Potential Surrogate Endpoints for Prostate Cancer Survival: Analysis of a Phase III Randomized Trial

Michael E. Ray, Kyounghwa Bae, Maha H. A. Hussain, Gerald E. Hanks, William U. Shipley, Howard M. Sandler

- **Background** The identification of surrogate endpoints for prostate cancer–specific survival may shorten the length of clinical trials for prostate cancer. We evaluated distant metastasis and general clinical treatment failure as potential surrogates for prostate cancer–specific survival by use of data from the Radiation Therapy and Oncology Group 92-02 randomized trial.
	- **Methods** Patients (n = 1554 randomly assigned and 1521 evaluable for this analysis) with locally advanced prostate cancer had been treated with 4 months of neoadjuvant and concurrent androgen deprivation therapy with external beam radiation therapy and then randomly assigned to no additional therapy (control arm) or 24 additional months of androgen deprivation therapy (experimental arm). Data from landmark analyses at 3 and 5 years for general clinical treatment failure (defined as documented local disease progression, regional or distant metastasis, initiation of androgen deprivation therapy, or a prostate-specific antigen level of 25 ng/mL or higher after radiation therapy) and/or distant metastasis were tested as surrogate endpoints for prostate cancer-specific survival at 10 years by use of Prentice's four criteria. All statistical tests were two-sided.
	- **Results** At 3 years, 1364 patients were alive and contributed data for analysis. Both distant metastasis and general clinical treatment failure at 3 years were consistent with all four of Prentice's criteria for being surrogate endpoints for prostate cancer-specific survival at 10 years. At 5 years, 1178 patients were alive and contributed data for analysis. Although prostate cancer–specific survival was not statistically significantly different between treatment arms at 5 years ($P = .08$), both endpoints were consistent with Prentice's remaining criteria.
- **Conclusions** Distant metastasis and general clinical treatment failure at 3 years may be candidate surrogate endpoints for prostate cancer-specific survival at 10 years. These endpoints, however, must be validated in other datasets.
	- J Natl Cancer Inst 2009;101: 228 236

Prostate cancer is a malignancy with a long natural history. Even men who are initially diagnosed with high-grade and locally advanced prostate cancer often survive for many years. Because of the long survival time, clinical trials of prostate cancer that are designed with primary endpoints of overall or prostate cancerspecific survival require long follow-up periods, especially those evaluating treatments for clinically localized disease. The time required for the conception, design, conduct, analysis, and initial reporting of a prostate cancer clinical trial often approaches 10 years (1-4). Identification of surrogate endpoints for prostate cancer cause-specific or overall survival would shorten the time required to conduct prostate cancer clinical trials and thus improve the chances of finding better treatments for prostate cancer. For patients with localized prostate cancer, the ideal surrogate endpoint for survival should use clinical information available as soon as possible after definitive local therapy that will identify patients highly likely to die of their disease. The findings from this study would most directly apply to patients treated with primary external

beam radiation therapy; further research is required to show that the findings could also apply to surgically treated patients.

The Radiation Therapy and Oncology Group (RTOG) 92-02 trial is a phase III, randomized multi-institutional clinical trial that was conducted between June 26, 1992, and April 15, 1995, during the

Affiliations of authors: Radiology Associates of Appleton, Appleton, WI (MER); Department of Statistics, Radiation Therapy Oncology Group, Philadelphia, PA (KB); Department of Internal Medicine (MHAH) and Department of Radiation Oncology (HMS), University of Michigan, Ann Arbor, MI; Department of Radiation Oncology, Fox Chase Cancer Center, Philadelphia, PA (GEH); Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA (WUS) .

Correspondence to: Michael E. Ray, MD, PhD, Radiology Associates of Appleton, 1818 North Meade St, Appleton, WI 54911 (e-mail: michael.ray@ thedacare.org).

See "Funding" and "Notes" following "References."

DOI: 10.1093/jnci/djn489

[©] The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org.

metastasis and general clinical treatment failure (as defined below) as surrogate endpoints for prostate cancer-specific survival by use of data from the phase III randomized RTOG 92-02 trial. **Patients and Methods Patients, Treatment, and Follow-up** Patients had histologically confirmed prostate adenocarcinoma that was clinical stage T2c to T4 (1992 American Joint Committee on Cancer Staging) with no evidence of lymph node metastasis in the common iliac or higher lymph node chains. Karnofsky performance scores were at least 70, and pretreatment PSA levels were less than 150 ng/mL. All patients received neoadjuvant and concurrent, short-term androgen deprivation therapy, with total androgen suppression from treatment with goserelin acetate (3.6 mg, subcutaneously

administered once each month) and flutamide (250 mg by mouth three times per day) for 2 months before and an additional 2 months during external beam radiation therapy. Radiation therapy was delivered to the whole pelvis followed by a boost to the prostate by use of a four-field technique and megavoltage equipment (minimum beam energy of 6 MV). Treatments were once daily, and doses were 1.8–2.0 Gy per fraction. Regional pelvic lymphatics received a total of 44–50 Gy, and the prostate received total doses of 65 – 70 Gy for T2c tumors or 67.5 – 70 Gy for T3 or T4 tumors. Patients were stratified by clinical stage, pretreatment PSA level, tumor grade, and lymph node status. Patients were randomly assigned to receive either no additional therapy (the control arm) or an additional 2 years of monthly subcutaneous injections (3.6 mg of goserelin acetate) in the long-term androgen deprivation therapy (the experimental arm).

so-called prostate-specific antigen (PSA) era, when serum PSA measurements for screening and monitoring of treatment response and failure became widespread. In this trial, patients with locally advanced prostate cancer were first treated with 4 months of neoadjuvant and concurrent androgen deprivation therapy and external beam radiation therapy and then randomly assigned to either no additional therapy (the short-term androgen deprivation or control arm) or an additional 2 years of androgen suppression (the long-term androgen deprivation or experimental arm). The initial results of this trial were published in 2003 (4,5); results of the 10-year follow-up have been reported (6). Both analyses found that patients in the experimental arm had statistically significantly better prostate cancer-specific survival than those in the control arm. The treatment regimen established in this trial is considered a current standard of care for men with locally advanced, high-risk prostate cancer who are treated with primary external beam radiation therapy. The purpose of this analysis was to evaluate distant

Patients were seen for follow-up visits every 3 months for year 1, every 4 months for year 2, every 6 months for years $3-5$, and once a year thereafter. At each visit, their history of symptoms or side effects since the preceding visit was obtained; a physical examination was performed; their Karnofsky performance score was determined; a sexual function assessment was obtained; and liver function tests, complete blood cell count, and PSA measurements were obtained. At each visit, the tumor status and toxic effects were recorded; however, obtaining radiographic imaging

CONTEXT AND CAVEATS

Prior knowledge

Surrogate endpoints for prostate cancer-specific survival may shorten the length of time required to conduct clinical trials for prostate cancer.

Study design

Prentice's four criteria were used to test 3- and 5-year results from a phase III randomized trial in patients with locally advanced prostate cancer to determine whether general treatment failure or distant metastasis at these times qualified as surrogate endpoints for prostate cancer-specific survival at 10 years.

Contribution

Analysis of data at 3 years from 1364 patients found that both general treatment failure or distant metastasis were consistent with all of Prentice's four criteria for being surrogate endpoints for prostate cancer-specific survival at 10 years.

Implications

General treatment failure or distant metastasis should be further validated as surrogate endpoints in other datasets.

Limitations

j

Data from only one trial were used in this study. Statistical methods for establishing and validating a surrogate endpoint, such as Prentice's criteria, must be applied with caution and with an appreciation of their limitations.

From the Editors

was neither dictated nor suggested by the protocol. The documentation of local, regional, or distant failures was based on clinical examination and radiographic or other testing that was obtained at the discretion of the follow-up physicians.

Study Endpoints

The failure event for prostate cancer-specific survival was prostate cancer – specific death, defined as death from prostate cancer or from complications that were related to the protocol treatment. The first potential surrogate endpoint examined was distant metastasis because it is frequently associated with the development of fatal prostate cancer. The failure event for distant metastasis was defined as documented clinical evidence of distant metastasis from a positive bone scan, a computed tomography scan, or other radiographic imaging. We also examined an additional endpoint, general clinical failure, to more broadly survey clinical, biochemical, and treatmentrelated data that may be associated with fatal prostate cancer. This composite endpoint includes distant metastasis, also surveys other potentially important patient data, and has been explored as a potential survival surrogate previously by other authors $(7,8)$. For this study, the failure event for general clinical treatment failure was defined as time to the first occurrence of 1) a documented local prostate cancer recurrence, 2) documented regional or distant prostate cancer metastasis, 3) initiation of androgen deprivation therapy after completion of the protocol treatment, or 4) a documented PSA level of 25 ng/mL or higher after completion of radiation therapy. The time to each endpoint was measured from the date of random assignment to the date of the failure event.

Statistical Methods

Comparisons of survival times of patients that are classified according to a surrogate endpoint may be subject to length bias because treatment for patients who have long survival times has probably not failed. A landmark analysis (9) was used to correct for this length bias. Only patients alive at the landmark time, 3 or 5 years, were included in the analysis and their surrogate endpoint failure status was assessed at the landmark time. The true endpoint was time to events for prostate cancer – specific survival. The surrogate endpoints, distant metastasis and general clinical treatment failure, were considered as binary variables (absence or presence of a failure event at 3 or 5 years). Prentice's four criteria (10) were used to evaluate whether distant metastasis and general clinical treatment failure by 3 or 5 years are surrogate endpoints for prostate cancer – specific survival at 10 years of follow-up in the RTOG 92-02 trial . Surrogacy requires that a surrogate for a true endpoint yields a valid test of the null hypothesis of no association between treatment and the true response. Prentice's criteria were validated by testing the following four criteria: 1) treatment has a statistically significant impact on the true endpoint; 2) treatment has a statistically significant impact on the surrogate endpoint; 3) the surrogate endpoint has a statistically significant impact on the true endpoint; and 4) the full effect of the treatment on the true endpoint should be captured by surrogate endpoint. We also showed that there is no statistical difference with respect to the true endpoint between the two arms in the trial, that is, the group for whom the surrogate endpoint failed to meet Prentice's criteria and the group for whom it did not. This argument suggested that the true endpoint benefit in favor of a particular arm is due to the nonfailure by the surrogate endpoint obtained with the treatment of this arm compared with the other arm. Specifically, the findings suggested that the prostate cancer – specific survival benefit in favor of the experimental arm was associated with the non-distant metastasis rates at 3 or 5 years in the experimental treatment arm, compared with those in the control treatment arm. The cumulative incidence method (11) was used to estimate rates of prostate cancer – specific survival, distant metastasis, and general clinical treatment failure because it specifically considers other competing causes of death. Gray's test statistic (12) was used to compare cumulative incidence rates. The competing risk for prostate cancer – specific survival is a death without a failure event for prostate cancer-specific survival. The competing risk for distant metastasis is a death without distant metastasis and that for general clinical treatment failure is a death. Cox proportional hazard models (13) were used to determine the hazard ratio (HR) between the two treatments with respect to the outcomes. The proportional hazards assumption was confirmed by linearity between the negative cumulative hazard function and survival time. All statistical tests were two-sided.

Results

The patient cohort for the RTOG 92-02 trial included 1521 eligible and evaluable patients from the 1554 patients who were randomly assigned to treatment (6). Among 1521 eligible and evaluable patients who were randomly assigned to treatment, there were 218 prostate cancer-specific deaths, 291 patients developed

distant metastasis , and 726 patients developed general clinical treatment failures during the available follow-up period. Among all patients, the median follow-up was 8.95 years for patients in the control arm and 9.18 years for patients in the experimental arm. Among patients who were still alive at the time of analysis, median follow-up was 11.31 years for patients in the control arm and 11.27 years for patients in the experimental arm. The 10-year prostate cancer – specific survival rate was 84.6% (95% confidence interval $|CI| = 81.9\%$ to 87.2%) in the control arm, compared with 88.9% (95% CI = 86.7% to 91.3%) in the experimental arm (HR of prostate cancer–specific death = $0.70, 95\%$ CI = 0.54 to $0.92; P = .01$). The 3- and 5-year distant metastasis rates were 10.7% (95% CI = 8.5% to 12.9%) and 13.9% (95% CI = 11.4% to 16.4%), respectively, for patients in the control arm, compared with 5.2% (95% $CI = 3.6\%$ to 6.8%) and 8.6% (95% $CI = 6.6\%$ to 10.6%), respectively, for patients in the experimental arm $(HR = 0.61,$ 95% CI = 0.48 to 0.77; *P* < .001). The 3- and 5-year general clinical treatment failure rates were 34.9% (95% CI = 31.5% to 38.3%) and 42.9% (95% CI = 39.4% to 46.5%), respectively, for patients in the control arm, compared with 26.3% (95% CI = 23.2% to 29.5%) and 32.7% (95% CI = 29.3% to 36.0%), respectively, for patients in the experimental arm (HR = 0.70, 95% CI = 0.61 to 0.81 ; $P < .001$).

Evaluation of the Surrogacy of Distant Metastasis and General Clinical Treatment Failure by 3 Years

A landmark analysis (8) at 3 years was performed by including only data from patients who were alive at 3 years from the date of random assignment (Table 1). Among the 1364 patients in this subgroup, the median follow-up was 9.58 years for the 685 patients in the control arm and 9.91 years for the 679 patients in the experimental arm. Among the 675 patients who were still alive at the time of the current analysis, median follow-up was 11.32 years for patients in the control arm and 11.30 years for those in the experimental arm. The numbers of events during available follow-up in this 3-year landmark group were as follows: 183 prostate cancer – specific deaths (107 in the control arm and 76 in the experimental arm), 259 distant metastases (158 in the control arm and 101 in the experimental arm), and 675 general clinical treatment failures (380 in the control arm and 295 in the experimental arm). Among the 259 patients who had distant metastasis, 88 (34%) had an event by 3 years, and among the 675 patients who had a general clinical treatment failure, 412 (61%) had an event by 3 years $(Table 1).$

There was no statistically significant difference with respect to the prostate cancer-specific survival between the two arms among those with (HR of prostate cancer-specific death $= 0.95$, 95% CI = 0.60 to 1.50) and those without (HR of prostate cancer-specific death = 0.76 , 95% CI = 0.51 to 1.11) distant metastasis (Table 2). This result suggested that the prostate cancer-specific survival benefit in the experimental arm, compared with the control arm, was due to the lower rate of distant metastasis at 3 years that was associated with the experimental treatment (HR = $0.69, 95\%$ CI = 0.52 to 0.93 ; Table 2). Similar results were found for groups with $(HR = 0.81, 95\% \text{ CI} = 0.58 \text{ to}$ 1.14) and without (HR = 0.88, 95% CI = 0.48 to 1.63) general clinical treatment failure (Table 2). Thus, these observations

Table 1. Distribution of endpoints from the landmark analysis at 3 years ($n = 1364$)*

	No. (%) of cumulative DM events			No. (%) of cumulative GCTF events			No. (%) of cumulative PCSDs		
Time to event, y	Control arm	Experimental arm	Total	Control arm	Experimental arm	Total	Control arm	Experimental arm	Total
\leq 1	2(1)	2(2)	4(1.5)	45 (12)	32(11)	77 (11.4)			
>1 to 2	25(16)	9(9)	34(13.1)	103 (27)	58 (20)	161 (23.9)			
>2 to 3	61 (39)	27(27)	88 (34.0)	235 (62)	177 (60)	412 (61.0)			
>3 to 4	74 (47)	38 (38)	112 (43.2)	267 (70)	199 (67)	466 (69.0)	17(16)	7(9)	24 (13.1)
>4 to 5	85 (54)	52(51)	137 (52.9)	295 (78)	224 (76)	519 (76.9)	33 (31)	20 (26)	53 (29.0)
>5 to 6	98 (62)	65 (64)	163 (62.9)	315 (83)	240 (81)	555 (82.2)	47(44)	31(41)	78 (42.6)
>6 to 7	114 (72)	70 (69)	184 (71.0)	341 (90)	256 (87)	597 (88.4)	55 (51)	37(49)	92(50.3)
>7 to 8	125 (79)	82 (81)	207 (79.9)	355 (93)	267 (91)	622 (92.2)	67 (63)	45 (59)	112(61.2)
>8 to 9	136 (86)	88 (87)	224 (86.5)	367 (97)	275 (93)	642 (95.1)	82 (77)	55 (72)	137 (74.9)
>9 to 10	147 (93)	95 (94)	242 (93.4)	375 (99)	284 (96)	659 (97.6)	92 (86)	64 (84)	156 (85.3)
>10 to 11	151 (96)	98 (97)	249 (96.1)	379 (99)	287 (97)	666 (98.7)	101 (94)	69 (91)	170 (92.9)
>11 to 12	157 (99)	99 (98)	256 (98.8)	379 (99)	293 (99)	672 (99.6)	104 (97)	73 (96)	177 (96.7)
>12	158 (100)	(100) 101	259 (100)	380 (100)	295 (100)	675 (100)	107 (100)	76 (100)	183 (100)

Time to event was from the date of random assignment to distant metastasis, general clinical treatment failure, or prostate cancer-specific death. DM = distant metastasis; GCTF = general clinical treatment failure; PCSDs = prostate cancer-specific deaths; —= no PCSDs.

indicate that distant metastasis at 3 years and also general clinical treatment failure at 3 years might be potential surrogate endpoints for prostate cancer-specific survival at 10 years, according to Prentice's definition.

Prentice's Criterion 1. The true endpoint, the 10-year prostate cancer-specific survival rate, was 86% (95% CI = 83.3% to 88.6%) for patients in the control arm and 90% (95% CI = 87.7% to 92.4%) for patients in the experimental arm $(P = .02;$ Table 3 and Figure 1). Therefore, Prentice's first criterion, that the investigational treatment (experimental) has a statistically significant impact on the true endpoint prostate cancer – specific survival, was satisfied.

Prentice's Criterion 2. Statistically significant treatment effects were observed on distant metastasis (odds ratio [OR] = 2.36, 95% $CI = 1.48$ to 2.76; Table 2 and Figure 2, A) and on general clinical treatment failure (OR = 1.48 , 95% CI = 1.17 to 1.88; Table 2 and Figure 2, B). Therefore, Prentice's second criterion for both distant metastasis and general clinical treatment failure, that the investigational treatment has a statistically significant impact on the surrogate endpoint, was satisfied.

Prentice's Criterion 3. Both distant metastasis and general clinical treatment failure had a statistically significant impact on time to prostate cancer-specific survival (for distant metastasis, $P < .001$; and for general clinical treatment failure, $P = .004$; Table 4).

Table 2. Prostate cancer-specific survival rate by surrogate endpoints at 3- (n = 1364) and 5-year (n = 1178) landmark analyses*

* PCSS = prostate cancer–specific survival; Control arm = short-term androgen deprivation and radiation therapy arm; Experimental arm = long-term androgen deprivation and radiation therapy arm; OR = odds ratio of distant metastasis or general clinical treatment failure event; HR = hazard ratio of prostate cancer–specific death; CI = confidence interval; DM = distant metastasis; GCTF = general clinical treatment failure.

* Control arm = short-term androgen deprivation and radiation therapy; experimental arm = long-term androgen deprivation and radiation therapy; HR = hazard ratio of death; CI = confidence interval.

† Death rates at 10 years.

‡ Median survival time for all patients in the 3-year landmark analysis (n = 1364) and the 5-year landmark analysis (n = 1178).

Therefore, Prentice's third criterion for both distant metastasis and general clinical treatment failure, that the surrogate endpoint has a statistically significant impact on the true endpoint, was satisfied.

Prentice's Criterion 4. The treatment effect on prostate cancerspecific survival was not statistically significant in either the group with distant metastasis or general clinical treatment failure (for distant metastasis, $P = .54$; and for general clinical failure, $P = .29$) or in the group without distant metastasis or general clinical treatment failure (for non-distant metastasis, $P = .39$; and for

Figure 1. Prostate cancer-specific survival for the 3-year landmark analysis of the Radiation Therapy and Oncology Group 92-02 trial. Patients in the control arm were treated with short-term androgen deprivation and radiation therapy; patients in the experimental arm were treated with long-term androgen deprivation and radiation therapy. Gray's test statistic was used, and all statistical tests were twosided. **Error bars** at 5, 8, and 10 years = 95% confidence intervals.

non-general clinical failure, $P = .98$; Table 4). These findings suggest that the full effect of treatment on prostate cancer – specific survival may be captured by the potential surrogate endpoints, independent of treatment, and so Prentice's fourth criterion was not rejected.

Evaluation of Surrogacy of Distant Metastasis and General Clinical Treatment Failure by 5 Years

The 5-year landmark analysis was performed by including only data from patients who were alive at 5 years from the date of random assignment (Table 5). Among the 1178 patients in this subgroup, the median follow-up was 10.35 years for the 587 patients in the control arm and 10.50 years for the 591 patients in the experimental arm. Among the 659 patients in this subgroup who were still alive at the time of the current analysis, median follow-up was 11.35 years in control arm and 11.33 years in the experimental arm. There were 130 prostate cancer – specific deaths (74 in the control arm and 56 in the experimental arm), 196 distant metastases (117 in the control arm and 79 in the experimental arm), and 574 general clinical treatment failures (324 in the control arm and 250 in the experimental arm) available for analysis in this subgroup. Among the 196 patients who had distant metastasis, 74 (38%) had their failure event by 5 years (44 in the control arm and 30 in the experimental arm), and among the 574 patients who had general clinical treatment failure, 418 (73%) had their failure event by 5 years (239 in the control arm and 179 in the experimental $arm)$ (Table 5).

There was no statistical difference at 5 years with respect to the prostate cancer-specific survival between the two arms among those with (HR of prostate cancer–specific death = 1.35 , 95% CI = 0.65 to 2.78) and without (HR of prostate cancer-specific death = 0.70, 95% CI = 0.47 to 1.04) distant metastasis (Table 2). These

Figure 2. Surrogate endpoint nonfailure rates for the 3-year landmark analysis of the Radiation Therapy and Oncology Group 92-02 trial. **A**) Distant metastasis nonfailure rates. **B**) General clinical treatment nonfailure rates. Patients in the control arm were treated with short-term androgen deprivation and radiation therapy; patients in the experimental arm were treated with long-term androgen deprivation and radiation therapy. Gray's test statistic was used, and all statistical tests were twosided. **Error bars** at 5, 8, and 10 years = 95% confidence intervals.

results suggest that the prostate cancer–specific benefit in the experimental arm, compared with the control arm, was due to the lower rate of distant metastasis at 5 years that was associated with the experimental treatment (HR = 0.73 , 95% CI = 0.52 to 1.04; Table 2), although the rate was not statistically significantly lower. Similar results were obtained for the groups with (HR = 0.85, 95% $CI = 0.56$ to 1.30) and without (HR = 0.84, 95% CI = 0.45 to 1.57) general clinical treatment failure. Thus, these observations indicate that distant metastasis at 5 years and also general clinical treatment failure at 5 years may be potential surrogate endpoints for prostate cancer-specific survival at 10 years, according to Prentice's definition.

Prentice's Criterion 1. The true endpoint, the 10-year prostate cancer-specific survival rate, was 89% (95% CI = 86.8% to 91.9%) for patients in the control arm, compared with 92% (95% CI = 89.9% to 94.4%) for patients in the experimental arm $(P = .08$; Table 3). Therefore, Prentice's first criterion, that the investigational treatment (in the experimental arm) has a statistically

Table 4. Prentice's criteria 3 and 4 for the 3-year landmark analysis (n = 1364)*

The following proportional model was used to validate the criteria. Prentice's third criterion, that the surrogate endpoint has a statistically significant impact on the true endpoint, can be tested by $H_0: \beta_1 = 0$. Prentice's fourth criterion can be tested by H_0 : $\beta_2 = 0$ and H_0 : $\beta_3 = 0$.

$$
\ln \frac{h(t)}{h_0(t)} = \beta_1 \times S(t) + \beta_2 \times \text{TRT} \times S(t) + \beta_3 \times \text{TRT} \times [1 - S(t)],
$$

where $h(t)$ = hazard rate of prostate cancer-specific death at time t; $S(t)$ = surrogate endpoint (failure = 1 and nonfailure = 0); $1 - S(t)$ = not having surrogate endpoint failure; $TRT =$ treatment arm analyzed (control $= 0$ and experimental = 1); $DM =$ distant metastasis; $GCTF =$ general clinical treatment failure; β_1 = coefficient for surrogate endpoint effect; β_2 = coefficient for treatment effect; β_3 = coefficient for the interaction between the not having a failure for surrogate endpoint and treatment.

† Chi-square test statistic. All statistical tests were two-sided.

significant impact on the true endpoint, prostate cancer-specific survival, was not satisfied.

Prentice's Criterion 2. The investigational treatment was not statistically significantly associated with distant metastasis (OR = 1.52, 95% CI = 0.94 to 2.45; Table 2) but was associated with general clinical treatment failure (OR = 1.58 , 95% CI = 1.24 to 2.01). Therefore, Prentice's second criterion, that the investigational treatment has a statistically significant impact on the surrogate endpoint, was satisfied for general clinical treatment failure but not for distant metastasis.

Prentice's Criterion 3. Both distant metastasis and general clinical treatment failure had a statistically significant impact on time to prostate cancer – specific survival (for distant metastasis, *P* = .004; and for general clinical treatment failure, $P = .02$; Table 6). Therefore, Prentice's third criterion, that the surrogate endpoint has a statistically significant impact on the true endpoint, was satisfied for both distant metastasis and general clinical treatment failure.

Prentice's Criterion 4. The treatment effect on prostate cancerspecific survival was not statistically significant either in the group with distant metastasis or general clinical treatment failure (for distant metastasis, $P = .17$; and for general clinical failure, $P = .58$) or in the group without distant metastasis or general clinical treatment failure (for non-distant metastasis, proportional hazard model coefficient = -0.59 , $P = .14$; and for non-general clinical failure, $P = .74$; Table 6). The full effect of treatment on prostate cancer-specific survival appears to have been captured by the candidate surrogate endpoints, independent of treatment. Therefore, Prentice's fourth criterion, that the full effect of the treatment on the true endpoint should be captured by the surrogate endpoint, was not rejected.

Table 5. Distribution of endpoints from the landmark analysis at 5 years ($n = 1178$)*

	No. (%) of cumulative DM events			No. (%) of cumulative GCTF events			No. (%) of cumulative PCSDs		
Time to	Control	Experiment		Control	Experiment			Experiment	
event, y	arm	arm	Total	arm	arm	Total	Control arm	arm	Total
\leq 1	1(1)	2(3)	3(1.5)	32(10)	28(11)	60 (10.5)			
>1 to 2	14(12)	6(8)	20(10.2)	70 (23)	44 (18)	114 (19.9)			
>2 to 3	27(23)	12(15)	39 (19.9)	182 (56)	141 (56)	323 (56.3)			
>3 to 4	35 (30)	18 (23)	53 (27.0)	211 (65)	155 (62)	366 (63.8)			
>4 to 5	44 (38)	30 (38)	74 (37.8)	239 (74)	179 (82)	418 (72.8)			
>5 to 6	57 (49)	43 (54)	100 (51.0)	259 (80)	195 (78)	454 (79.1)	14 (19)	11(20)	25(19.2)
>6 to 7	73 (62)	48 (61)	121 (61.7)	285 (88)	211 (84)	496 (86.4)	22 (30)	17 (30)	39(30.0)
>7 to 8	84 (72)	60 (76)	144 (73.5)	299 (92)	222 (89)	521 (90.8)	34 (46)	25 (45)	59 (45.8)
>8 to 9	95 (81)	66 (84)	161 (84.1)	311 (96)	230 (92)	541 (94.3)	49 (66)	35 (63)	84 (64.6)
>9 to 10	106 (91)	73 (92)	179 (91.3)	319 (98)	239 (96)	558 (97.2)	59 (80)	44 (79)	103 (79.2)
>10 to 11	110 (94)	76 (96)	186 (94.9)	323 (99)	242 (97)	565 (98.4)	68 (92)	49 (88)	117 (90.0)
>11 to 12	116 (99)	77 (97)	193 (98.5)	323 (99)	248 (99)	571 (99.5)	71 (96)	53 (95)	124 (95.4)
>12	117 (100)	79 (100)	196 (100)	324 (100)	250 (100)	574 (100)	74 (100)	56 (100)	130 (100)

* Time to event was from the date of random assignment to distant metastasis, general clinical treatment failure, or prostate cancer–specific death.

DM = distant metastasis; GCTF = general clinical treatment failure; PCSDs = prostate cancer–specific deaths; —= no PCSDs.

Discussion

This study is a secondary analysis of the randomized, cooperative group phase III trail, RTOG 92-02. The purpose of this analysis was to evaluate distant metastasis and general clinical treatment failure as surrogate endpoints for prostate cancer – specific survival by using the surrogacy criteria of Prentice (10). Both distant metastasis and general clinical treatment failure at 3 years were consistent with all four of Prentice's criteria for being surrogate endpoints for prostate cancer – specific survival at 10 years. Although prostate cancer – specific survival was not statistically significantly different between treatment arms at 5 years $(P = .08)$, both endpoints were consistent with Prentice's three remaining criteria.

Table 6. Prentice's criteria 3 and 4 for the 5-year landmark analysis (n = 1178)*

Variable	Coefficient	P valuet
$S(t) = DM$		
DM (H ₀ : $\beta_1 = 0$)	1.93	.004
DM \times TRT (H ₀ : $\beta_2 = 0$)	0.27	.17
Non-DM \times TRT (H ₀ : $\beta_2 = 0$)	-0.59	.14
$S(t) = GCTF$		
GCTF $(H_0: \beta_1 = 0)$	2.61	.02
GCTF \times TRT (H ₀ : $\beta_2 = 0$)	-0.10	.58
Non-GCTF \times TRT (H ₀ : $\beta_2 = 0$)	-0.24	-74

* The following proportional model was used to validate the criteria. Prentice's third criterion, that the surrogate endpoint has a statistically significant impact on the true endpoint, can be tested by $H_0: \beta_1 = 0$. Prentice's fourth criterion can be tested by $H_0: \beta_2 = 0$ and $H_0: \beta_3 = 0$.

$$
\ln \frac{b(t)}{b_0(t)} = \beta_1 \times S(t) + \beta_2 \times \text{TRT} \times S(t) + \beta_3 \times \text{TRT} \times [1 - S(t)],
$$

where $h(t)$ = hazard rate of prostate cancer-specific death at time t; $S(t)$ = surrogate endpoint (failure = 1 and nonfailure = 0); $1 - S(t)$ = not having surrogate endpoint failure; TRT = treatment arm analyzed (control = 0 and $experimental = 1$; $DM = distant metastasis$; $GCTF = general clinical treat$ ment failure; β_1 = coefficient for surrogate endpoint effect; β_2 = coefficient for treatment effect; β_3 = coefficient for the interaction between the not having a failure for surrogate endpoint and treatment.

† Chi-square test statistic. All statistical tests were two-sided.

These results suggest that distant metastasis and general clinical treatment failure at 3 and 5 years may be useful surrogate endpoints for prostate cancer – specific survival.

An increasing number of randomized clinical studies of modern local therapies for patients with clinically localized prostate cancer that found improved biochemical outcomes have been reported (14-17). However, relatively few prostate cancer trials have demonstrated statistically significant differences in prostate cancer-specific survival or overall survival $(1-4)$. Because of the long natural history of prostate cancer, randomized trials designed to detect statistically significant differences in mortality often are prohibitive in terms of size, duration, and cost. In contrast, colorectal cancers have a relatively rapid natural history, and many trials of adjuvant therapy have demonstrated statistically significant survival differences $(18, 19)$. By using meta-analysis of multiple randomized experiments, investigators have demonstrated that disease-free survival is a valid surrogate endpoint for overall survival in colorectal cancer $(20-22)$. The identification and rigorous validation of surrogate endpoints for overall or disease-specific survival are of great utility in clinical cancer research because the use of surrogate endpoints substantially decreases the size and duration of clinical trials, allowing more rapid prospective testing of hypotheses and potentially accelerating development of improved cancer treatment.

This study had several limitations arising from the challenges of proving Prentice's key fourth criterion and the fact that data from only one trial were available for this analysis. Statistical methods for establishing and validating a surrogate endpoint, such as Prentice's criteria, must be applied with caution and with an appreciation of their limitations (23). In particular, the failure to reject the critical Prentice's fourth criterion is not necessarily definite evidence that the criterion holds. Furthermore, surrogate endpoint analyses are ideally conducted with meta-analytic approaches and include data from many randomized trials. Unfortunately, few randomized trials of treatment for localized prostate cancer have yet observed statistically significant differences in survival endpoints, limiting the opportunity for comprehensive meta-analysis.

Although this analysis included patients from only one clinical trial, the findings are strengthened because they derive from data on a large group of patients who were treated in a multiinstitutional setting under the direction of a large cooperative group. The RTOG 92-02 trial was conducted with standardized data collection, treatment, and follow-up. All patients were treated between June 26, 1992, and April 15, 1995, with subsequent follow-up, all occurring within the PSA era. Although this study is a retrospective secondary analysis, data are from a prospective randomized trial, which should remove potential sources of selection or other accidental bias. The experimental treatment that was evaluated in the RTOG 92-02 trial has become a standard of care, and the experimental arm of the RTOG 92-02 trial is now the standard arm in the successor combined modality trial, RTOG 05-21, for high-risk prostate cancer. The RTOG 05-21 trial is a randomized comparison between long-term androgen deprivation therapy and external beam radiation therapy with or without adjuvant docetaxel-based chemotherapy. The continued prospective use of data from the experimental arm of the RTOG 92-02 trial indicates that the analysis is contemporary and relevant to ongoing clinical research.

Previous attempts at identification of a surrogate endpoint for prostate cancer survival have focused on the kinetics of increasing posttreatment serum PSA levels. Many studies have found that rapid PSA increases, or short posttreatment PSA doubling times, after local therapy are associated with the development of metastatic disease and prostate cancer-specific mortality (24-28). D'Amico et al. (26) , by use of a large database of retrospective patient data, demonstrated that a short PSA doubling time of less than 3 months meets statistical criteria for being a surrogate endpoint for prostate cancer-specific survival. However, metrics such as PSA doubling time and PSA velocity can be confounded by the methods and frequency of PSA measurement and can be difficult to systematically assess.

Data from the RTOG 92-02 trial have been the subject of a previous secondary analysis that evaluated posttreatment PSA doubling time as a potential surrogate endpoint for prostate cancer survival (29). Valicenti et al. (29) confirmed that treatment was statistically significantly associated with PSA doubling times that were shorter than 6, 9, and 12 months and that these PSA doubling times were statistically significantly associated with prostate cancerspecific survival, meeting Prentice's second and third criteria. However, prostate cancer-specific survival was not independent of treatment for PSA doubling times longer than any cut point in the study (eg, $P = .89$ for PSA doubling time of <6 months but $P = .014$ for PSA doubling time of ≥6 months). Because PSA doubling time had been previously analyzed in this dataset and Prentice's fourth criteria for surrogacy was rejected and also because PSA kinetics have not been validated as an independent predictor of prostate cancer survival, we chose not to include PSA kinetics as a component of a surrogate endpoint definition in this study.

In patients with metastatic, androgen-independent prostate cancer, Petrylak et al. (30) reported that several metrics, including 3-month declines in the PSA level of 20% – 40%, 2-month decline in the PSA level of 30%, and PSA velocities at 2 and 3 months after chemotherapy, did meet surrogacy criteria for survival in an analysis of the Southwest Oncology Group 99-16 trial. The rate or

degree of decline in the serum PSA level after initiation of androgen deprivation therapy would not likely constitute a viable endpoint for patients with hormonally naive, clinically localized prostate cancer, such as the patients in the RTOG 92-02 trial; therefore, PSA declines were also not analyzed as potential survival surrogates in this study.

The potential surrogate endpoints that were selected for this study included distant metastasis, which is a straightforward and objective finding in patients with progressive prostate cancer and has been associated with increased prostate cancer-specific death (31). However, the detection of distant metastasis may be subject to variation in how often imaging studies (eg, bone scan or other x-ray studies) are obtained, with some clinicians using asymptomatic PSA changes to request an imaging study and others using the development of symptomatic disease progression, such as the onset of bone pain. To address this potential concern, we sought an additional candidate surrogate endpoint that could capture a wider range of indicators of clinically significant disease progression that might suggest fatal disease. We identified a composite endpoint, general clinical treatment failure, that has been applied previously $(7,8)$ and that incorporates clinical (identification of local, regional, or distant failure), biochemical (measurement of a substantial elevation in the PSA level of more than 25 ng/mL), and treatment-related (initiation of androgen deprivation therapy) indicators of treatment failure that are likely to be associated with fatal prostate cancer. Although there is considerable overlap between the two candidate surrogate endpoints evaluated in the study, the definition that we used for general clinical treatment failure broadens the failure criteria by including local and regional disease recurrence at any clinically apparent site and the stringent biochemical treatment failure criterion of a PSA level of more than 25 ng/mL. In addition, this endpoint captures other factors judged by the treating physician as clinically significant enough to warrant systemic treatment by initiating androgen deprivation therapy after the completion of curative intent local therapy. In addition to distant metastasis, we feel that the definition for the composite endpoint, general clinical treatment failure, is also a straightforward measure that can be applied to many patients with prostate cancer at various stages and may more completely capture treatment failures that are likely to be associated with increased prostate cancer-specific death.

Validation of distant metastasis and/or general clinical treatment failure, as defined in this study, as surrogate endpoints for prostate cancer-specific survival clearly requires additional study of other patient data. Although few randomized trials of treatment for localized prostate cancer have demonstrated statistically significant improvements in prostate cancer-specific survival or overall survival, certain trials conducted by the Southwest Oncology Group, the European Organization for Research and Treatment of Cancer, and other groups contain data that may be useful for validation of these factors as surrogate endpoints for prostate cancer-specific survival and that could potentially support a meta-analysis similar to the approach used for colorectal cancer $(20-22)$. This study represents an initial attempt to explore intermediate outcome criteria as possible surrogate endpoints for prostate cancer-specific survival, and its results will require validation in prospective trials, preferably involving several different treatment types, to determine whether the findings can be applied only to combination treatment

with androgen deprivation therapy and external beam radiotherapy or whether they can be more broadly applied to other treatment strategies, such as those that incorporate different local and systemic therapies. The potential surrogate endpoints, distant metastasis and general clinical treatment failure, should be further and prospectively evaluated in future clinical trials along with overall and prostate cancer-specific survival endpoints.

References

- 1. Pilepich MV, Krall JM, Al-Sarraf M, et al. Androgen deprivation with radiation therapy compared with radiation therapy alone for locally advanced prostatic carcinoma: a randomized comparative trial of the Radiation Therapy Oncology Group. *Urology*. 1995;45(4):616-623.
- 2. Pilepich MV, Caplan R, Byhardt RW, et al. Phase III trial of androgen suppression using goserelin in unfavorable-prognosis carcinoma of the prostate treated with definitive radiotherapy: report of Radiation Therapy Oncology Group Protocol 85-31. *J Clin Oncol.* 1997;15(3):1013-1021.
- 3. Bolla M, Gonzalez D, Warde P, et al. Improved survival in patients with locally advanced prostate cancer treated with radiotherapy and goserelin. *N Engl I Med.* 1997;337(5):295-300.
- 4. Hanks GE, Pajak TF, Porter A, et al. Phase III trial of long-term androgen deprivation after neoadjuvant hormonal cytoreduction and radiotherapy in locally advanced carcinoma of the prostate: the Radiation Therapy Oncology Group Protocol 92-02. *J Clin Oncol.* 2003;21(21):3972-3978.
- 5. Horwitz EM, Bae KY, Hanks GE, et al. Ten-year follow-up of RTOG 92-02: a phase III trial of the duration of elective androgen deprivation in locally advanced prostate cancer. *J Clin Oncol*. 2008;26(15): 2497-2504.
- 6. Horwitz EM, Bae K, Porter A, et al. Ten-year follow-up of RTOG 92-02: a phase III trial of the duration of elective androgen deprivation in locally advanced prostate cancer. *Int J Radiat Oncol Biol Phys.* 2006;66:S13. Abstract 22 .
- 7. Taylor JM, Griffith KA, Sandler HM. Definitions of biochemical failure in prostate cancer following radiation therapy . *Int J Radiat Oncol Biol Phys.* 2001:50(5):1212-1219.
- 8. Thames H, Kuban D, Levy L, et al. Comparison of alternative biochemical failure definitions based on clinical outcome in 4839 prostate cancer patients treated by external beam radiotherapy between 1986 and 1995 . *Int J Radiat Oncol Biol Phys.* 2003 ; 57 (4): 929 – 943 .
- 9. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol.* 1983;1(11):710-719.
- 10. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med.* 1989;8(4):431-440.
- 11. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York, NY: John Wiley & Sons; 1980.
- 12. Gray RJ . A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat*. 1988;16:1141-1154.
- 13. Cox D. Regression models and life tables. *J Royal Stat Soc.* 1972;34: 187-229.
- 14. Kuban DA, Tucker SL, Dong L, et al. Long-term results of the M.D. Anderson randomized dose-escalation trial for prostate cancer . *Int J Radiat Oncol Biol Phys.* 2008;70(1):67-74.
- 15. Zietman AL, DeSilvio ML, Slater JD, et al. Comparison of conventionaldose vs high-dose conformal radiation therapy in clinically localized adenocarcinoma of the prostate: a randomized controlled trial . *JAMA .* 2005;294(10):1233-1239.
- 16. Bolla M, van Poppel H, Collette L, et al. Postoperative radiotherapy after radical prostatectomy: a randomised controlled trial (EORTC trial 22911) . Lancet. 2005;366(9485):572-578.
- 17. Thompson IM Jr, Tangen CM, Paradelo J, et al. Adjuvant radiotherapy for pathologically advanced prostate cancer: a randomized clinical trial . *JAMA*. 2006;296(19):2329-2335.
- 18. Moertel CG, Fleming TR, Macdonald JS, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *N Engl J Med.* 1990;322(6):352-358.
- 19. Wolmark N, Rockette H, Fisher B, et al. The benefit of leucovorinmodulated fluorouracil as postoperative adjuvant therapy for primary colon cancer: results from the National Surgical Adjuvant Breast and Bowel Project protocol C-03. *7 Clin Oncol.* 1993;11(10):1879-1887.
- 20. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoint in meta-analyses of randomized experiments . *Biostatistics.* 2000;1(1):49-67.
- 21. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol*. 2007;25(33):5218-5224.
- 22. Sargent DJ, Wieand HS, Haller DG, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol.* 2005;23(34):8664-8670.
- 23. Baker SG . Surrogate endpoints: wishful thinking or reality? *J Natl Cancer* Inst. 2006;98(8):502-503.
- 24. Okotie OT, Aronson WJ, Wieder JA, et al. Predictors of metastatic disease in men with biochemical failure following radical prostatectomy. *J Urol.* 2004;171(6):2260-2264.
- 25. Freedland SJ, Humphreys EB, Mangold LA, et al. Risk of prostate cancerspecific mortality following biochemical recurrence after radical prostatectomy. $JAMA$. 2005;294(4):433-439.
- 26. D'Amico AV, Cote K, Loffedo M, Renshaw AA, Schultz D. Determinants of prostate cancer-specific survival after radiation therapy for patients with clinically localized prostate cancer. *J Clin Oncol.* 2002;20(23):4567-4573.
- 27. Hanlon AL, Diratzouian H, Hanks GE. Posttreatment prostate-specific antigen nadir highly predictive of distant failure and death from prostate cancer. *Int J Radiat Oncol Biol Phys.* 2002;53(2):297-303.
- 28. Albertsen PC, Hanley JA, Penson DF, Fine J. Validation of increasing prostate specific antigen as a predictor of prostate cancer death after treatment of localized prostate cancer with surgery or radiation. *J Urol.* 2004;171(6): 2221-2225.
- 29. Valicenti RK, DeSilvio M, Hanks GE, et al. Posttreatment prostaticspecific antigen doubling time as a surrogate endpoint for prostate cancerspecific survival: an analysis of Radiation Therapy Oncology Group protocol 92-02. *Int J Radiat Oncol Biol Phys.* 2006;66(4):1064-1071.
- 30. Petrylak DP, Ankerst DP, Jiang CS, Tangen CM, Hussain MHA, Lara PN Jr. Evaluation of prostate-specific antigen declines for surrogacy in patients treated on SWOG 99-16. *J Natl Cancer Inst.* 2006;98(8): 516 – 521.
- 31. Zeitman AL, Dallow KC, McManus PL, Heney NM, Shipley WU. Time to second prostate cancer specific failure is a surrogate endpoint for prostate cancer death in a prospective trial of therapy for localized disease. *Urology*. 1996;47(2):236-239.

Funding

Supported by RTOG U10 CA21661, CCOP U10 CA37422, and Stat U10 CA32115 from the National Cancer Institute.

Notes

Presented at the 49th Annual Meeting of the American Society for Therapeutic Radiology and Oncology, Los Angeles, CA, 2007.

The contents of this article are the sole responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute. The sponsor(s) or funding agency had no role in the design of the study; the collection, analysis, and interpretation of the data; the decision to submit the manuscript for publication; and the writing of the manuscript.

Dr H. M. Sandler has been a consultant for AstraZeneca.

Manuscript received May 5, 2008; revised November 20, 2008; accepted December 5, 2008.