Is Prostate-Specific Antigen a Valid Surrogate End Point for Survival in Hormonally Treated Patients With Metastatic Prostate Cancer? Joint Research of the European Organisation for Research and Treatment of Cancer, the Limburgs Universitair Centrum, and AstraZeneca Pharmaceuticals

Laurence Collette, Tomasz Burzykowski, Kevin J. Carroll, Don Newling, Tom Morris, and Fritz H. Schröder

From the European Organisation for Research and Treatment of Cancer Data Center, Brussels; Limburgs Universitair Centrum, Diepenbeek, Belgium; AstraZeneca Pharmaceuticals, Macclesfield, United Kingdom; and Erasmus Medical Centrum, Rotterdam, the Netherlands.

Submitted August 27, 2004; accepted May 11, 2005.

Authors' disclosures of potential conflicts of interest are found at the end of this article.

Address reprint requests to Laurence Collette, MSc, European Organization for Research and Treatment of Cancer Data Center, Avenue Emmanuel Mounier 83/11, B-1200 Brussels, Belgium; e-mail: laurence.collette@eortc.be.

0732-183X/05/2325-6139/$20.00

DOI: 10.1200/JCO.2005.08.156

ABSTRACT

Purpose
The long duration of phase III clinical trials of overall survival (OS) slows down the treatment-development process. It could be shortened by using surrogate end points. Prostate-specific antigen (PSA) is the most studied biomarker in prostate cancer (PCa). This study attempts to validate PSA end points as surrogates for OS in advanced PCa.

Patients and Methods
Individual data from 2,161 advanced PCa patients treated in studies comparing bicalutamide to castration were used in a meta-analytic approach to surrogate end-point validation. PSA response, PSA normalization, time to PSA progression, and longitudinal PSA measurements were considered.

Results
The known association between PSA and OS at the individual patient level was confirmed. The association between the effect of intervention on any PSA end point and on OS was generally low (determination coefficient, < 0.69).

Conclusion
It is a common misconception that high correlation between biomarkers and true end point justify the use of the former as surrogates. To statistically validate surrogate end points, a high correlation between the treatment effects on the surrogate and true end point needs to be established across groups of patients treated with two alternative interventions. The levels of association observed in this study indicate that the effect of hormonal treatment on OS cannot be predicted with a high degree of precision from observed treatment effects on PSA end points, and thus statistical validity is unproven. In practice, non-null treatment effects on OS can be predicted only from precisely estimated large effects on time to PSA progression (TTPP; hazard ratio, < 0.50).

J Clin Oncol 23:6139-6148. © 2005 by American Society of Clinical Oncology

INTRODUCTION

Phase III cancer clinical trials that evaluate the clinical benefit of new treatment options often require large patient numbers and long follow-up. Recent advances in the understanding of the biologic mechanisms of disease development have resulted in the emergence of a large number of potentially effective new agents. There is also increasing public pressure for promising new drugs to receive marketing approval as rapidly as

possible, in particular for life-threatening diseases such as cancer. For these reasons, there is an urgent need to find ways of shortening the duration of cancer clinical trials. The duration of phase III trials results from the use of long-term clinical end points (clinical progression and survival). Therefore, to replace this end point (the "true" end point) by another ("surrogate") end point that could be measured earlier, more conveniently, or more frequently and would adequately reflect the benefit of new treatments on the clinical end point(s) seems to be an attractive solution.

"Biomarkers" (ie, physical signs or laboratory measurements that occur in association with a pathological process or that have putative diagnostic and/or prognostic utility[1]) are generally regarded as the best candidate surrogate end points. A biomarker is an intermediate outcome that is correlated with the true clinical outcome for an individual patient. It may be useful for diagnostic or prognostic information on a particular patient. It is a common misconception that established biomarkers necessarily make valid surrogate end points. To this aim, it is required that "the effect of treatment on a surrogate end point must be reasonably likely to predict clinical benefit."[2] Thus, "surrogacy" is a concept that relates to groups of patients. To demonstrate surrogacy, a strong association between the treatment effects on the surrogate and on the true end point needs to be established across groups of patients treated with the new and standard interventions.

The validation of a candidate surrogate end point is not straightforward. Until recently, the statistical methods developed for this purpose used the data from a single trial.[3-5] These methods suffer from numerous drawbacks: some of them are too stringent to be of practical value, whereas others are based on nontestable assumptions.[6,7] To overcome these limitations, a new methodology, known as the "meta-analytic" validation approach, was developed recently.[8-10] This method uses large databases from multiple randomized clinical trials and aims at measuring directly the association between the treatment effects on the surrogate and the true end point.

In the field of prostate cancer (PCa), prostate-specific antigen (PSA) has probably been the most studied biomarker. It has been investigated as a potential surrogate end point across disease stages,[11-14] and in hormone-refractory patients in particular.[15-18] In a recent article, Buyse et al[19] considered several PSA-based end points in androgen-independent patients treated with liarozole (an imidazole-like compound that causes elevation of retinoic acid, postulated to have antitumor activity), cyproterone acetate, or flutamide. They showed that despite a strong association at the individual patient level, none of the end points qualified as a surrogate for overall survival (OS). In early PCa, Newling et al[20] found a modest correlation between the effect of Casodex on time to PSA progression (TTPP) and on objectively confirmed progression. In primary metastatic PCa, several studies demonstrated some level of association between a post-therapy fall in PSA or a PSA relapse on treatment and long-term survival prognosis.[21-25] However, this merely qualifies PSA as a biomarker. In trial NCI-INT-105, treatment differences in post-therapy PSA levels did not translate into survival differences.[26] Thus, whether PSA is a valid surrogate for survival in hormonally treated PCa remains an open question. This question is of importance, because the use of PSA could shorten the time to the end point from between several months in advanced disease[27] to several years in early disease.[28]

The objective of the present research is to assess PSA-based end points as surrogates for OS in hormone-naïve metastatic PCa using the meta-analytic approach. The data from $>$ 2,000 patients treated with bicalutamide (Casodex) that were made available by AstraZeneca Pharmaceuticals were used for this purpose.

PATIENTS AND METHODS

Individual data from three large international randomized trials of AstraZeneca's Casodex Development Program were used (301/302,[29,30] 306/307,[31] and US trial 0001[32,33]; Table 1). In studies 301/302 and 306/307, Casodex monotherapy (50 and 150 mg/day, respectively) was compared to medical or surgical castration. In the US trial, Casodex (50 mg/day) in combination with goserelin or leuprolide acetate was compared to the combination of flutamides (750 mg/day) and castration in a 2 $\times$ 2 factorial design. All patients were newly diagnosed with metastatic PCa. Four hundred eighty patients with T3-4 M0 disease and elevated PSA from trial 306/307 were excluded. Survival was an end point in all studies, although time to treatment failure (Table 1) was the primary end point in most. PSA was monitored at months 1, 2 (except US trial), and 3 and then every 3 months until month 18 (trial 301/302) or death (other trials). For the analysis, the PSA test date was assumed to be the visit date.

End Points

We considered OS calculated from randomization to the date of death or last visit as the true end point. PCa-specific survival was defined similarly but with deaths unrelated to PCa or treatment censored at the last visit. PSA response, PSA normalization, TTPP, and the complete series of PSA measurements ("PSA profile") were successively assessed as potential surrogate end points for OS.

Patients who had a baseline PSA level at least five times above the normal range ($>$ 20 ng/mL) were included in the analyses of PSA response and PSA normalization. Patients qualified for PSA response if their PSA declined by at least 50% from baseline level at two subsequent observations at least 4 weeks apart. Patients in whom the decline reached a value below or equal to normal (4 ng/mL) qualified for PSA normalization.[25]

Two definitions of TTPP were assessed: (1) For TTPP-1, PSA progression was defined as a PSA value above normal (4 ng/mL), representing a first increase $\geq$ 20% above the nadir[25] (eg, with a PSA nadir of 2 ng/mL, a minimum increase to 4 ng/mL [100% increase] is required, whereas with a PSA nadir of 3.5 ng/mL, a 20% increase to 4.2 ng/mL is enough). (2) For TTPP-2, PSA progression was defined as a PSA value $>$ 2.5 times the normal range (10 ng/mL), representing a first increase $\geq$ 50% above the moving average (based on three

Table 1. Trials Used in the Analysis

Trial	
301/302[28,29]	
Patients	Stage D2, fit for orchidectomy; ECOG performance status 0-2; no prior systemic therapy for prostate cancer, no previous radiotherapy to the prostate within 3 months of entry
Treatments	Bicalutamide (50 mg/d) *v* castration (orchidectomy in trial 301, orchidectomy or goserelin 3.6 mg monthly injection in trial 302)
Design	Open two-arm randomization
Objective	To compare bicalutamide to castration in a pooled analysis
Efficacy end points	Time to treatment failure (objective progression, change of treatment, death as a result of any cause)*; overall survival
Results	Bicalutamide (50 mg/d) demonstrated significantly worse time to progression and survival in trial 301; the trend was not significant in trial 302; by pooled analysis, both end points were significantly worse with bicalutamide than with castration
306/307[30]	
Patients	Metastatic (M1) or locally advanced with PSA five-fold in excess of the upper normal limit (T3-4 M0); only the M1 patients were included in the presently reported analyses; fit for orchidectomy; ECOG performance status 0-2; no prior systemic therapy for prostate cancer, no previous radiotherapy to the prostate within 3 months of entry
Treatments	Bicalutamide (100 or 150 mg/d) or castration (medical or surgical at the patient's discretion)
Design	Initially 2 (Casodex 100 mg):2 (Casodex 150 mg):1 (castration) then changed to 2:1 randomization between Casodex 150 mg and castration
Objective	To demonstrate noninferiority of Casodex 150 mg in comparison to castration by excluding a risk increase of 25%
Efficacy end points	Time to treatment failure (addition of systemic therapy or withdrawal from therapy, objective progression, or death)*; overall survival; objective response
Results (in M1)	Significant differences in favor of castration were found for time to treatment failure (HR, 1.43; 95% CI, 1.20 to 1.71 in favor of castration) and overall survival (HR, 1.30; 95% CI, 1.04 to 1.64)
US trial[31,32]	
Patients	Stage D2 only; ECOG performance status 0-2; no prior systemic therapy for prostate cancer
Treatments	Bicalutamide (50 mg/d) *v* flutamide (250 mg tid) in combination with goserelin acetate (3.6-mg monthly injection) or leuprolide acetate (7.5-mg monthly injection)
Design	2 × 2 factorial design, blinding for the LHRH-A randomization
Objective	To demonstrate noninferiority of bicalutamide + LHRH-A relative flutamide + LHRH-A by excluding a relative-risk increase of 25%
Efficacy end points	Time to treatment failure (addition of systemic therapy or withdrawal from therapy, objective progression, or death)*; overall survival
Results	Noninferior time to treatment failure (HR, 0.93 in favor of bicalutamide; 95% CI, 0.79 to 1.10) Noninferior overall survival (HR, 0.87 in favor of bicalutamide; 95% CI, 0.72 to 1.05)

Abbreviations: ECOG, Eastern Cooperative Oncology Group; PSA, prostate-specific antigen; HR, hazard ratio; LHRH-A, luteinizing hormone-releasing hormone agonist.
*A rising PSA was not considered a sign of progression in any of the studies.

consecutive measurements) nadir. This increase had to be either the last observed value or be sustained for at least 4 weeks[19] (eg, with a nadir of 2 ng/mL at three consecutive occasions, a 500% increase to 10 ng/mL is needed to reach the end point, whereas after a nadir of 7 ng/mL, a 50% increase to 10.5 ng/mL is enough).

Patients who died or are alive without PSA progression were censored at the time of death or last visit, respectively.

Statistical Methods

The meta-analytic approach to surrogate end-point validation has been detailed extensively elsewhere.[6,9,34-36] Thus, we shall only summarize the key features. The method is rooted in the concept that a valid surrogate end point must enable one to predict with sufficient precision the treatment effect on the true clinical end point (OS) from the observed treatment effect on the surrogate (PSA-based) end point. Unlike traditional validation methods such as the Prentice criteria,[3] this new methodology does not require that any of those effects be statistically significant. Indeed, when data from several trials are available, the method consists of simultaneously estimating the relative treatment effects on the survival end point and on the PSA end point (log odds ratio of PSA response or normalization, log hazard ratio [HR] of PSA progression, treatment effect on the longitudinal PSA measurements) in each trial. A model that estimates the association between the treatment effects on the true end point and the corresponding effects on the PSA end points (PSA response,[34] TTPP,[35] or longitudinal PSA measurements[36]) in a way similar to standard linear regression (although mathematically more sophisticated) is then adjusted. As in linear regression, the strength of the association is measured by the squared correlation coefficient that we shall denote R^2_{trial}. This coefficient also indicates the precision with which the treatment effect on the survival end point can be predicted from the observed treatment effect on the surrogate. The maximal possible value of R^2_{trial} is 1, which indicates a perfect prediction. In practice, observing $R^2_{trial} = 1$ is not possible, and one rather seeks a value close to 1, which indicates a strong association between the treatment effects and thus a relatively precise prediction.[9,35] Additionally, the model quantifies the association between the PSA-based end point and the survival end point at the individual patient level. Parameters quantifying the strength of the association at this level will be denoted by the subscript "patient." They can be regarded as measures of validity of the PSA end point as a biomarker for predicting duration of survival.

Only three trials were available, which is too few to allow a precise estimation of R^2_{trial}. Therefore, the patients were grouped

by the trial they entered and their country of residence, as done by Buyse et al.[19] These groups will be henceforth referred to as "trial units."

RESULTS

After excluding nonmetastatic patients and those with no baseline or follow-up PSA measurements, the individual data from 2,161 patients classified into 21 trial units were available for the analysis (Table 2). Their baseline and treatment characteristics are listed in Table 3. More than half of the patients presented with six or more bone metastases. After a median follow-up of 3.25 years, 1,018 patients (52.9%) had died, 815 (71.3%) as a result of PCa (Table 4). The median OS was 2.2 years (95% CI, 2.1 to 2.5) for the Casodex-treated patients and 2.3 years (95% CI, 2.1 to 2.6) in the pooled control groups (Fig 1). The average number of PSA assessments per patient was 6.9 (range, 1 to 23)

PSA Response (≥ 50% Decline From Baseline) and PSA Normalization

PSA response could be assessed for 1,853 patients. A total of 974 (89.4%) and 687 (90.0%) assessable patients on the Casodex and control groups, respectively, achieved a PSA response (Table 4). Only thirteen trial units representing 1,606 patients were used in the analysis: two trial units were removed because no deaths were observed in the castration group, and six were removed because all patients responded in one or both treatment arms. At the individual level, PSA response was a strong predictor of prolonged survival with a survival odds ratio $\theta_{patient}$ of 1.94 (SE, 0.33), representing a two-fold increase in the odds of surviving beyond any specified time t for the PSA responders compared to the nonresponders. At the trial level, the effects of hormonal intervention on PSA response and on OS were poorly correlated with $R^2_{trial} = 0.08$ (SE, 0.14; 95% CI, 0.0 to 0.49). Figure 2A presents the estimated treatment effects on the response (log odds ratio) and OS (log HR).

One should be careful in interpreting these results, because eight trial units with extreme results were excluded from the analysis.

In 399 (36.6%) and 380 (49.8%) of the assessable patients, the PSA declined to a value ≤ 4 ng/mL. Seventeen trial units representing 1,778 patients could be used for this analysis: four were excluded for same reasons as above. At the individual level, the survival odds ratio $\theta_{patient}$ for patients with PSA normalization compared to those without was 4.90 (SE, 0.52), indicating a 4.9-fold greater odds of surviving any specified time t for the patients whose PSA normalized. At the trial level, the treatment effects on PSA and on OS were moderately correlated with $R^2_{trial} = 0.41$ (SE, 0.18; 95% CI, 0.05 to 0.72; Table 5). Figure 2B presents the estimated treatment effects on PSA normalization and OS.

Table 2. Trial Units Available for the Analysis (N = 2,161)

Trial	Country	N
US	Canada	114
US	United States	647
301	Denmark	158
301	Norway	75
301	Sweden	63
302	Austria	46
302	The Netherlands	29
302	United Kingdom	159
306	Denmark	83
306	Finland	69
306	Norway	83
306	Sweden	86
307	Australia	35
307	Austria	14
307	Belgium	95
307	Germany	47
307	The Netherlands	35
307	Italy	11
307	Republic of South Africa	48
307	Spain	22
307	United Kingdom	242

PSA Progression

Nineteen trial units (2,070 patients) and 18 trial units (2,043 patients) could be used for the analysis of TTPP-1 and TTPP-2, respectively (two trial units were excluded from both analyses because of absence of deaths in the castration arm and one from the TTPP-2 analysis because of the absence of PSA progressions in both treatment arms).

The TTPP-1 is presented in Figure 3A: 54.6% of the patients progressed according to this definition (Table 4) within a median time of 11.1 months after being randomly assigned. TTPP-1 was somewhat shorter for the pooled Casodex group than for the control group. TTPP-1 was moderately associated with OS at the individual patient level: the concordance coefficient $\tau_{patient} = 0.52$ (SE, 0.004) indicates that for each individual patient there is an approximately 50% chance to observe a long (short) OS given a long (short) TTPP. At the trial-unit level, the association between the effects of Casodex on TTPP-1 and on OS was low, with $R^2_{trial} = 0.21$ (SE, 0.17; 95% CI, 0.0 to 0.56; Table 5). This analysis is depicted in Figure 4A, where the treatment effect on survival is regressed against the treatment effect on TTPP-1: the size of the circles represents the trial-unit size. The low trial-level association may be partly because of the outlying data from one trial unit. Excluding this unit from the analysis leaves the individual-level association unchanged ($\tau_{patient} = 0.52$; SE, 0.004) but increases R^2_{trial} to 0.58 (SE, 0.15; 95% CI, 0.20 to 0.81).

Only 31.8% of the patients met the more stringent criterion TTPP-2 (Table 4) at a median time of 24.9 months (Fig 3B). At the patient level, the association of TTPP-2 and OS was somewhat stronger than for TTPP-1, with a

Table 3. Patient Characteristics

	Age					Baseline PSA			
	Mean	SD	Median	First and Third Quartiles	Performance Status 0/1/2/3/4, %	Mean	SD	Median	First and Third Quartiles
301/302									
Total (N = 530)			Data not available		Data not available	839.1	1,551.3	267.9	98.6, 784.7
Casodex 50 mg (N = 262)			Data not available		Data not available	811.2	1,477.8	273.2	98.3, 840.0
Castration (N = 268)			Data not available		Data not available	866.3	1,622.2	266.7	99.4, 713.3
306/307 (UICC M1 pts.)									
Total (N = 870)	71.6	8.2	72	66, 78	53.8/32.8/13.3/0/0.1	747.3	1,657.2	179.1	65.7, 634.7
Casodex 100/150 mg (N = 617)	71.2	8.2	72	66, 77	54.0/31.9/14.1/0/0	772.6	1,772.5	189.8	64.5, 658.4
Castration (N = 253)	72.7	8.1	73	67, 78	53.4/34.8/11.5/0/0.4	685.6	1,336.0	156.0	67.0, 587.3
US (D2 pts.)									
Total (761)	70.2	8.7	70	65, 76	51.4/37.2/11.4/0/0	694.2	1,444.2	174.3	45.6, 580.6
Casodex + castration (N = 377)	69.8	8.2	70	65, 75	53.8/36.1/10.1/0/0	650.4	1,382.8	170.0	53.8, 588.1
Flutamide + castration (N = 384)	70.5	9.2	71	65, 77	49.0/38.3/12.8/0/0	737.3	1,502.6	178.3	38.7, 576.5

Abbreviations: PSA, prostate-specific antigen; SD, standard deviation.

concordance coefficient $\tau_{patient} = 0.61$ (SE, 0.02). The association between the treatment effects on TTPP-2 and OS was somewhat higher than for TTPP-1, with $R^2_{trial} = 0.66$ (SE, 0.13; 95% CI, 0.30 to 0.85; Fig 4B and Table 5).

Longitudinal Measurements of PSA

All previously considered PSA-based end points are summary measures derived from the longitudinal PSA measurements and use only a limited amount of the available information. It thus seemed logical to investigate if the longitudinal series of PSA measurements would not be a better surrogate end point for OS. Figure 5A presents the mean profiles of log-transformed PSA measurements for groups of patients with similar observation time: all profiles eventually end with a PSA increase (progression), and patients with an early progression tend to have a higher initial PSA that does not decrease as much early on.

Figure 5B displays the mean PSA profiles per treatment group: starting from week 52 the curves show a relatively stable linear decrease rather than the increasing curvature observed in Figure 4A. This distortion results from attrition: progressive patients, in whom PSA increases, tend to leave the study, and thus the curve in Figure 5B reflects only those with stable PSA.

In view of Figure 5A, the treatment effect on the log-transformed PSA levels was expressed as a function of time and its square root in a joint model of PSA measurements and survival times. In that model, the individual patient-level association between the PSA process and the hazard of dying is a function of time and cannot be easily summarized into a single measure.[35] The results indicated that the correlation between the individual PSA and mortality hazard processes was > 0.90 at any time > 7 months, which suggests a strong association between the PSA profile and the hazard of dying for individual

Table 4. Survival and Prostate-Specific Antigen Outcome

	Casodex (n = 1,256)		Control (n = 905)		Total (N = 2,161)	
	No.	%	No.	%	No.	%
Alive	571		447		1,018	
Dead	685		458		1,143	52.9
Because of prostate cancer	496		319		815	71.3
Because of another cause	189		139		328	28.7
PSA response						
Evaluable	1,090		763		1,853	
Decline to ≤ 4 ng/mL	399	36.6	380	49.8	779	42.0
Decline by ≥ 50% of baseline	575	52.8	307	40.2	882	47.6
No response	116	10.6	76	10.0	192	10.4
Not evaluable	142		166		308	
PSA progression (TTPP-1)	415		729		1,144	
PSA progression (TTPP-2)	432		233		665	
Not evaluable for PSA progression	35		32		67	

Abbreviations: PSA, prostate-specific antigen; TTPP, time to PSA progression.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.