	19	17	0.91	0.94	(.03)
11. These days, do you wear a wig?	11	11	1.00	1.00	(.00)
12. Have you had desire for sexual intercourse?	44	* 38	0.71	0.82	(.04)
13. How frequently have you had sexual intercourse? (3)	1.39	1.47	0.85	0.93	(.03)
14. Have you been bothered by vaginal dryness?	14	15	0.84	0.85	(.05)
15. Have you been bothered by vaginal discharge?	11	14	0.91	0.88	(.05)
16. Do you have paid employment?	76	76	1.00	1.00	(.00)
17. How many hours a week have you worked outside your home during the last month?	20	19	0.76	-	-

1 Untransformed score; 1='Much less', 4='Unchanged', 7='Much greater'

2 Per cent reporting weight gain.

3 Transformed score; 0='Have not had sexual intercourse', 4='7 times or more'

\*) Scores of patients and observer significantly different ( $p = .03$  and  $p = .02$ , respectively).

#### APPENDIX C. HYPOTHESES CONCERNING PAPER VIII

Table C1

Staff members' (N=36) expectations concerning the comparison of chemotherapy versus ovarian ablation. For each issue, we asked 'Which group does - all things being equal - have the symptom/problem to the largest extent?' The issues are ranked according to the number of respondents expecting more symptomatology in patients receiving chemotherapy.

Issue	Chemotherapy	No difference (2)	Ovarian ablation
Nausea and vomiting	28	8	0
Hair loss	23	13	0
Fatigue	22	12	2
Reduced ability to work/do household jobs	14	21	0
Overall physical condition/quality of life impaired (1)	14	20	2
Financial difficulties due to disease/treatment	10	26	0
Interference with family life/social activities	10	24	2
Reduced ability to concentrate and remember things	10	24	2
Weight gain	10	19	6
Reduced 'physical function' (being able to take a walk, take care of oneself) (1)	9	27	0
Reduced 'emotional function' (anxiety, depression, etc.)	7	25	3
Impaired sexual life	6	25	4
Pain	2	33	0
Dissatisfaction with appearance	2	31	2
Hot flushes/sweats	1	11	23
Irregular bleedings/menostasia	1	11	23
Dyspnea	0	36	0

1 Separate issues covered each of the two items in the scale. The figure is the average of the responses to these two issues.

2 'No difference' means that no difference between treatments was expected.

**APPENDIX D. GENERAL POPULATION DATA FOR THE DBCG 89 QUESTIONNAIRE**

**Table D1**

Age-stratified mean values (SD) for the DBCG 89 Questionnaire in the general population sample. Only the 12 items that were meaningful in a non-cancer population were applied.

	30-39 yrs. (N=145)	40-49 yrs. (N=138)	50-59 yrs. (N=141)	60-69 yrs. (N=116)	70-75 yrs. (N=68)	All (N=608)
Sore mouth	4 (14)	5 (17)	11 (24)	9 (22)	6 (14)	7 (19)
Hot flushes	15 (27)	20 (28)	31 (32)	22 (27)	12 (25)	21 (29)
Satisfied appearance	64 (24)	62 (26)	60 (25)	61 (28)	62 (23)	62 (26)
Regular bleedings	85 (35)	77 (42)	16 (37)	2 (13)	3 (17)	43 (49)
Bleedings	94 (23)	84 (37)	28 (45)	3 (16)	2 (12)	49 (50)
Incontinence	9 (17)	11 (19)	17 (25)	16 (23)	18 (27)	14 (22)
Sexual interest	54 (27)	56 (28)	38 (28)	23 (27)	14 (24)	41 (31)
Had sex last month	82 (39)	81 (39)	63 (48)	39 (49)	19 (40)	63 (48)
Vaginal dryness	5 (16)	5 (15)	16 (28)	21 (34)	9 (20)	11 (24)
Vaginal discharge	22 (25)	12 (19)	8 (18)	6 (17)	6 (17)	12 (21)
Employed full time	49 (50)	49 (50)	37 (49)	11 (31)	0 (0)	34 (47)
Employed	74 (44)	77 (42)	71 (46)	18 (38)	5 (21)	56 (50)