The Prostate 38:100–109 (1999)

Quality-Adjusted Survival (Q-TWiST) Analysis of EORTC Trial 30853: Comparing Goserelin Acetate and Flutamide With Bilateral Orchiectomy in Patients With Metastatic Prostate Cancer

Ingvar Rosendahl,[1]* Gwendoline M. Kiebert,[1] Desmond Curran,[1] Bernhard F. Cole,[2] Jane C. Weeks,[2] Louis J. Denis,[3] and Reginald R. Hall[4]

[1]*Data Center, European Organization for Research and Treatment of Cancer, Brussels, Belgium*

[2]*Dana-Farber Cancer Institute, Boston, Massachusetts*

[3]*Department of Urology, Academic Hospital Middelheim, Antwerp, Belgium*

[4]*Northern Cancer Network, Freeman Hospital, Newcastle-upon-Tyne, United Kingdom*

BACKGROUND. The first data analysis of the European Organization for Research and Treatment of Cancer (EORTC) 30853 trial indicated a significantly longer time to progression and duration of survival for the maximal androgen blockade (MAB) treatment arm. However, the MAB treatment arm had a higher frequency of reported side effects.

METHODS. The quality-adjusted survival (Q-TWiST) method was applied to perform a secondary analysis of the EORTC 30853 trial in order to obtain a quality-adjusted survival (QAS) analysis. Two models with different definitions of the progression health state were used for the analysis. In the first model, progression was defined by both objective and subjective criteria, and in the second model only by increase in pain score. The approach was also extended to include an analysis using actual utility scores (Q-tility) of patients in the relevant health states.

RESULTS. Based on Q-tility scores obtained from a separate study of a cohort of prostate cancer patients, the QAS analysis resulted in a 5.2-month difference (95% CI, -1.1; 11.5 months) in favor of zoladex and flutamide, equal in magnitude to the benefit found in the unadjusted survival analysis.

CONCLUSIONS. A QAS analysis such as the Q-TWiST method may be preferred over the unadjusted approach in clinical trials where the health states are clearly distinct, and differ significantly in either duration or quality of life (QOL), or both. The second model, with progression defined as increase in pain score, made no difference to the results in this study because of the small difference in duration of the pain-progression health state between treatment arms. However, Q-tility scores from the separate cross-sectional study that was

Grant sponsor: EORTC Lady Grierson Research Fellowship; Grant sponsor: American Cancer Society; Grant number: PBR-53.

Gwendoline M. Kiebert is now at MEDTAB International Inc., London, UK.

Bernhard F. Cole is now at Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire.

*Correspondence to: Ingvar Rosendahl, now at the Oncologic Center, Karolinska Hospital, M8:01, S-171 76 Stockholm, Sweden. E-mail: ingvar@onkc.ks.se

Received 28 January 1998; Accepted 26 June 1998

used in this Q-TWiST analysis showed that a subjective definition of health states better reflects differences in QOL between the health states that the patients experience during follow-up. *Prostate 38:100–109, 1999.* © 1999 Wiley-Liss, Inc.

KEY WORDS: **clinical trials; quality of life; utility; Q-TWiST; prostate cancer; secondary data analysis**

INTRODUCTION

Prostate cancer is the first or second most common malignancy in men in industrialized countries [1]. In the United States it is the most frequent cancer diagnosed among men, and is second only to lung cancer as a cause of cancer death [2]. Mortality and particularly incidence rates from prostate cancer have risen steadily during the last decades [3]. The main reason for this increase is that the population has become older [1]. Prostate cancer is a disease mostly observed in elderly men, with 73% of new cases and 80% of prostate cancer deaths occurring in men over age 70 years. Increases in incidence rates are also attributable to enhanced use of diagnostic procedures, including transurethral prostatectomy for benign prostatic hypertrophy, and prostate-specific antigen testing [4]. While more men are diagnosed in the early stages of the disease due to a broader participation in screening and early detection programs, there are still 20–30% of patients who have advanced disease at the time of diagnosis and who thus have only a 24–36-month median survival [5].

Although early stages of prostatic cancer may be cured by local modalities, the treatment of metastatic disease is much less satisfactory and constitutes a major problem in the management of the disease. Unfortunately, the treatment is always palliative or life-extending and never curative, with the primary aim to improve or maintain a good quality of life (QOL) for the patient [6]. Androgen deprivation is the therapy of choice in advanced carcinoma of the prostate. The options available to achieve androgen elimination include surgical removal of the testes (orchiectomy) and medical drug therapy with a luteinizing hormone-releasing hormone-agonist (LHRH-A). Orchiectomy and LHRH-A therapy are equally effective, with approximately 70% of patients responding [7]. No difference has been found in terms of overall or progression-free survival [8]. Orchiectomy is an inexpensive surgical procedure with a low percentage of adverse events and mortality due to surgery. However, orchiectomy is irreversible and disfiguring [9]. LHRH-A therapy offers the psychological advantage of avoiding surgical castration, but is expensive. The total cost until disease progression has been estimated to be $3,810–$14,050 higher than for orchiectomy [10]. Side effects are the same for orchiectomy and LHRH-A therapy, i.e., lack of libido, impotence, and hot flashes [8,11].

Although castration, either by orchiectomy or LHRH-A, causes a reduction of about 90% in serum testosterone concentrations, androgen biosynthesis in the adrenals is unaffected [9]. Addition of nonsteroidal (pure) antiandrogen to either medical or surgical castration, a combination known as MAB, is used to neutralize the remaining androgens of adrenal origin. The side effects of pure antiandrogens such as flutamide or nilutamide include gastrointestinal problems, hepatotoxicity, hot flashes, and gynecomastia, which in some cases can be painful [8,11]. Antiandrogens are also expensive: the cost of 1 year of flutamide treatment has been estimated to be $3,200 [12].

Usually, treatment comparisons are made by analyzing the different endpoints separately (e.g., survival, progression-free survival). This approach may not be ideal when there are important tradeoffs between different endpoints such as an increased time with side effects of treatment, a longer time to progression, and/or a longer duration of survival in one arm than in the other. In recent years there has been a growing recognition of the need to explore new methods to integrate the classical endpoints of clinical trials with QOL outcomes. One method which enables an integration of these outcomes is the quality-adjusted time without symptoms of disease and toxicity of treatment (Q-TWiST) method. This method adjusts survival data by the QOL that patients experience in the various health states during and following treatment in a cancer clinical trial. The quality adjustment weights can be prospectively obtained from patients in the trial or through cross-sectional surveys of non-trial patients in the health states of interest, or can be assigned an arbitrary value. This method of measuring QOL is different from the classical psychometric approach. The latter provides a description or profile of the various aspects of QOL. However, it does not provide a single ''valuation'' of the patient's overall health state, which is needed to integrate QOL with length of life.

In 1985, the European Organization for Research and Treatment of Cancer (EORTC) started a randomized clinical trial in which a combination of goserelin acetate (zoladex), an LHRH-A depot formulation, and flutamide was compared with bilateral orchiectomy. The objectives of this study were: 1) to compare the

TABLE I. Inclusion and Exclusion Criteria for Patient Selection in the EORTC (30853) Trial

Criteria for inclusion
1. Patients with histologically proven carcinoma of the prostate.
2. Patients in all T, N, and G categories are accepted but must have M1 category disease without previous systemic treatment. Bone metastases may be diagnosed by bone scan and/or X-rays, but questionable metastases should be biopsied. In doubt, call bone scan committee.
3. Patients must have performance status WHO 0–2 with a minimum life expectancy of 3 months.

Criteria for exclusion
1. Patients with previous hormonal and/or chemotherapy. Prior surgery (total prostatectomy or transurethral resection) and radiotherapy are not a cause for exclusion if there is proven progressive metastatic disease outside the field of irradiation.
2. Patients with another neoplasia (except skin, excluding melanoma).
3. Patients with expected difficulties of follow-up related to psychiatric disorders, marked senility, or too large a distance between patient's home and investigator's center.
4. Patients with obvious liver disease (a rise of twice-normal SGOT and SGPT values).
5. Patients over age 80 years.

efficacy and safety of zoladex and flutamide vs. orchiectomy in delaying progression of metastatic carcinoma of the prostate (M1 disease) and in prolonging survival; and 2) to determine the incidence and duration of response for each treatment arm [11]. The ''classical'' data analysis of this trial indicated a significantly longer time to progression and duration of survival for the group of patients who received zoladex and flutamide. Hot flashes and gynecomastia were the side effects reported most frequently by patients in both treatment arms, and occurred at higher rates among patients in the zoladex and flutamide treatment arm.

The aim of this study was to perform a secondary analysis of these data by applying the Q-TWiST method to get a summary measure of the tradeoff between side effects, time to progression, and duration of survival for zoladex and flutamide vs. bilateral orchiectomy. We applied the standard approach of the Q-TWiST method, using ''objective'' clinical data to define the various health states, as well as an approach using a ''subjective'' definition. The source of the QOL valuations of these health states in our study was a separate study of a cohort of prostate cancer patients; however, we also performed a sensitivity analysis for the quality valuations.

MATERIALS AND METHODS

A full description of the design and results of the EORTC 30853 trial can be found elsewhere [11,13]. Inclusion and exclusion criteria for this study are presented in Table I. According to the protocol, treatment was to continue for a minimum of 3 months or until objective progression (Table II). Upon objective progression the patient was to leave the study, and further treatment was at the investigator's discretion. The schedule of clinical evaluations was as follows: pretreatment, subsequently every 4 weeks for 12 weeks, and then every 12 weeks for 48 weeks and every 24 weeks thereafter.

A total of 327 patients was entered (163 to orchiectomy and 164 to MAB) between March 1986–May 1988. For 30 patients, no clinical follow-up form was available (15 patients in each treatment arm). These 30 patients were excluded from the Q-TWiST analysis, since information on progression and side effects was not available. Thus, 297 patients were included in the Q-TWiST analysis.

The Q-TWiST method is a modification made by Goldhirsch et al.[14] of the TWiST method presented by Gelber et al. [15]. It was originally developed to evaluate adjuvant therapies for breast cancer where there is often a tradeoff between toxicity of treatment and delayed recurrence of disease. Recently it has also been applied in studies of other disease settings such as AIDS, rectal cancer, and melanoma [16–18]. For a full description of the method we refer to the paper by Gelber et al. [19].

As stated in the Introduction, the basic principle of the method is to adjust survival in cancer clinical trials by the QOL that patients experience in distinct health states. Each health state is assigned a weight, called a utility score, that corresponds to its value in terms of QOL.

The QOL health states in the study were defined as time with hot flashes as a result of treatment (TOX), time without progression of disease and side effects of treatment (TWiST), and time spent with progression of disease (PROG).

In the primary analysis, the most frequently reported side effects of treatment were hot flashes and gynecomastia. However, only the presence of hot

TABLE II. Criteria of Subjective and Objective Progression for the EORTC (30853) Trial

Subjective criteria of progression
WHO performance status
Pain score
Progression: increase of two categories (from the lowest value), or from 3 to 4 on two successive occasions
Weight
Progression: ≥10% decrease within 1 year
Urology symptoms
Progression: severe symptoms requiring surgical relief or catheterization
Hemoglobin
Progression: decrease of >25% from highest value
Alkaline phosphates
Prostatic acid phosphates
Phosphates
≤1.25 × N*
1.26–2.5 × N
2.6–5 × N
5.1–10 × N
>10 × N
Progression: increase of two categories, or from 3 to 4
Objective criteria of progression
Digital examination of primary tumor
Progression: increase of >50% (from lowest value) of product of largest perpendicular diameters; patients with product of <9 cm^2 at entry were excluded
Regional lymph nodes
Distant lymph nodes
Lung metastases
Liver metastases
Progression: increase of ≥25% of sum of products of largest perpendicular diameters or appearance of new metastases
Bone metastases
Progression: appearance of new hot spots on bone scan or new lesions on X-ray film, or (osteolytic lesions): increase of ≥25% of sum of products of largest perpendicular diameters.

*N, upper limit of normal.

flashes was taken to define the TOX state for two reasons, first of all because these were reported more frequently than gynecomastia. Hot flashes were reported by 64% and gynecomastia by 15% of patients. Most patients who reported gynecomastia also had hot flashes; only 5 patients reported gynecomastia without hot flashes. Secondly, according to the opinion of urologists participating in this study, hot flashes were considered to be more of a burden to patients than gynecomastia. Time with treatment toxicity was calculated by partitioning the follow-up into time intervals according to the schedule of clinical evaluations. The percentage of patients at each time interval with hot flashes and without progression of the disease was plotted as a bar plot under the progression-free survival curve (Figs. 1–4). Some patients did not have follow-up forms during each time interval. To ensure that the proportion of symptoms was not underestimated, we assumed that patients with no follow-up forms had the same proportion of symptoms as the patients for whom follow-up forms were available.

Overall survival (OS) and progression-free survival (PFS) curves were estimated by the Kaplan-Meier product limit method for each treatment arm. The area under a Kaplan-Meier survival curve gives an estimate of mean survival time, and areas between curves represent the estimated mean duration in each clinical health state. The estimates are restricted to an upper limit determined by the median follow-up time and are referred to as restricted means [20]. For this study, the upper limit was restricted to 7 years (median follow-up, 7.2 years).

In this way, the total sum of the areas of the bars represents the mean duration of TOX. The mean duration of TWiST was calculated as TWiST = PFS − TOX, and the mean duration of PROG was calculated as PROG = OS − PFS.

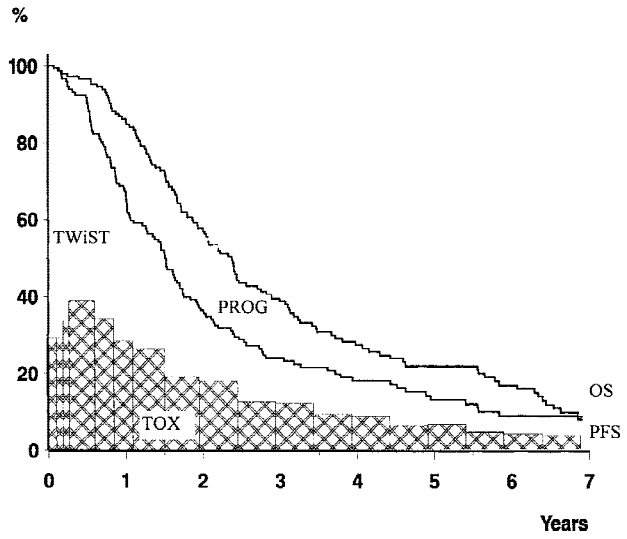

Fig. 1. Mean duration of health states for orchiectomy, model 1.

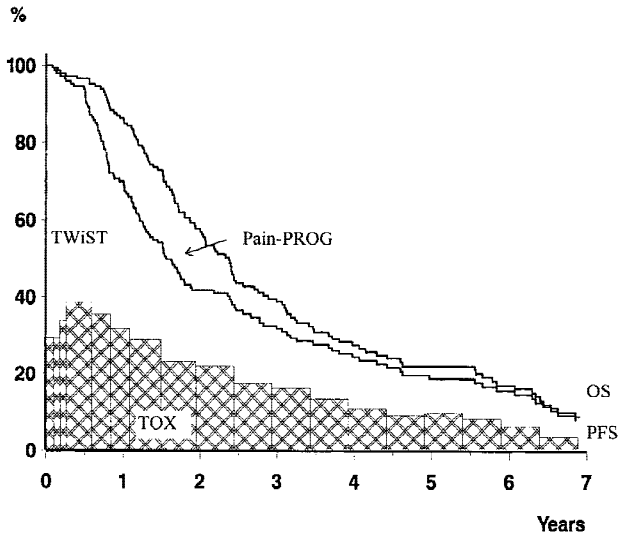

Fig. 3. Mean duration of health states for orchiectomy, model 2.

The Q-TWiST was calculated as a linear combination of the restricted mean duration of each clinical health state and its utility score: Q – TWiST = u_{tox} × TOX + TWiST + u_{prog} × PROG, for each treatment arm, where u_{tox} and u_{prog} are utility scores that reflect the relative value for the TOX and PROG states, respectively. The utility scores are on a scale from zero to one, where zero denotes a health state ''as bad as death'' and one denotes ''as good as best possible health for the patients.''

For patients with metastatic prostate cancer it is obvious that even in the TWiST state, patients are far from perfectly healthy. Although a utility score of one in the TWiST state does not mean a state of perfect health, it can serve as a reference state where the patient is free from the symptoms that are studied in the analysis.

To obtain utility scores, QOL data were obtained from a cross-sectional study of 113 patients with metastatic prostate cancer, 60 in remission and 53 with disease progression [21]. These patients completed several QOL questionnaires, including EORTC QLQ-C30(+3), the SF-36 Health Survey, and Spitzer's Quality of Life Index (QLI), a well-validated, cancer-specific QOL instrument. The QLI requires patients to describe their level of functioning in each of five domains (work, self-care, health perception, social support, and outlook). For each domain there are three

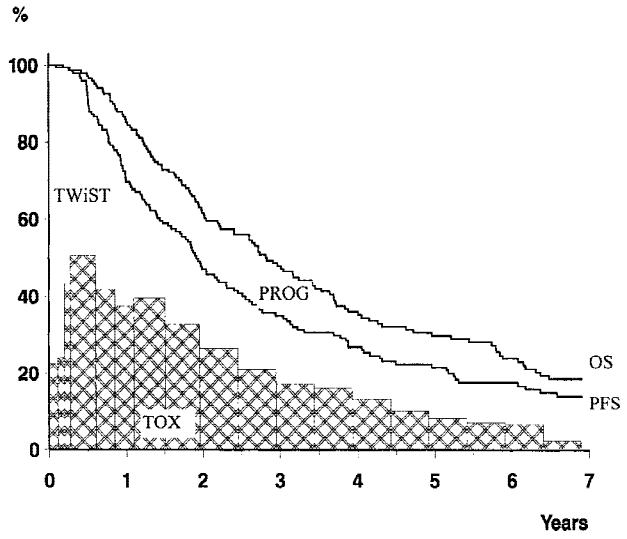

Fig. 2. Mean duration of health states for zoladex and flutamide, model 1.

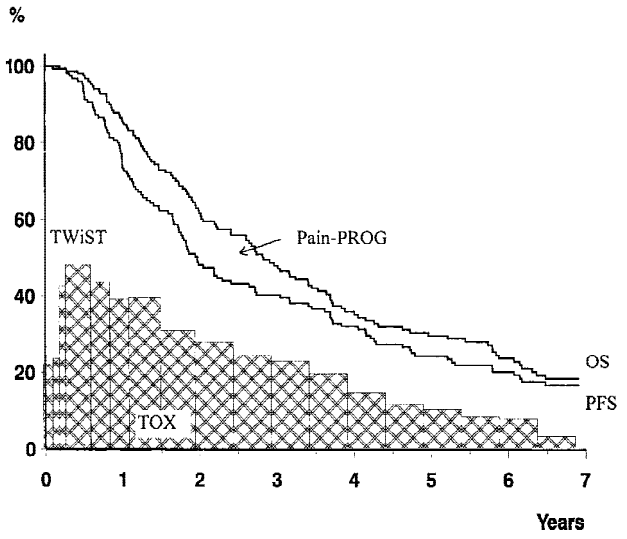

Fig. 4. Mean duration of health states for zoladex and flutamide, model 2.

TABLE III. Q-tility Estimates

Health states	Estimates	
	Crude	Standardized[a]
TOX		
n	44	
$\bar{x}$	.89	.89/.91 = .98
SD	.15	.22
TWiST		
n	16	
$\bar{x}$	.91	.91/.91 = 1
SD	.14	.21
PROG		
n	52	
$\bar{x}$	.85	.85/.91 = .93
SD	.16	.22
Pain-PROG		
n	24	
$\bar{x}$	.76	.76/.91 = .84
SD	.16	.22

[a]Each health states mean value, relative to the mean value for TWiST.

possible levels of functioning. Therefore, there are 243 unique possible health states that can be captured by the QLI [22].

The Q-tility index, developed using multiattribute utility theory, was used to assign utility estimates to each patient's QLI profile [23]. The Q-tility index was developed using a societal reference population to obtain preference weights for each of the QLI health state profiles, using time tradeoff questions. Q-tility scores for having hot flashes in a state of remission, being in a state of progression, and being in a state of progression with pain were set relative to being in a state of remission without hot flashes (TWiST) and used as utility scores in our study. Both crude as well as standardized estimates are shown in Table III.

Comparisons were made by subtracting the weighted sum for the orchiectomy arm from the weighted sum for the zoladex and flutamide arm. Ninety-five percent confidence intervals were calculated, assuming normality based on observed differences between treatment arms. Standard errors of means and covariances between health states were estimated by using the bootstrap method [24].

Every Q-TWiST analysis will contain some degree of uncertainty about the utility scores for the health states. To test the sensitivity of the results and conclusions to changes in the utility scores, a threshold (sensitivity) analysis was performed. The variance/covariance matrix was used to calculate the 95% confidence limits around the threshold line (i.e., the line which represents equality of the two treatment arms). The threshold analysis for the utility scores identified pairs of values for u_{tox} and u_{prog} for which zoladex and flutamide provided more Q-TWiST than orchiectomy, and vice versa. Similarly, pairs of values for u_{tox} and u_{prog} for which the treatment comparison was statistically significant (i.e., $P < 0.05$) were also identified [25].

In this trial, bone metastases were the most common site of first progression (33%). However, if the bone metastases were detected radiologically in patients without bone pain, progression might not be associated with any change in patient symptoms. In this case, it might not be appropriate to assign a lower utility to the PROG state than to the TWiST state. For this reason we defined two models, where in model 1 the PROG state began at the date of first progression according to all of the criteria in Table II. In model 2, the PROG state began at the date of pain progression. Even though pain scores are useful for defining a health state of progression, they do not quantify how patients value time in the health state. Model 2 implies a smaller number of patients in the PROG state (see Table IV) compared to model 1, but the difference between the PROG state and the TWiST state in terms of utility scores is increased.

RESULTS

Figures 1 and 2 present the Q-TWiST plots (model 1) for orchiectomy and zoladex plus flutamide, respectively. As shown in Figures 3 and 4, the PFS curves for model 2 result in a smaller area of pain-PROG than of PROG in model 1; the bars for TOX are only slightly different between models 1 and 2. The bar plots of TOX show that hot flashes were more frequent for the orchiectomy arm in the beginning of follow-up and more frequent for the zoladex and flutamide arm later on.

Restricted means for OS, PFS, each clinical health state, and the Q-TWiST outcome are presented by treatment arm in Table V. Differences between zoladex plus flutamide and orchiectomy plus the 95% confidence intervals are included in Table V. A confidence interval including zero implies that the difference is not statistically significant. After 7 years' follow-up, the difference in mean duration of survival between the two treatment arms was 5.3 months in favor of zoladex and flutamide. With progression defined according to model 1, the average PFS was 6.4 months longer for zoladex and flutamide. However, patients receiving zoladex and flutamide spent on average 4.2 months longer in TOX, 2.2 months longer in TWiST, and 1.1 months less in PROG. When Q-TWiST is calculated using the utilities from the cross-sectional study ($u_{tox} = 0.98$, $u_{prog} = 0.93$), zoladex plus flutamide

TABLE IV. Number of Patients, Progressions, and Deaths

Treatment	Patients	First progression	Pain progression	Deaths
Orchiectomy	148	119	67	119
Zoladex and flutamide	149	103	60	116
Total	297	222	127	235

TABLE V. Overall Means (Months, Restricted to 7 Years) Based on Q-TWiST Analysis

Variable	Treatment		Difference (95% CI)
	Zoladex and flutamide	Orchiectomy	
Model 1, first progression (objective or subjective)			
OS	41.5	36.2	5.3 (−1.0, 11.6)
PFS	33.3	26.9	6.4 (0.4, 12.4)
TOX	15.2	11.0	4.2 (0.5, 7.9)
TWiST	18.1	15.9	2.2 (−2.8, 7.2)
PROG	8.2	9.3	−1.1 (−4.2, 2.0)
Q-TWiST (u_{tox} = 0.98, u_{prog} = 0.93)	40.6	35.3	5.3 (−0.9, 11.5)
Model 2, pain progression			
OS	41.5	36.2	5.3 (−1.0, 11.6)
PFS	36.1	30.9	5.2 (−1.4, 11.8)
TOX	16.4	13.0	3.4 (−0.2, 7.0)
TWiST	19.7	17.9	1.8 (−3.8, 7.4)
Pain-PROG	5.4	5.3	0.1 (−2.4, 2.6)
Q-TWiST (u_{tox} = 0.98, $u_{pain\text{-}prog}$ = 0.88)	40.6	35.4	5.2 (−1.1, 11.5)

has a Q-TWiST that is 5.3 months longer than orchiectomy alone.

With progression defined according to model 2, the average PFS was increased for both treatment arms, with a 5.2-month difference in favor for the zoladex and flutamide arm. However, with this second definition of progression, patients receiving zoladex and flutamide spent an average of 3.4 months longer in TOX, 1.8 months longer in TWiST, and 0.1 month longer in pain-PROG. Using utilities for u_{tox} of 0.98 and $u_{pain\text{-}prog}$ of 0.88, zoladex plus flutamide has a Q-TWiST that is 5.2 months longer than for orchiectomy alone.

A sensitivity analysis was also performed to examine the treatment effect in terms of Q-TWiST, as values for the utility scores (u_{tox} and u_{prog}) vary between 0–1. Figure 5 illustrates this for model 1 in a threshold utility analysis. For all possible pairs of values for u_{tox} and u_{prog}, we determined the quality-adjusted (Q-TWiST) time gained (or lost) for zoladex and flutamide as compared to orchiectomy alone. We found that for all combinations of values, the zoladex and flutamide arm experienced more Q-TWiST on average as compared to orchiectomy alone, and for some combinations of values (i.e., those above the dashed line in Fig. 5), zoladex and flutamide provided statistically significantly (i.e., $P < 0.05$) more Q-TWiST as compared to orchiectomy alone.

DISCUSSION

The primary analysis of EORTC trial 30853, using the ''classical'' approach to calculate the net benefit of two treatment arms in a randomized clinical trial, showed that zoladex and flutamide produced better overall and progression-free survival, but resulted in higher rates of hot flashes and gynecomastia. The benefit in length of life did not take into consideration the ''loss'' in terms of QOL as a result of these treatment-related side effects. We reanalyzed this trial to examine the effect of zoladex and flutamide, taking into account not just survival but also QOL. We applied the Q-TWiST method to adjust OS for the QOL effects due to hot flashes and disease progression, defined in two ways. The results of this analysis show a benefit for zoladex and flutamide almost exactly equal in magnitude to the benefit found in the classical calculation of OS.

The main reason for this small difference between OS and Q-TWiST is the relatively low impact on QOL

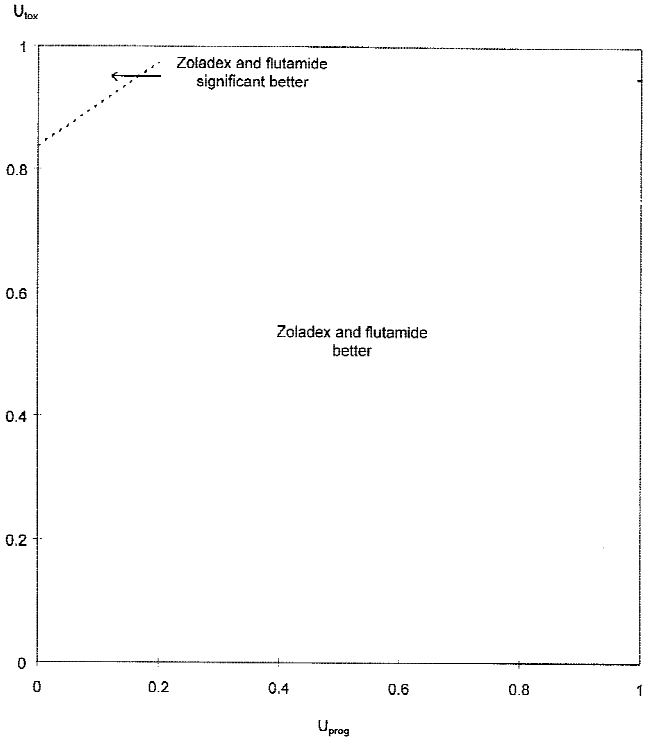

Fig. 5. Threshold utility plot for 7 years' follow-up, model 1.

of hot flashes in terms of utility score for the health state TOX. In other words, the utility score of the health state that contributes to the difference in OS does not differ essentially from the utility score for TWiST. Hot flashes are not such severe side effects that they have a substantial negative impact on the value of overall QOL. As a result, the additional time spent in TOX for the zoladex and flutamide treatment arm is discounted very little in calculating overall Q-TWiST. However, we have to bear in mind that hot flashes constitute only one factor which influences QOL. Gastrointestinal toxicity, for instance, was not recorded on the follow-up form and is a limitation of the study, since physicians may be concerned that gastrointestinal toxicity may limit the usefulness of flutamide. However, gastrointestinal toxicity occurs in 10–15% of prostate carcinoma patients treated with flutamide [12]. Furthermore, by a comparison of the utility score for a health state of gastrointestinal side effects in the study of Bennett et al. [12] with the crude (i.e., compared to perfect health) Q-tility score for toxicity due to hot flashes in Table III, we see a difference of 0.05 (0.84 vs. 0.89). Finally, it should be noted that utility weights necessarily place a single overall value on a state of health, and measure that value in terms of willingness to accept shortened life to improve QOL. This is a fundamentally different approach from traditional descriptive QOL measures that offer much more insight into functioning and distress in the full array of dimensions inherent in health-related QOL, but that are not appropriate for use in quality-adjusted survival (QAS) analysis.

To date, Q-TWiST analyses have defined health states by the usual objective criteria used to determine TOX and PROG in clinical trials. Health states defined in this fashion may not optimally reflect patients' subjective experiences. Data from the cross-sectional survey of prostate cancer patients, which was the source of the utility values used in our analysis, showed a significant difference in utility for symptomatic and asymptomatic progression [26]. We therefore performed the analysis using two different definitions of progression, one based on objective criteria and the other on patient symptoms. The analysis based on patient symptoms found that more Q-TWiST was spent in TOX and TWiST and with less in PROG in both groups, but did not change the conclusions about the benefits of zoladex and flutamide.

The Q-TWiST model incorporated Q-tility scores to adjust the time spent in TOX and PROG according to the value of time in these health states from a QOL point of view. Ideally, it is preferable to obtain QOL evaluations from patients themselves. However, since this was not done prospectively in this trial, it was necessary to obtain such data from another source in order to estimate appropriate values for the utility scores. To accomplish this, we obtained data from a cross-sectional sample of patients with metastatic prostate cancer, using established techniques for utility assessment. Utility scores based on physician focus groups have been reported: Bennett et al. [12] found that the utility scores were 0.84 for toxicity due to flutamide, 0.92 for stable disease, 0.83 for early progressive disease, and 0.42 for late progressive disease. These estimates were relative to a state of perfect health, whereas the Q-tility scores in our model were relative to a state of best possible health given that the patient has prostate cancer (i.e., the TWiST state). For that reason, should the utility scores from Bennett et al. [12] be compared to the crude Q-tility scores in Table III. Our Q-tility scores for TOX and PROG are, even if one looks upon the crude estimates, somewhat higher than those of Bennett et al. [12], but have the advantage of not being based on physician focus groups.

The overview analysis of MAB in advanced prostate cancer indicates that the data from 22 randomized trials of 5,710 patients do not show that MAB results in longer survival than castration alone [27]. Unfortunately, the overview analysis does not address the possibility that MAB may improve time to progression, and may thus possibly represent a potential improvement in QOL if not quantity of life. Our Q-TWiST analysis has the advantage of evaluating the

tradeoff of a possible improvement in time to progression and increased time with treatment-related toxicity. It is especially important to evaluate this tradeoff, if the results of the overview analysis are correct.

CONCLUSIONS

In general, Q-TWiST will approximate OS when the difference in the duration of the various health states between two treatment arms is small, or if the toxicity of the treatment has only a limited negative impact on the value of overall QOL. Thus, studies where a QAS analysis, such as the Q-TWiST method, may provide additional insights to those gained by a classical approach (to investigate the superiority of one treatment over another) are clinical trials where the health states are clearly distinct, and differ significantly in duration, utility score, or both. A consideration of QAS is particularly helpful when the toxicity of treatment is so severe that a patient might be willing to trade off some survival to avoid it.

A frequently expressed concern about performing a QAS analysis is the uncertainty about how to obtain utilities for the various health states. There is controversy not only about whose utilities to use (patients, proxies, or members of the general public), but also which methods to employ to measure those utilities. To date, Q-TWiST analyses have not incorporated specific patient-derived utility scores. Instead, a threshold utility analysis has been used to present treatment comparisons for the entire range of plausible values. A central point in the Q-TWiST method is that it was not designed to provide a single number representing the overall treatment benefit, but to illustrate to decision makers, including patients, the various tradeoffs between QOL and survival.

However, in this evaluation we extended the Q-TWiST approach to include an analysis using actual utility scores from patients in the relevant health state. The use of cross-sectional utility data is the best approach to receive information on patients' utility weights in a retrospective Q-TWiST analysis, since we will not have the opportunity to rerun the clinical trial to get these utility weights. Given the results of the sensitivity analysis showing that zoladex plus flutamide is the preferred strategy over wide ranges of possible utility values, this relatively simple approach proved quite acceptable for this study. For patients who place a high utility score on toxicity and a low utility score on progression (e.g., $u_{tox} = 0.95$ and $u_{prog} = 0.15$), the benefit of 6.03 months in QAS is statistically significant (i.e., $P < 0.05$). On the other hand, for most combinations of values (including the utility scores derived by the cross-sectional study), our sensitivity analysis indicated that the benefit of zoladex and flutamide is not statistically significant (i.e., $P > 0.05$).

ACKNOWLEDGMENTS

This Q-TWiST analysis was made possible by a grant from the EORTC Lady Grierson Research Fellowship. Partial support was provided by American Cancer Society grant PBR-53, which funds the Q-TWiST work of B.F.C.

REFERENCES

1. Denis LJ. Trends in the diagnosis and treatment of prostate cancer, part 1: diagnosis, incidence and pathology. Oncol Pract 1994;1:14–17.
2. Silverberg E, Boring CC, Squires TS. Cancer statistics, 1990. CA 1990;40:9–26.
3. Chamberlain J, Melia J. Screening for prostate cancer. In: Chamberlain J, Moss S, editors. Evaluation of cancer screening. London: Springer; 1996. p 117–136.
4. Rubens RD. Metastatic bone disease in prostate cancer. Oncol Pract 1996;2:2–3.
5. Mettlin CJ, Murphy GP. Prostate cancer. In: Steele GD, Winchester DP, Menck HR, Murphy GP, editors. National cancer database. Annual review of patient care, 1993. Atlanta: American Cancer Society; 1993. p 37–51.
6. James ND, Glaholm J. Prostate cancer—what should be studied? Eur J Cancer 1995;31:1565–1566.
7. Hanks GE, Myers CE, Scardino PT. Cancer of the prostate. In: DeVita VT, Hellman S, Rosenberg SA, editors. Cancer: principles and practice of oncology. Philadelphia: J.B. Lippincott; 1993. p 1073–1113.
8. Denis LJ. Trends in prostate cancer, part 2: treatment options and prognosis. Oncol Pract 1994;2:15–18.
9. Debruyne FMJ, Dijkman GA. Advances and trends in hormonal therapy for advanced prostate cancer. Eur Urol 1995;28:177.
10. Hillner BE, McLeod DG, Crawford ED, Bennett CL. Estimating the cost effectiveness of total androgen blockade with flutamide in M1 prostate cancer. J Urol 1995;45:633–640.
11. EORTC urological group protocol no. 30853.
12. Bennett CL, Matchar D, McCrory D, McLeod DG, Crawford ED, Hillner BE. Cost-effective models for flutamide for prostate carcinoma patients. Cancer 1996;77:1854–1861.
13. Denis LJ, Carneiro de Moura JL, Bono A, Sylvester R, Whelan P, Newling D, Depauw M, Members of the EORTC GU Group and EORTC Data Centre. Goserelin acetate and flutamide versus bilateral orchiectomy: a phase III EORTC trial (30853). J Urol 1993;42:119–130.
14. Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS, for the Ludwig Breast Cancer Study Group. Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis. J Clin Oncol 1989;7:36–44.
15. Gelber RD, Goldhirsch A, for the Ludwig Breast Cancer Study Group. A new endpoint for the assessment of adjuvant therapy in postmenopausal women with operable breast cancer. J Clin Oncol 1986;4:1772–1779.
16. Lenderking WR, Gelber RD, Cotton DJ, Cole BF, Goldhirsch A, Volberding PA, Testa MT, for the AIDS Clinical Trials Group. Evaluation of the quality of life associated with zidovudine treatment in asymptomatic human immunodeficiency virus infection. N Engl J Med 1994;330:738–743.

17. Gelber RD, Goldhirsch A, Cole BF, Wieand HS, Schroeder G, Krook JE. A quality-adjusted time without symptoms or toxicity (Q-TWiST) analysis of adjuvant radiation therapy and chemotherapy for resectable rectal cancer. J Natl Cancer Inst 1996;88: 1039–1045.
18. Cole BF, Gelber RD, Kirkwood JM, Goldhirsch A, Barylak E, Borden E. Quality-of-life-adjusted survival analysis of interferon alfa-2b adjuvant treatment of high-risk resected cutaneous melanoma: an Eastern Cooperative Oncology Group study. J Clin Oncol 1996;14:2666–2673.
19. Gelber RD, Cole BF, Gelber S, Goldhirsch A. Comparing treatments using quality-adjusted survival: the Q-TWiST method. Am Stat 1995;49:161–169.
20. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;54:457–481.
21. Albertsen PC, Aaronson NK, Muller MJ, Keller SD, Ware JE Jr. Health related quality of life among patients with metastatic prostate cancer. J Urol 1997;49:207–217.
22. Spitzer WO, Dobson AJ, Hall J, Chesterman E, Levi J, Shepherd R, Battista RN, Catchlove BR. Measuring the quality of life of cancer patients: a concise QL-index for use by physicians. J Chronic Dis 1981;34:585–597.
23. Weeks JC, O'Leary J, Fairclough D, Paltiel D, Weinstein M. The ''Q-tility index'': a new tool for assessing health related quality of life and utilities in clinical trials and clinical practice. Proc Am Soc Clin Oncol 1994;13:436.
24. Efron B, Tibshirani RJ. The accuracy of a sample mean. In: Efron B, Tibshirani RJ, editors. An introduction to the bootstrap. New York: Chapman & Hall; 1993. p 10–16.
25. Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. Stat Med 1990;9:1259–1276.
26. Weeks JC, Gelber S, Albertsen PC. Impact of progression on quality of life in patients with metastatic prostate cancer (MPC): implications for medical decision-making and cost-effectiveness analyses. J Urol 1997;157:1312A.
27. Prostate Cancer Trialists' Collaborative Group. Maximal androgen blockade in advanced prostate cancer: an overview of 22 randomised trials with 3283 deaths in 5710 patients. Lancet 1995;346:265–269.