Time to biochemical failure and PSA doubling time as surrogates for prostate cancer-specific mortality: evidence from the TROG 96.01 randomized controlled trial

Authors:

Professor James W. Denham, FRANZCR ^{*1}, Allison Steigler BMath¹, Chantelle Wilcox BBiomedSc¹, David S. Lamb FRANZCR², Professor David Joseph FRANZCR³, Chris Atkinson FRANZCR⁴, John Matthews FRANZCR⁵, Keen-Hun Tai FRANZCR⁶, Professor Nigel A. Spry FRANZCR³, Assoc. Professor David Christie FRANZCR⁷, Paul S. Gleeson¹, Peter B. Greer PhD⁸ Professor Catherine D'Este PhD⁹, on behalf of the Trans-Tasman Radiation Oncology Group 96.01 trial participants

- ¹ School of Medicine and Public Health, The University of Newcastle, Newcastle, New South Wales, Australia
- ² Wellington Cancer Centre, Wellington, New Zealand
- ³ Sir Charles Gairdner Hospital, Perth, Western Australia, Australia
- ⁴ Christchurch Hospital, Christchurch, New Zealand
- ⁵ Auckland Hospital, Auckland, New Zealand
- ⁶ Peter MacCallum Cancer Institute, Melbourne, Victoria, Australia
- ⁷ Premion, Tugun, Queensland, Australia
- ⁸ School of Physical and Mathematical Sciences, The University of Newcastle, Newcastle, New South Wales, Australia
- ⁹ Centre for Clinical Epidemiology and Biostatistics, The University of Newcastle, Newcastle, New South Wales, Australia

*Corresponding author
Prof James W Denham
Calvary Mater Newcastle Hospital
Locked Bag 7
Hunter Region Mail Centre NSW 2310
Ph. +61 2 4921 1174
Fax +61 2 4921 1465
Email Jim.Denham@newcastle.edu.au

Manuscript category: Original report

Body text word count: 4918

Abstract word count: 305

Reference count: 33

Table count: 2

Figure count: 6

Running head: Surrogates for prostate cancer-specific mortality

Subject area: Cancer Prevention/Control/Trials

Keywords: Biomarker & Diagnosis/Prevention

Hormones, Hormone Receptors & Therapy

Prostate Cancer

ABSTRACT

Background: Surrogate endpoints for prostate cancer-specific mortality (PCSM) following curative primary treatment are not well established. We sought to assess time to biochemical failure (TTBF) and prostate-specific antigen (PSA) doubling time after failure as candidates.

Methods: PSA and survival data from the Trans Tasman Radiation Oncology Group (TROG) 96.01 trial (Australian New Zealand Clinical Trials Registry, number ACTRN12607000237482) were used to assess the surrogate candidates. Between June 28, 1996 and February 16, 2000, 802 eligible men with locally advanced prostate cancer (PC) were randomly allocated prostatic irradiation alone (XRT), 3 or 6 months of short term maximal androgen deprivation (STAD) prior to and during radiation. Successful surrogates were required to satisfy the Prentice criteria and to predict the trial result. The TROG 96.01 Trial is registered with the Australian New Zealand Clinical Trials Registry, number ACTRN12607000237482.

Results: The trial results indicated that 6 months STAD significantly reduced PCSM when compared with radiation alone, but 3 months STAD did not. Relative to XRT the hazard ratio of PCSM from randomization in the 3 month STAD trial arm was 0.95 (95% confidence interval [CI] = 0.63 to 1.41; $p=0.79$), and in the 6 month arm was 0.56 (CI = 0.36 to 0.88; $p=0.01$). PSA doubling time predicted the trial result and satisfied all four Prentice criteria at the <12 and <15 months cutpoints, with proportion of treatment effect (PTE) ratios between 0.36 and 0.56. TTBF performed better, predicting the trial result and

satisfying all four Prentice criteria at cutpoints <1.5, <2 and <2.5 years, with PTE ratios between 0.45 and 0.64.

Conclusion: This study provides proof of the principle that TTBF and PSA doubling time can be useful as surrogate endpoints. TTBF, in particular, and PSA doubling time are promising surrogate endpoint candidates that now require evaluation in multi-trial meta-analytic studies before use in clinical trials.

Funding: National Health and Medical Research Council (Australia) (Project Grant Applications 9936572, 209801, 455520); Hunter Medical Research Institute (Newcastle, Australia); AstraZeneca (Sydney, Australia); and Schering-Plough, (Sydney, Australia).

INTRODUCTION

Progress in the curative management of localized prostate cancer (PC) is seriously hampered by the need for large scale trials with prolonged follow-up. The identification of surrogate endpoints for prostate cancer specific mortality (PCSM) is therefore one of the highest priorities in clinical PC research.

Several studies have shown that parameters based on longitudinal changes in serum levels of prostate specific antigen (PSA) are potent prognostic variables. In particular rapid PSA doubling time after failure of curative treatment (1-7) and shorter time to biochemical failure (TTBF) (8-12) are associated with distant failure and predict early PC death whether curative treatment is by surgery or radiation. It must be pointed out however, that although a certain prognostic variable is associated with a strong likelihood of subsequent PC death, it cannot be used as a surrogate endpoint in clinical trials unless it can be shown that a specific range of values of the variable predict a very strong likelihood of PC death, while the remaining values do not. Encouragement that PSA based endpoints have a role as surrogate endpoints for PCSM following the failure of primary treatment has come from recent trials (13-14), and in community practice (1). Progress in identifying suitable surrogate endpoint candidates in the curative treatment of localized PC has been less encouraging. Data from only one trial has been reported (15) and the surrogate candidate selected, PSA doubling time, did not satisfy all four Prentice criteria, which is the usual test applied to surrogate candidates in single controlled trials (16).

The Trans-Tasman Radiation Oncology Group (TROG) 96.01 trial has provided an excellent opportunity to test the value of both PSA doubling time

and TTBF as surrogate candidates for PCSM following primary therapy. In this trial 802 evaluable patients receiving radiotherapy for locally advanced, non-metastatic PC were randomized to receive 0, 3 or 6 months neo-adjuvant maximal androgen deprivation between 1996 and 2000. Recently updated data confirm preliminary findings published in 2005 (17) that 6 months STAD significantly reduced PCSM when compared with radiation alone, but 3 months STAD did not. We reasoned that a surrogate endpoint would need to be a very good candidate indeed if it were to provide an accurate prediction of the trial result to date **and** satisfy all four Prentice criteria.

These criteria require four conditions to be satisfied: (1) treatment is prognostic for the true endpoint; (2) treatment is prognostic for the surrogate endpoint; (3) the surrogate is prognostic for the true endpoint; and (4) the surrogate mediates the effect of treatment on the true endpoint. Several authors (1; 18-19) have drawn attention to the fact that an inability to account for the impact of secondary therapeutic intervention (STI) on survival compromises efforts to determine firstly, that primary treatment really is prognostic for the true endpoint (PCSM) (i.e. Prentice criterion 1) and secondly, that a surrogate candidate mediates the effect of the primary treatment on the true endpoint (PCSM) (i.e. Prentice criterion 4). Two factors need to be taken into account: (a) the time between biochemical failure and initiation of STI, and (b) the efficacy of STI once commenced. The availability of detailed STI data from the 96.01 trial that define its timing after failure of primary therapy, its type and its efficacy has therefore allowed us to address these important issues.

METHODS

Study Design

The trial opened in 1996 following ethical approval at participating Australian and New Zealand sites (17). 802 eligible patients receiving radiotherapy (66Gy delivered using 33 daily fractions) for locally advanced prostate cancer (PC) were randomized to receive 0, 3 or 6 months short term maximal androgen deprivation (STAD) prior to and during radiation, after providing written informed consent and stratification by age (<70/70-80/>80 years), stage (T2b,c/T3,T4), tumor differentiation (well/moderate/poor) and initial PSA level (<20/≥20 ng/mL). STAD comprised goserelin¹ 3.6mg given every month sub-cutaneously and flutamide² 250mg given three times a day orally. Three months of STAD commenced 2 months prior to radiation and 6 months commenced 5 months prior.

Endpoints

Endpoints used for this report were: (a) time from randomization to death due to PC or in the context of progressive PC; (b) time from the end of radiation to biochemical failure (TTBF); and (c) PSA doubling time.

Although the ASTRO method was used prior to 2005, biochemical failure was defined in this report using a variation of the more accurate Phoenix method (i.e. time from end of radiotherapy to a rise in PSA of 2 ng/mL above the post-treatment nadir value) (20-23). In the variation, failure is confirmed by further rising PSA values over a 6 month period (24).

PSA doubling time estimates were derived using PSA values between biochemical failure and initiation of secondary therapeutic intervention (STI), or

¹ Zoladex® AstraZeneca Pty Ltd, Sydney Australia

last contact if STI had not been commenced. Categorization and estimation of the doubling times of the various types of PSA ascent observed were achieved using a best-fitting spline program designed by DB and executed on Stata Version 9. In this procedure log PSA ascent data were plotted against time from biochemical failure. Best fit regression lines between the biochemical failure value and a progressively increasing number of values up to initiation of STI or last contact were extrapolated back to the log PSA axis. A non-exponential PSA ascent was defined by an intercept estimate at this axis whose confidence intervals did not include the PSA at biochemical failure. A single exponential ascent was therefore defined by an intercept estimate close to the PSA at biochemical failure.

Follow-up

The trial's routine follow-up schedule has been described previously (17). Rising PSA levels were routinely monitored at 2-3 monthly intervals. Investigations, including prostate biopsy, computed tomography scanning, chest x-ray and isotope bone scan, were mandated if symptoms or signs suggested a need, or if the PSA level had reached 20 ng/mL without symptoms or signs suggestive of recurrence.

² Eulexin® Schering-Plough Pty Ltd, Sydney Australia

Analytic Strategy

Prior to assessment of TTBF and PSA doubling time as surrogate candidates for PCSM, errors in the measurement of these variables were estimated. Estimation of these errors is described in the on-line supplement to this report.

The relationships between key pre-treatment variables (age, clinical stage, Gleason score and initial PSA [iPSA]) and treatment arm and the surrogate candidates were then investigated in logistic regression models to establish the potential for interactions to confound subsequent surrogacy models. In these models cutpoints of TTBF < 2 years and PSA doubling time < 6 months were selected as these are established predictors of early PC death.

Finally, before constructing surrogacy models, another important confounding factor mentioned in the Introduction, STI, was addressed. To do this, the influence of treatment arm on: 1- time to initiation of STI from biochemical failure, 2- time to PC death from biochemical failure, and 3- time to PC death from commencement of STI, was assessed in univariate and multivariate analyses. In addition, the influence of treatment arm, and the surrogate candidates on the type of STI used were compared to rule out biases introduced by differences in the efficacy of the various types of STI used.

Because the data used originate from a randomized controlled trial where treatment arm has influenced time to PC death, the Prentice criteria (16) were used to assess various cutpoints of TTBF and PSA doubling time as surrogate endpoints for PCSM.

The criteria were addressed as follows:

1. Criterion 1 : by a proportional hazards model (PHM) of time to PC death by trial treatment arm, adjusting for the pre-treatment stratification variables.
2. Criterion 2 : by PHMs of (a) TTBF and (b) PSA doubling time by trial treatment arm, adjusting for the pre-treatment stratification. In models assessing various cutpoints of TTBF, an event was defined as a biochemical failure up to the cutpoint time. The cutpoints selected for testing were: <1, <1.5, <2, <2.5, <3, <4 and <5 years. All other men were censored at the time of biochemical failure or last follow-up if biochemical failure did not occur. In models assessing various cutpoints of PSA doubling time, an event was defined as a PSA doubling time value up to the cutpoint value, and its timing was defined as the time of biochemical failure. The cutpoints selected for testing were: <3, <6, <9, <12, <15, <18 and <24 months. All other men were censored as described above. Men who experienced biochemical failure, but whose PSA doubling time could not be estimated (usually because STI was commenced immediately), had a hazard of PCSM similar to those with PSA doubling times in the range of 3-6 months and were therefore included as an event at all PSA doubling time cutpoints except <3 months.
3. Criterion 3 : by PHMs of time to PC death by TTBF and PSA doubling time treated as time dependent covariates using the cutpoints described above, adjusting for the pre-treatment stratification variables.
4. Prentice criterion 4 : in PHMs of time to PC death including the candidate surrogate variable treated as a time dependent covariate, the trial treatment arm, and the pre-treatment stratification variables. Proportion

of treatment effect (PTE) mediated by the surrogate endpoint was determined using the formula derived by Lin et al, namely 1 minus the ratio of the regression coefficient of the 6 month treatment arm in models adjusted for the surrogate endpoint, and the regression coefficient of the 6 month arm in models not adjusted for the surrogate (25). PTE is a quantitative measure of how much the treatment effect on clinical outcome is explained by the surrogate outcome. For a perfect surrogate PTE is 1, however in practice there is no perfect surrogate so a “high” PTE (0.75 or greater) is suitable in clinical trials where secondary therapies do not prolong survival. Where secondary therapies do prolong survival, as can occur in prostate cancer trials, lower PTE values may be acceptable. PTE confidence intervals have not been derived. For discussion of their derivation the reader is referred to Chen et al (26).

To assess the robustness of the results we conducted a series of sensitivity analyses: (1) models using randomization and end of radiotherapy as the starting point for all time to event measures (presented in Results); (2) PSA doubling time models excluding men with missing PSA doubling times due to insufficient PSA's (Table S3a); (3) PSA doubling time models with all PSA doubling times recalculated using the PSA value prior to biochemical failure in addition to post failure values (to enable PSA doubling time to be calculated for all men with biochemical failure) (Table S3b); and (4) models in analysis (3) repeated using PSA DT estimates based on a maximum of 6, 12 and 24 months of PSA values (Table S3c,d,e).

Statistical Methods

Time to all endpoint events was derived by the Kaplan-Meier technique (27). In cause-specific survival analyses patients lost to follow-up, or who died of intercurrent, unrelated medical problems due to unknown causes were censored at time of last follow-up or death if they had remained continuously free of any sign (physical or biochemical) of PC since treatment. Unadjusted and adjusted hazard ratios and 95% confidence intervals were obtained from Cox PHMs (28).

In models requiring adjustment for pretreatment factors: age at randomization (continuous), Gleason score (2-6, 7 or 8-10), initial PSA (continuous), and T stage (T2b, T2c or T3/T4) were used.

Significance was determined by the Wald test in Cox analyses. All analyses involving trial arm were conducted on an “intention to treat” basis. Two sided probability levels below 0.05 were considered statistically significant. All endpoint data were audited between December 1, 2006 and August 1, 2007. Closeout date for data inclusion was June 30, 2007.

Role of the funding source

No sponsor or funding source was involved in the study design or in the collection, analysis, and interpretation of the data, the writing of the report or the decision to submit for publication. JD, AS and CW had full access to all of the raw data. The corresponding author had the final responsibility to submit for publication.

RESULTS

The pre-treatment prognostic characteristics of the 802 eligible men randomized in the 96.01 trial are presented by trial arm in Table 1. Table 2 shows the distribution of time to biochemical failure and PSA doubling times by trial arm for the 436 men experiencing biochemical failure. A breakdown according to number of biochemical and/or clinical failures (n=454), secondary therapeutic intervention (STI) (n=342), PC deaths (n=125) and losses to follow-up (n=16) up to June 30, 2007 is shown in Figure 1. An additional 6 patients (2 unknown deaths and 4 losses to follow-up) have been reclassified as PC deaths due to histories of progressive PC.

PSA doubling time was estimated in 360 (83%) of 436 men who experienced biochemical failure. The Venn diagram in Figure 2 displays median and range values of PSA doubling time (months) and TTBF (years) for subgroups of men based on their history of biochemical failure, clinical failure and STI. It can be seen that PSA doubling times were more rapid and TTBF shorter in men who received STI than those who did not, regardless of whether anatomical site of failure was diagnosed or not.

Further details regarding the derivation of TTBF and PSA doubling time are shown in the on-line supplement.

Using multiple logistic regression (Table S1), risk factors for cutpoint TTBF < 2 years were higher initial PSA, higher stage and higher Gleason score. All four trial stratification factors were predictive of PSA doubling time < 6 months (Table S2). When treatment arm was factored into these models the only finding to change was that men assigned 6 months STAD were at decreased risk of TTBF compared to men assigned to the other treatment arms. No interactions were identified.

Of the total 342 men who received STI, breakdown according to treatment arm was 146 (43%) XRT alone, 108 (31%) 3 month STAD and 88 (26%) 6 month STAD. Androgen deprivation medications were the most common type of first STI employed (n=297 [87%]) with the remainder treated by orchiectomy, strontium or radiotherapy.

To satisfy Prentice criteria 1 and 4 that treatment is prognostic for the true endpoint and that a surrogate candidate mediates the effect of treatment on this endpoint, it was first necessary to exclude confounding attributable to the type, timing and efficacy of STI in prolonging survival after failure.

- (a) Treatment arm was **not** found to influence time to prostate cancer death from biochemical failure in univariate (Figure 3A) or multivariate analysis (not shown). For 3 months STAD compared to XRT alone, HR = 1.47, 95% CI = 0.99-2.19, p=0.06, and for 6 months STAD HR = 1.02, 95% CI = 0.65-1.60, p=0.95.
- (b) The probabilities of initiating STI after biochemical failure were very similar in all three trial arms in univariate (Figure 3B) and multivariate analysis (not shown). For 3 months STAD compared to XRT alone, HR = 1.28, 95% CI = 0.99-1.64, p=0.06, and for 6 months STAD HR = 1.10, 95% CI = 0.85-1.44, p=0.47.
- (c) Time to cancer death from initiation of STI was not influenced by trial arm either in univariate (Figure 3C) or multivariate analysis (not shown). For 3 months STAD compared to XRT alone, HR = 1.33, 95% CI = 0.89-2.01, p=0.17, and for 6 months STAD HR = 0.89, 95% CI = 0.55-1.44, p=0.64.
- (d) The choice of STI was not influenced by trial arm or increasing values of the two candidate surrogate endpoints (data not shown).

Having concluded that the type, timing and efficacy of STI was influenced minimally by trial arm, proportional hazards models without adjustments for STI were constructed to determine whether PSA doubling time and TTBF, at various cutpoints, satisfied the Prentice criteria.

Assessment of Prentice Criteria is summarized in Tables 3 and 4:

Prentice criterion 1 ("Treatment is prognostic for the true endpoint [i.e. PCSM]") was satisfied for the 6 month STAD arm of the trial but **not** by the 3 month arm in both univariate (Figure 4) and multivariate analyses. Compared to the control arm (i.e. radiotherapy alone), the hazard of PCSM from randomization was significantly reduced by 6 month STAD (HR = 0.56, 95% CI = 0.36 to 0.88; p=0.01) but not by 3 month STAD (HR= 0.95, 95% CI = 0.63 to 1.41; p=0.79), and from the end of radiotherapy by 6 month STAD (HR = 0.62, 95% CI = 0.39 to 0.98; p=0.04) but not by 3 month STAD (HR = 1.01, 95% CI = 0.67 to 1.52; p=0.96).

Prentice criterion 2 ("Treatment is prognostic for the surrogate endpoints") was satisfied by TTBF at all cutpoints for the 6 month STAD trial arm but in contrast to the trial result, at the <1, <3, <4 and <5 year cutpoints for the 3 month STAD arm in models starting from randomization and <3, <4 and <5 year cutpoints in models from the end of radiotherapy. Criterion 2 was also satisfied by PSA doubling time for the 6 month STAD arm at the <9, <12, <15, <18 and <24 month cutpoints in models from randomization but only <12, <15, <18 and <24 month cutpoints in models from the end of radiotherapy. In contrast to the trial result, criterion 2 was satisfied for the 3 month arm at <18 and <24 months in models from both starting times. In summary the trial result was therefore predicted successfully **and** criterion 2 satisfied by (a) TTBF at all

cutpoints between <1.5 and <2.5 years and (b) PSA doubling time at the <12 and <15 month cutpoints in models from both starting times.

Prentice Criterion 3 (“The surrogate is prognostic for the true endpoint [i.e. PCSM]”) was satisfied by both TTBF and PSA doubling time at **all** cutpoints examined and from both starting times.

Prentice criterion 4 (“The surrogate mediates the influence of initial treatment on the true endpoint [i.e. PCSM]”) was satisfied by TTBF at all cutpoints for the 6 month STAD arm of the trial in models from both starting times. PTE values ranged between 0.50 and 0.78 in models starting from randomization and 0.45 and 0.83 from the end of radiotherapy. PSA doubling time satisfied criterion 4 at the <3, <12, <15, <18 and <24 months cutpoints from both starting times. However PTE values were lower than those estimated for TTBF cutpoints, ranging between 0.34 and 0.64 in models starting from randomization and 0.10 and 0.68 in models from the end of radiotherapy.

The sensitivity analyses performed to assess the impact on the results of varying the ranges of PSA values used to derive PSA doubling time estimates (Table S3) provide confirmation that PSA doubling time cutpoints <12 and <15 months are good candidates. Moreover they suggest that the <9 month cutpoint may also be a candidate that should be examined in meta-analytic studies.

In summary all **four** Prentice criteria were satisfied **and** the trial result was predicted successfully at all TTBF cutpoints between 1.5 and 2.5 years, and by PSA doubling time at the <12, <15 month cutpoints. Best fits to the trial result were provided by TTBF at the <1.5, <2 year and PSA doubling time <12 month cutpoints. The predictive value of the TTBF <2 years, and PSA doubling time <12 months surrogate candidates is depicted in the cancer specific survival

plots presented in Figure 5. For TTBF cutpoint <2 years, HR = 10.64, 95% CI = 6.27-18.04, $p < 0.0001$, and for PSA doubling time cutpoint <12 months HR = 11.82, 95% CI = 7.85-17.80, $p < 0.0001$.

DISCUSSION

The surrogacy study reported herein, has provided important proof of the principle that TTBF and PSA doubling time are prognostic variables potent enough to produce surrogate endpoint candidates for PCSM that satisfy rigorous statistical testing within a randomized controlled trial. In particular TTBF in a cutpoint band between <1.5 and <2.5 years, and PSA doubling time in a cutpoint band between <12 and <15 months have predicted the results of the trial, including all three of its treatment arms, almost perfectly. The results are robust, and as shown by our sensitivity analyses, are applicable regardless of whether TTBF is measured from the end of radiotherapy (as is current standard practice) or from randomization. It must also be pointed out that the 96.01 trial, whose data this study draws from, produced a result that was **not** influenced by secondary therapy. It was a conclusive one, even though the trial was of modest size, involving 800 men, of whom 131 were deaths due to prostate cancer.

The questions that these results raise are; 1- how they compare with results from previous studies, and 2- whether the results can be used in trials addressing new treatments in locally advanced cancer or other scenarios where cure is hoped for.

The first study to seriously address the surrogacy issue was reported by D'Amico in 2003 and focused on PSA doubling time as a candidate. Since data came from a very different clinical setting to the present one, it is difficult to compare the results of this study with our own. It made use of data from a community treatment outcomes database, that were not collected prospectively. Alternative treatment options were not randomly allocated and, as a result, numerous treatment selection biases could have been operative. In particular

the pressure on clinicians in the geographic regions covered to commence secondary therapies shortly after PSA values began to rise following attempted curative treatment with surgery or radiation with or without a short course of androgen deprivation, could not be adjusted for easily. Nevertheless, PSA doubling time at a cutpoint of <3 months emerged from a sophisticated analysis as a promising surrogate endpoint candidate for PCSM, and the place of PSA doubling time in future surrogacy analyses was established.

PSA doubling time was focused on as a surrogate endpoint again using data from the RTOG 92.02 randomized trial (15;19). This trial compared 4 months short term androgen deprivation (STAD) starting 2 months before radiotherapy with the same treatment followed by 24 months goserelin (LTAD) in the curative management of locally advanced PC. Results of the surrogacy studies reported in 2003 and 2006 are of direct relevance to the present report because this trial was open to a similar group of men as the TROG 96.01 trial and because different durations of androgen deprivation therapy were compared. It is relevant to note, therefore, that a PSA doubling time at a cutpoint <12 months was found to satisfy all but Prentice's fourth criterion, which it narrowly missed (15). This was thought to be due to the observation that secondary therapy was more successful in men failing STAD who, as a result, survived longer than men who failed LTAD (19). Now that 10 year follow-up data recently became available (29) which confirm that the benefits of secondary therapy were largely confined to men with tumours of Gleason score 7 and below, it would be interesting to know whether the performance of the surrogacy candidates tested has changed with increasing follow-up.

To our knowledge the value of time to biochemical failure as a surrogate endpoint for PCSM from the time of curative treatment has not been addressed.

However, although not directly comparable with the present study, “time to biochemical progression” has been evaluated as a surrogate candidate for PCSM in a series of randomized trials of various new hormonal and chemotherapeutic agents in the management of men with hormone sensitive recurrent and metastatic disease. These studies, which were well reviewed by Collette et al in 2006 (18), indicated that time to PSA progression offered the most promise as a surrogate candidate but could not be validated as an endpoint. In advanced hormone-refractory prostate cancer Petrylak et al found that a decrease in PSA of 30% or greater over a 3 month period satisfied surrogacy criteria using data from the SWOG 99-16 randomized trial comparing docetaxel and estramustine with mitoxantrone and prednisone (14).

Turning to the question of the application of the results of this study, the first issue is whether the results can be used in a new randomized trial in locally advanced prostate cancer. The short answer is no. We agree with Collette (13) and Buyse (30) that a multi-trial meta-analysis of a series of promising surrogate candidates is necessary to validate the effectiveness of one or more of the candidates as surrogate endpoints in a wide range of clinical scenarios. In particular it is important to establish the range of scenarios in which a particular surrogate is useful. The present study has shown that the two variables contain ranges of values that are good surrogate candidates. It has not shown, however, which range of values might be useful in scenarios other than locally advanced disease. In fact it is not yet clear that the range of values identified in this report will be useful in locally advanced cancer trials where markedly different durations of androgen deprivation are being compared. This is because the time to testosterone recovery may have an important influence on time to biochemical failure. In this study this is not an issue because our

findings were near identical regardless of whether TTBF was taken from the end of radiation or from randomization. However the androgen deprivation regimes used in the 96.01 trial were relatively short, which means that the results may not be generalisable to a trial like RTOG 92.02. Fortunately, however, a range of promising new treatment options may allow clinical practice to move away from the use of long term androgen deprivation with its well known complications and almost prohibitive costs. It may not matter therefore whether successful surrogate candidates are identified for this scenario.

Of course surrogate endpoints are of little value if they are difficult to estimate. PSA doubling time has a problem in this regard. In the on-line supplement to this paper we demonstrate PSA doubling time reproducibility in this dataset was strongly dependent on the number of PSA values used to estimate it. Fortunately this reproducibility issue had a limited impact only on the results of the surrogacy testing process. In particular, the sensitivity analyses performed confirm that PSA doubling time is a good surrogate candidate when derived using a relatively wide set of PSA values. Nevertheless it should be remembered that we are not alone in identifying the problems of PSA doubling time estimation (31-33), and international consensus on how this variable is to be derived remains necessary. **It might be of interest to note that in the successor trial to 96.01 (TROG 03.04 Randomized Androgen Deprivation and Radiation [RADAR] Trial), relapse diagnosis guidelines have been designed to instruct clinicians on how to measure PSA doubling time more reliably.** If PSA doubling time is to fulfil its role as a surrogate endpoint, then guidelines on the timing of its estimation after biochemical failure (e.g. within 6 months) will be a priority for development. TTBF is less difficult to estimate accurately, but significant errors are possible particularly if biochemical failure is

not anticipated, as reported in the on-line supplement. An obvious question is whether the same prognostic significance can be attached to TTBF values derived using the markedly different ways of defining biochemical failure after radical prostatectomy. The large study conducted by Pound et al (8), which identified that a TTBF of two years was an important prognostic cutpoint after prostatectomy, suggests that the differences may not be great. The only way to be sure, however, is to undertake an appropriately designed multi-trial meta-analysis which we hope will be stimulated by this report.

This study suggests that if the cutpoints TTBF <2 years or PSA doubling time <12 months proved to be successful surrogates and could be estimated accurately within 6 months of time of biochemical failure, then trials like 96.01, requiring a minimum follow-up period of 7 years to produce mature PCSM results, would be predicted accurately in a trial of similar case composition and sample size with a minimum follow-up of only 2.5 years, i.e. 4.5 years earlier (Figure 6 a,b). In the case of 96.01 this would have been early 2003, only 6½ years after the first patient was randomized. These surrogate candidates therefore have the potential to accelerate clinical trials progress substantially. Meta-analytic testing of these surrogate candidates is keenly anticipated.

Contributors

JD chaired the trial and is the main author of the paper. AS and CW took part in writing the report. JD, DL, DJ, CA JM, KT, NS, and DC took part in data collection and gave editorial help with the paper. AS was responsible for the collection and assembly of data and the coordination of the trial. JD, AS, CW, PSG, PBG and CD participated in the interpretation and analysis of the data. AS, CW and CD did the statistical analysis and gave editorial assistance. All authors reviewed and approved the report.

Conflict of interest

X received an honorarium from y. The other authors declared no conflict of interest.

Acknowledgements

We are grateful for the significant clinical contributions to this trial of Drs C Jose (Auckland, New Zealand), I Franklin, G Bryant, K Gogna, G Dickie, R Allison, L Kenny (Brisbane, Australia), M Jackson, G Morgan, S Cooper, A Kneebone (Sydney, Australia), S Turner (Westmead, Australia), P Jeal, P Fon, M Collins (Townsville, Australia), D Byram, R Lynch (Geelong, Australia), C Johnson (Wellington, New Zealand), C Hamilton, P O'Brien and M Back (Newcastle, Australia). Ms Kristen Palmer (Newcastle, Australia) is thanked for her contribution to the centralised data management of the trial in Newcastle, Ms Dana Bill (Newcastle, Australia) for her design of the Stata program to assess PSA doubling times, and Ms Rosemary Bradford (Newcastle, Australia) for her skilful preparation of the manuscript.

FIGURE LEGENDS

Figure 1. Consort diagram showing outcomes of initial and secondary therapeutic interventions.

* Biochemical (Phoenix) or clinical fail.

† Includes 21 patients whose clinical fail occurred after secondary therapeutic intervention.

‡ Confirmed prostate cancer deaths (excludes a total of 2 unknown deaths and 4 losses to follow-up reclassified as PC deaths due to history of progressive disease).

Figure 2. Venn diagram indicating median and ranges for PSA doubling time (months) and time to Phoenix biochemical failure (years) according to site of failure and secondary therapeutic intervention (STI).

Figure 3. Outcomes after biochemical failure and secondary therapeutic intervention (STI) by treatment arm:

A Time to prostate cancer death from biochemical failure

B Time to STI from biochemical failure

C Time to prostate cancer death from STI

Figure 4. Time to prostate cancer death from randomization by treatment arm.

Figure 5. Time to prostate cancer death from randomization:

A Stratified by TTBF (<2 years vs $\geq$ 2 years)

B Stratified by PSA doubling time (<12 months vs $\geq$ 12 months)

Figure 6. Time to surrogate endpoint by treatment arm using 2003 data:

A Time to biochemical failure if timing is less than 2 years after radiation
(i.e. TTBF <2 years).

B Time to biochemical failure if PSA doubling time is less than 12 months
(i.e. PSA doubling time <12 months).

REFERENCES

- (1) D'Amico AV, Moul JW, Carroll PR, Sun L, Lubeck D, Chen M-H. Surrogate end point for prostate cancer-specific mortality after radical prostatectomy or radiation therapy. *J Natl Cancer Inst* 2003;95(18):1376-83.
- (2) Partin AW, Pearson JD, Landis PK, et al. Evaluation of serum prostate-specific antigen velocity after radical prostatectomy to distinguish local recurrence from distant metastases. *Urology* 1994;43(5):649-59.
- (3) 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 Pt 1):2221-5.
- (4) Zhou P, Chen M-H, McLeod D, Carroll PR, Moul JW, D'Amico AV. Predictors of prostate cancer-specific mortality after radical prostatectomy or radiation therapy. *J Clin Oncol* 2005;23:6992-8.
- (5) Freedland SJ, Humphreys EB, Mangold LA, et al. Risk of prostate cancer-specific mortality following biochemical recurrence after radical prostatectomy. *J A M A* 2005;294(4):433-9.
- (6) Hanks GE, Hanlon AL, Lee WR, Slivjak A, Schultheiss TE. Pretreatment prostate-specific antigen doubling times: clinical utility of this predictor of prostate cancer behaviour. *Int J Radiat Oncol Biol Phys* 1996;34(3):549-53.
- (7) Ward JF, Blute ML, Slezak J, Bergstralh EJ, Zincke H. The long-term clinical impact of biochemical recurrence of prostate cancer 5 or more years after radical prostatectomy. *J Urol* 2003;170:1872-6.
- (8) Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC. Natural history of progression after PSA elevation following radical prostatectomy. *J A M A* 1999;281(17):1591-7.
- (9) Pollack A, Hanlon AL, Movsas B, Hanks GE, Uzzo R, Horwitz EM. Biochemical failure as a determinant of distant metastasis and death in prostate cancer treated with radiotherapy. *Int J Radiat Oncol Biol Phys* 2003;57(1):19-23.
- (10) Zelefsky MJ, Ben-Porat L, Scher HI, et al. Outcome predictors for increasing PSA state after definitive external-beam radiotherapy for prostate cancer. *J Clin Oncol* 2005;23(4):826-31.
- (11) Alcantara P, Hanlon A, Buyyounouski MK, Horwitz EM, Pollack A. Prostate-specific antigen nadir within 12 months of prostate cancer radiotherapy predicts metastasis and death. *Cancer* 2007;109:41-7.
- (12) Buyyounouski MK, Hanlon AL, Horwitz EM, Pollack A. Interval to biochemical failure high prognostic for distant metastasis and prostate cancer-specific mortality after radiotherapy. *Int J Radiat Oncol Biol Phys* 2008;70(1):59-66.
- (13) Collette L, Burzykowski T, Carroll KJ, Newling D, Morris T, Schroder FH. 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. *J Clin Oncol* 2005;23:6139-48.

- (14) Petrylak DP, Ankerst DP, Jiang CS, et al. Evaluation of prostate-specific antigen declines for surrogacy in patients treated on SWOG 99-16. *J Natl Cancer Inst* 2006;98(8):516-21.
- (15) Valicenti RK, DeSilvio M, Hanks GE, et al. Posttreatment prostatic-specific antigen doubling time as a surrogate endpoint for prostate cancer-specific survival: an analysis of radiation therapy oncology group protocol 92-02. *Int J Radiat Oncol Biol Phys* 2006;66(4):1064-71.
- (16) Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431-40.
- (17) Denham JW, Steigler A, Lamb DS, et al. Short-term androgen deprivation and radiotherapy for locally advanced prostate cancer: results from the Trans-Tasman Radiation Oncology Group 96.01 randomised controlled trial. *Lancet Oncol* 2005;6:841-50.
- (18) Collette L, Burzykowski T, Schroder FH. Prostate-specific antigen (PSA) alone is not an appropriate surrogate marker of long-term therapeutic benefit in prostate cancer trials. *Eur J Cancer* 2006;42:1344-50.
- (19) Sandler HM, Pajak TF, Hanks GE, et al. Can biochemical failure (ASTRO definition) be used as a surrogate endpoint for prostate cancer survival in phase III localized prostate cancer clinical trials? Analysis of RTOG protocol 92-02. 2003; 2003 p. Abs 1529.
- (20) 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):929-43.
- (21) American Society for Therapeutic Radiology and Oncology Consensus Panel. Consensus statement: Guidelines for PSA following radiation therapy. *Int J Radiat Oncol Biol Phys* 1997;37:1035-41.
- (22) Pickles T, Kim-Sing C, Morris WJ, Tyldesley S, Paltiel C. Evaluation of the Houston biochemical relapse definition in men treated with prolonged neoadjuvant and adjuvant androgen ablation and assessment of follow-up lead-time bias. *Int J Radiat Oncol Biol Phys* 2003;57(1):11-8.

- (23) Roach MI, Hanks G, Thames HJ, et al. Defining biochemical failure following radiotherapy with or without hormonal therapy in men with clinically localized prostate cancer: recommendations of the RTOG-ASTRO Phoenix Consensus Conference. *Int J Radiat Oncol Biol Phys* 2006;65(4):965-74.
- (24) Denham JW, Kumar M, Gleeson PS, et al. Recognizing false biochemical failure calls after radiation with or without neo-adjuvant androgen deprivation for prostate cancer. *Int J Radiat Oncol Biol Phys* (in press).
- (25) Lin DY, Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. *Stat Med* 1997;16:1515-27.
- (26) Chen C, Wang H, Snapinn SM. Proportion of treatment effect (PTE) explained by a surrogate marker. *Stat Med* 2003;22:3449-59.
- (27) Kaplan EL, Meier P. Non-parametric estimation from incomplete observation. *J Am Stat Assoc* 1958;53:457-81.
- (28) Cox DR. Regression models and life tables. *J R Stat Soc* 1972;34:187-220.
- (29) Horwitz EM, Bae K, Hanks GE, et al. Ten-Year Follow-Up of Radiation Therapy Oncology Group Protocol 92-02: A Phase III Trial of the Duration of Elective Androgen Deprivation in Locally Advanced Prostate Cancer. *J Clin Oncol* 2008;26(15):2497-504.
- (30) Svatek RS, Shulman M, Choudhary PK, Benaim E. Critical analysis of prostate-specific antigen doubling time calculation methodology. *Cancer* 2006;106:1047-53.
- (31) Vollmer RT, Montana GS. The dynamics of prostate-specific antigen after definitive radiation therapy for prostate cancer. *Clin Cancer Res* 1999;5:4119-25.
- (32) Oh WK, Daskivich T. Practical rules for calculating PSA doubling time. *Urology Times* . 1-9-2007. (Magazine Article).
- (33) 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):5218-24.

Table 1. Breakdown by trial arm of prognostic variables at randomization**(n=802)***

	XRT Only (n=270)	3 month STAD (n=265)	6 month STAD (n=267)
Age (years)			
Median (range)	67 (51-80)	68 (47-80)	68 (41-87)
Stage			
T2b,c	164 (61%)	154 (58%)	162 (61%)
T3,4	106 (39%)	111 (42%)	105 (39%)
Gleason score			
2-6	113 (42%)	117 (44%)	122 (46%)
7	114 (42%)	92 (35%)	99 (37%)
8-10	41 (15%)	53 (20%)	43 (16%)
Missing data	2 (1%)	3 (1%)	3 (1%)
PSA (ng/mL)			
<4	10 (4%)	12 (5%)	9 (3%)
≥4 and <10	54 (20%)	69 (26%)	69 (26%)
≥10 and <20	90 (33%)	86 (33%)	99 (37%)
≥20 and <50	95 (35%)	79 (29%)	66 (25%)
≥50 and <100	18 (7%)	18 (7%)	18 (7%)
≥100	3 (1%)	1 (0.4%)	6 (2%)

* XRT = radiotherapy; STAD = short term androgen deprivation; PSA = prostate-specific antigen.

Table 2. Breakdown by trial arm of time to Phoenix fail and PSA doubling time at Phoenix fail (n=436)*

	XRT Only (n=182)	3 month STAD (n=135)	6 month STAD (n=119)
Time to Phoenix fail (years)			
<1	31 (17%)	29 (22%)	13 (11%)
≥1 and <2	38 (21%)	27 (20%)	30 (25%)
≥2 and <3	41 (23%)	23 (17%)	29 (24%)
≥3 and <4	28 (15%)	19 (14%)	13 (11%)
≥4	44 (24%)	37 (27%)	34 (29%)
PSA doubling time (months)			
<3	16 (9%)	23 (17%)	10 (8%)
≥3 and <6	21 (12%)	27 (20%)	25 (21%)
≥6 and <12	43 (24%)	38 (28%)	29 (24%)
≥12 and <24	44 (24%)	19 (14%)	21 (18%)
≥24	25 (14%)	8 (6%)	11 (9%)
Missing	33 (18%)	20 (15%)	23 (19%)

* XRT = radiotherapy; STAD = short term androgen deprivation; PSA = prostate-specific antigen.

Table 3. Application of the Prentice criteria to the surrogate candidates (time to biochemical failure and PSA doubling time) at various cutpoints with time taken from randomization*

PCSM Surrogate	Criterion 2								Criterion 3				Criterion 4								Surrogate Performance	
	3 month STAD				6 month STAD				Surrogate				6 month STAD				Surrogate				+ve No.	Good surrogate #
	HR† (95% CI)	p‡	+ve §		HR† (95% CI)	p‡	+ve §	CTR	HR† (95% CI)	p‡	+ve e §		HR† (95% CI)	p‡	+ve §	PTE	HR† (95% CI)	p‡				
TTBF<1	0.48	0.03	Y		0.10	<0.001	Y	N	9.63	<0.001	Y		0.80		0.36	Y	0.62	8.97 (0.00-0.00)	<0.001		4	N
TTBF<1.5	0.94	0.79	N		0.30	<0.001	Y	Y	13.20	<0.001	Y		0.81		0.37	Y	0.64	13.10	<0.001		4	+
TTBF<2	0.86	0.46	N		0.46	0.001	Y	Y	13.14	<0.001	Y		0.75		0.23	Y	0.50	12.66	<0.001		4	+++
TTBF<2.5	0.81	0.24	N		0.49	<0.001	Y	Y	15.09	<0.001	Y		0.79		0.30	Y	0.59	14.67	<0.001		4	++
TTBF<3	0.71	0.03	Y		0.51	<0.001	Y	N	26.62	<0.001	Y		0.84		0.46	Y	0.70	26.62	<0.001		4	N
TTBF<4	0.64	0.002	Y		0.48	<0.001	Y	N	36.30	<0.001	Y		0.88		0.57	Y	0.78	36.50	<0.001		4	N
TTBF<5	0.62	<0.001	Y		0.50	<0.001	Y	N	65.48	<0.001	Y		0.84		0.45	Y	0.70	65.43	<0.001		4	N
PSA DT<3	1.39	0.32	N		0.48	0.08	N	N	9.79	<0.001	Y		0.68		0.10	Y	0.34	9.23	<0.001		3	N
PSA DT<6	1.04	0.80	N		0.70	0.05	N	N	14.07	<0.001	Y		0.62		0.04	N	-	13.85	<0.001		2	N
PSA DT<9	0.93	0.63	N		0.71	0.03	Y	Y	20.33	<0.001	Y		0.60		0.03	N	-	20.23	<0.001		2	N
PSA DT<12	0.91	0.47	N		0.62	0.001	Y	Y	23.49	<0.001	Y		0.69		0.11	Y	0.36	22.87	<0.001		4	+++
PSA DT<15	0.77	0.05	N		0.57	<0.001	Y	Y	28.96	<0.001	Y		0.75		0.22	Y	0.50	28.33	<0.001		4	++
PSA DT<18	0.73	0.02	Y		0.55	<0.001	Y	N	40.34	<0.001	Y		0.79		0.31	Y	0.59	39.70	<0.001		4	N
PSA DT<24	0.74	0.02	Y		0.53	<0.001	Y	N	50.67	<0.001	Y		0.81		0.37	Y	0.64	49.49	<0.001		4	N

* XRT = radiation therapy; STAD = short term androgen deprivation; HR = hazard ratio; PCSM = prostate cancer-specific mortality; TTBF = time to biochemical (Phoenix) fail (years); PSA DT = PSA doubling time (months); PTE = proportion of treatment effect; PC = Prentice criteria; CTR = comparable to trial result.

† Hazard ratio (model adjusted for pre-treatment prognostic factors age, PSA, tumor stage and Gleason score)

‡ p value from the Cox proportional hazards model

§ Prentice criterion satisfied (Y/N)

|| Comparable to trial result (Y/N) (i.e. treatment effect on surrogate is same as treatment effect on PCSM)

¶ Number of Prentice criteria (1-4) satisfied by 6 month STAD arm

All 4 Prentice criteria satisfied and surrogate predicts trial result for both STAD arms as follows:-

Comparison of HR's for criterion 2 versus HR's for criterion 1 (i.e. for trial result):

+++ Both HR's within 0.1 of each other

++ One HR within 0.1

+ Both more than 0.1

Table 4. Application of the Prentice criteria to the surrogate candidates (time to biochemical failure and PSA doubling time) at various cutpoints with time taken from end of radiation therapy*

PCSM Surrogate	Criterion 2							Criterion 3				Criterion 4								Surrogate			
	3 month STAD			6 month STAD				Surrogate				6 month STAD				Surrogate				Performance			
	HR†	(95% CI)	p‡	+ve §	HR†	(95% CI)	p‡	+ve §	CTR	HR†	(95% CI)	p‡	+ve §	HR†	(95% CI)	p‡	+ve §	PTE	HR†	(95% CI)	p‡	+ve No.	Good surrogate #
TTBF<1	0.96		0.87	N	0.40		0.007	Y	Y	11.02		<0.001	Y	0.80		0.35	Y	0.53	10.60		<0.001	4	++
TTBF<1.5	0.93		0.74	N	0.53		0.009	Y	Y	12.86		<0.001	Y	0.77		0.26	Y	0.45	12.48		<0.001	4	+++
TTBF<2	0.83		0.29	N	0.56		0.004	Y	Y	13.90		<0.001	Y	0.81		0.37	Y	0.56	13.69		<0.001	4	++
TTBF<2.5	0.76		0.10	N	0.61		0.004	Y	Y	22.75		<0.001	Y	0.81		0.38	Y	0.56	22.90		<0.001	4	++
TTBF<3	0.70		0.02	Y	0.59		0.001	Y	N	32.87		<0.001	Y	0.86		0.51	Y	0.68	33.85		<0.001	4	N
TTBF<4	0.67		0.003	Y	0.52		<0.001	Y	N	66.15		<0.001	Y	0.92		0.72	Y	0.83	67.32		<0.001	4	N
TTBF<5	0.65		0.001	Y	0.55		<0.001	Y	N	128.38		<0.001	Y	0.89		0.61	Y	0.76	130.90		<0.001	4	N
PSA DT<3	1.43		0.28	N	0.52		0.12	N	N	9.61		<0.001	Y	0.75		0.24	Y	0.40	9.18		<0.001	3	N
PSA DT<6	1.08		0.66	N	0.78		0.16	N	N	14.17		<0.001	Y	0.68		0.11	Y	0.19	14.01		<0.001	3	N
PSA DT<9	0.96		0.79	N	0.78		0.10	N	N	20.81		<0.001	Y	0.65		0.07	Y	0.10	20.79		<0.001	3	N
PSA DT<12	0.94		0.63	N	0.68		0.008	Y	Y	24.16		<0.001	Y	0.75		0.22	Y	0.40	23.76		<0.001	4	+++
PSA DT<15	0.80		0.09	N	0.61		<0.001	Y	Y	30.19		<0.001	Y	0.81		0.38	Y	0.56	29.94		<0.001	4	++
PSA DT<18	0.76		0.03	Y	0.59		<0.001	Y	N	44.04		<0.001	Y	0.84		0.47	Y	0.64	44.02		<0.001	4	N
PSA DT<24	0.77		0.03	Y	0.58		<0.001	Y	N	59.29		<0.001	Y	0.86		0.53	Y	0.68	58.89		<0.001	4	N

* XRT = radiation therapy; STAD = short term androgen deprivation; HR = hazard ratio; PCSM = prostate cancer-specific mortality; TTBF = time to biochemical (Phoenix) fail (years); PSA DT = PSA doubling time (months); PTE = proportion of treatment effect; PC = Prentice criteria; CTR = comparable to trial result.

† Hazard ratio (model adjusted for pre-treatment prognostic factors age, PSA, tumor stage and Gleason score)

‡ p value from the Cox proportional hazards model

§ Prentice criterion satisfied (Y/N)

|| Comparable to trial result (Y/N) (i.e. treatment effect on surrogate is same as treatment effect on PCSM)

¶ Number of Prentice criteria (1-4) satisfied by 6 month STAD arm

All 4 Prentice criteria satisfied and surrogate predicts trial result for both STAD arms as follows:-

Comparison of HR's for criterion 2 versus HR's for criterion 1 (i.e. for trial result):

+++ Both within 0.1 of each other

++ One within 0.1

+ Both more than 0.1

Figure 1

Figure 2

Figure 3A

Figure 3B

Figure 3C

Figure 4

Figure 5A

Figure 5B

Figure 6A

Figure 6B

On-Line Supplement

Time to biochemical failure (TTBF) using a variation of the Phoenix method.

1. Errors in estimation

Assigned TTBF values are usually overestimates because PSA values at declaration of biochemical failure are almost inevitably greater than the nadir + 2 ng/mL. Estimation of these errors has been derived from interpolations between the dates and values of PSA at time of biochemical failure and the immediately preceding PSA measure.

Time to declaration of biochemical failure exceeded estimated (actual) time to biochemical failure by a median of 2.7 months (range = 0-14.2 months), which translated into an overestimate in median measured time to biochemical failure of 9.5% (range = 0-69.2%). Importantly, substitution of “actual” time to biochemical failure for measured time made minimal difference to the prognostic importance of this variable in **any** of our models.

Although Phoenix biochemical failure was not an initial trial endpoint, this omission was not an important cause for missing Phoenix biochemical failures. ASTRO failures were identified in the absence of Phoenix fail in 106 men. Secondary therapeutic intervention (STI) was commenced in only one of these men prior to the possibility of subsequent Phoenix biochemical failure. Initiation of secondary therapeutic intervention prevented Phoenix biochemical failure being identified in a further 8 men. None of these men had sufficient PSA readings to identify an ASTRO biochemical failure either.

2. Relationship with pre-treatment variables

Multiple logistic regression models (Table S1) confirmed risk factors for cutpoint TTBF < 2 years were higher initial PSA, higher stage and higher Gleason score. When treatment arm was added to this model, initial PSA, stage and Gleason score were retained as significant prognostic factors. In addition, the 6 month STAD arm was shown to be at decreased risk of biochemical failure within the first 2 post treatment years. No significant difference was observed between the 3 month STAD and XRT alone arms.

Table S1. Risk factors for cutpoint TTBF <2 years

Covariate	Wald	Odds ratio	95% CIs	p Value
Age	2.13	0.98	0.95 1.01	0.15
T stage		(T2b=1)		
T2c	1.54	1.41	0.82 2.42	0.22
T3,4	12.27	2.44	1.48 4.02	<0.0001
Gleason grade group		(GS 2-6=1)		
7 only	14.34	2.33	1.5 3.6	<0.0001
8 to 10	29.52	3.94	2.4 6.46	<0.0001
iPSA	19.48	1.02	1.01 1.03	<0.0001

Table S1. Risk factors for cutpoint TTBF<2 years including treatment arm

Covariate	Wald	Odds ratio	95% CIs	p Value
Age	2.23	0.98	0.95 1.01	0.14
T stage		(T2b=1)		
T2c	1.51	1.41	0.82 2.42	0.22
T3,4	12.02	2.43	1.47 4.01	0.001
Gleason grade group		(GS 2-6=1)		
7 only	13.27	2.26	1.46 3.51	<0.0001
8 to 10	28.53	3.9	2.37 6.43	<0.0001
iPSA	20.51	1.02	1.01 1.03	<0.0001
Treatment arm		(XRTonly=1)		
3 mths STAD	1.002	0.8	0.52 1.24	0.32
6 mths STAD	7.56	0.52	0.33 0.83	0.006

PSA doubling time

1. Errors in estimation

Variations in estimating PSA doubling time were derived in five subgroups of at least 20 men with exponential PSA ascents, defined by their assigned PSA doubling time (i.e. 0-6, 6-12, 12-18, 18-24 and 24-48 months). To determine how many PSA values were necessary to reproduce a PSA doubling time estimate near to the assigned PSA doubling time using all PSA values from time of biochemical failure, estimates obtained from a successively increasing number of PSA measures from biochemical failure were compared with the assigned PSA doubling time in each of the 182 men selected for study.

It was found that the number of PSA measures needed to derive a PSA doubling time estimate within 10% of the value derived using all PSA values available (the “assigned” value) were 6 in men with assigned PSA doubling times in the range 0-6 months, 9 in the 6-12 month and 12-18 month subgroups, 10 in the 18-24 month subgroup and 14 in the 24-48 month subgroup. However, if variations within 30% are considered tolerable, then four values were found to be sufficient to ensure that the proportion of men in each subgroup with such variations was 85%.

PSA doubling time estimates were derived using only 2 PSA measures in 82 (23%) men, and 3 measures in a further 67 (19%). In 211 (58%) men, 4 or more PSAs were available for PSA doubling time estimation.

In the 360 men with PSA doubling time estimations, 20 (6%) men were observed to have two distinct doubling times (PSADT1 and PSADT2): in 14 men PSA doubling time became more rapid and in 6 slower. To determine

whether doubling time should be estimated using PSADT1 or PSADT2, a PHM of cause specific survival in all failing men using PSADT1 as a covariate were compared with an identical model using a composite variant of PSA doubling time composed by replacing PSADT1 with PSADT2 for those 20 men with changing doubling times. This comparison indicated that the association between PSA doubling time and PCSM was not altered by using the composite variant of PSA doubling time (models not shown) and PSADT1 (initial PSA DT) is therefore used in all subsequent analyses.

2. Relationship with pre-treatment variables

Logistic regression models (Table S2) confirmed all four trial stratification factors were predictive of PSA doubling time < 6 months. When treatment arm was factored into this model, all four stratification factors remained strongly predictive of PSA doubling time < 6 months however treatment arm itself was not significant.

Table S2. Predictors of PSA doubling time < 6 months

Covariate	Wald	Odds ratio	95% CIs	p Value
Age	5.01	0.97	0.94 0.99	0.025
T stage		(T2b=1)		
T2c	8.03	2.13	1.26 3.59	0.005
T3,4	21.53	3.23	1.97 5.31	<0.0001
Gleason grade group		(GS 2-6=1)		
7 only	21.12	2.57	1.72 3.85	<0.0001
8 to 10	22.43	3.18	1.97 5.13	<0.0001
iPSA	13.78	1.01	1.01 1.02	<0.0001

Table S2. Predictors of PSA doubling time < 6 months including treatment arm

Covariate	Wald	Odds ratio	95% CIs	p Value
Age	5.26	0.97	0.94 0.99	0.022
T stage		(T2b=1)		
T2c	8.1	2.14	1.27 3.61	0.004
T3,4	21.36	3.22	1.96 5.29	<0.0001
Gleason grade group		(GS 2-6=1)		
7 only	20.86	2.57	1.71 3.85	<0.0001
8 to 10	21.8	3.14	1.94 5.07	<0.0001
iPSA	14.36	1.01	1.007 1.02	<0.0001
Treatment arm		(XRTonly=1)		
3 mths STAD	0.32	1.13	0.74 1.71	0.57
6 mths STAD	0.8	0.82	0.53 1.27	0.37

3. Impact of varying PSA range used to estimate PSA DT

Because PSA doubling time estimates are sensitive to the number of PSA values used to derive them, we sought to determine whether the results of the surrogacy testing process summarized in Tables 3 and 4 would be sensitive to the range of PSA values used to determine PSA doubling time.

Briefly summing up the results of the sensitivity analyses (with reference to Table S3):-

- (a) Removal of the 76 men with missing PSA doubling time data has not influenced the Prentice criteria testing. Cutpoints <12 and 15 months still satisfy all four (Table S3a).
- (b) Reconstitution of the dataset with PSA doubling times derived using the PSA value prior to biochemical failure in addition to post failure values to eliminate all missing PSA doubling time data as described below produced the following:-

- i. Using an indefinite number of PSAs (as in the original analysis presented in Tables 3 and 4): All four criteria satisfied and the trial result predicted by cutpoints <12 and <15 months but **in addition** <9 months (Table S3b).
- ii. Using a maximum of 6 months PSA data: All four criteria satisfied by cutpoints <9 months and above, but the trial result predicted too by the <9 month cutpoint alone (Table S3c).
- iii. Using a maximum of 12 months PSA values: All four criteria satisfied and the trial result predicted by cutpoints <12 and <15 months **plus** <9 months (Table S3d).
- iv. Using a maximum of 24 months PSA values: All four criteria satisfied and the trial result predicted by cutpoints <12 and <15 months **plus** <9 months (Table S3e).

Comment: These sensitivity analyses confirm that PSA doubling time at cutpoints <12 and <15 months are good candidates for meta-analysis. In addition <9 months may also be a candidate worth testing. Additional PSAs in the first 6 months of biochemical failure may provide PSA doubling time estimates that are reliable enough for surrogate endpoint use.

Table S3. Sensitivity analysis

PCSM Surrogate	Criterion 2								Criterion 3				Criterion 4						Surrogate Performance			
	3 month STAD				6 month STAD					Surrogate				6 month STAD								
	HR†	(95% CI)	p‡	PC§	HR†	(95% CI)	p‡	PC§	CTR	HR†	(95% CI)	p‡	PC§	HR†	(95% CI)	p‡	PC§	HR†	(95% CI)	p‡	PC No.	Good surrogate#
a.																						
PSA DT<3	1.39		0.32	N	0.47		0.08	N	N	12.15		<0.001	Y	0.75		0.28	Y	11.57		<0.001	3	N
PSA DT<6	1.32		0.21	N	0.76		0.26	N	N	12.64		<0.001	Y	0.58		0.04	N	12.50		<0.001	2	N
PSA DT<9	1.04		0.81	N	0.75		0.13	N	N	16.94		<0.001	Y	0.54		0.02	N	17.09		<0.001	2	N
PSA DT<12	0.99		0.93	N	0.63		0.006	Y	Y	18.52		<0.001	Y	0.65		0.10	Y	18.09		<0.001	4	+++
PSA DT<15	0.80		0.14	N	0.56		<0.001	Y	Y	22.20		<0.001	Y	0.71		0.18	Y	21.70		<0.001	4	++
PSA DT<18	0.75		0.04	Y	0.53		<0.001	Y	N	30.76		<0.001	Y	0.75		0.27	Y	30.24		<0.001	4	N
PSA DT<24	0.76		0.04	Y	0.52		<0.001	Y	N	37.91		<0.001	Y	0.78		0.33	Y	36.95		<0.001	4	N
b.																						
PSA DT<3	1.63		0.13	N	1.10		0.77	N	N	10.7		<0.001	Y	0.48		0.002	N	11.68		<0.001	2	N
PSA DT<6	1.20		0.33	N	0.80		0.28	N	N	13.17		<0.001	Y	0.54		0.009	N	13.06		<0.001	2	N
PSA DT<9	0.87		0.36	N	0.62		0.003	Y	Y	32.64		<0.001	Y	0.67		0.08	Y	31.82		<0.001	4	+++
PSA DT<12	0.84		0.22	N	0.61		0.001	Y	Y	47.88		<0.001	Y	0.68		0.09	Y	46.48		<0.001	4	++
PSA DT<15	0.78		0.06	N	0.56		<0.001	Y	Y	51.11		<0.001	Y	0.74		0.20	Y	49.85		<0.001	4	++
PSA DT<18	0.75		0.03	Y	0.54		<0.001	Y	N	56.46		<0.001	Y	0.78		0.29	Y	55.05		<0.001	4	N
PSA DT<24	0.75		0.02	Y	0.53		<0.001	Y	N	62.99		<0.001	Y	0.82		0.40	Y	61.43		<0.001	4	N
c.																						
PSA DT<3	1.80		0.07	N	1.27		0.47	N	N	10.23		<0.001	Y	0.46		0.001	N	11.32		<0.001	2	N
PSA DT<6	1.05		0.76	N	0.80		0.21	N	N	17.56		<0.001	Y	0.51		0.004	N	17.71		<0.001	2	N
PSA DT<9	0.81		0.16	N	0.64		0.002	Y	Y	29.68		<0.001	Y	0.65		0.06	Y	29.04		<0.001	4	++
PSA DT<12	0.71		0.01	Y	0.57		<0.001	Y	N	45.92		<0.001	Y	0.75		0.20	Y	45.09		<0.001	4	N
PSA DT<15	0.69		0.003	Y	0.52		<0.001	Y	N	96.43		<0.001	Y	0.84		0.44	Y	95.63		<0.001	4	N
PSA DT<18	0.65		<0.001	Y	0.49		<0.001	Y	N	88.19		<0.001	Y	0.86		0.51	Y	87.76		<0.001	4	N
PSA DT<24	0.66		0.001	Y	0.49		<0.001	Y	N	85.32		<0.001	Y	0.87		0.53	Y	84.87		<0.001	4	N
d.																						
PSA DT<3	1.71		0.09	N	1.11		0.76	N	N	10.77		<0.001	Y	0.47		0.002	N	11.72		<0.001	2	N
PSA DT<6	1.09		0.64	N	0.80		0.24	N	N	15.27		<0.001	Y	0.53		0.007	N	15.35		<0.001	2	N
PSA DT<9	0.77		0.08	N	0.62		0.001	Y	Y	25.79		<0.001	Y	0.65		0.06	Y	25.46		<0.001	4	++
PSA DT<12	0.78		0.07	N	0.57		<0.001	Y	Y	42.95		<0.001	Y	0.76		0.23	Y	41.97		<0.001	4	++

PCSM Surrogate	Criterion 2								Criterion 3				Criterion 4						Surrogate Performance			
	3 month STAD				6 month STAD					Surrogate				6 month STAD								
	HR†	(95% CI)	p‡	PC§	HR†	(95% CI)	p‡	PC§	CTR	HR†	(95% CI)	p‡	PC§	HR†	(95% CI)	p‡	PC§	HR†	(95% CI)	p‡	PC No.	Good surrogate#
PSA DT<15	0.78		0.05	N	0.54		<0.001	Y	Y	70.83		<0.001	Y	0.82		0.41	Y	69.16		<0.001	4	++
PSA DT<18	0.73		0.01	Y	0.53		<0.001	Y	N	91.68		<0.001	Y	0.81		0.37	Y	89.29		<0.001	4	N
PSA DT<24	0.69		0.002	Y	0.51		<0.001	Y	N	85.19		<0.001	Y	0.85		0.48	Y	83.70		<0.001	4	N
e.																						
PSA DT<3	1.63		0.13	N	1.10		0.77	N	N	10.70		<0.001	Y	0.48		0.002	N	11.68		<0.001	2	N
PSA DT<6	1.20		0.33	N	0.77		0.20	N	N	13.63		<0.001	Y	0.55		0.01	N	13.43		<0.001	2	N
PSA DT<9	0.84		0.24	N	0.62		0.002	Y	Y	30.37		<0.001	Y	0.65		0.06	Y	29.74		<0.001	4	++
PSA DT<12	0.82		0.16	N	0.61		0.001	Y	Y	46.86		<0.001	Y	0.68		0.09	Y	45.77		<0.001	4	++
PSA DT<15	0.79		0.07	N	0.55		<0.001	Y	Y	73.34		<0.001	Y	0.79		0.30	Y	71.32		<0.001	4	++
PSA DT<18	0.74		0.02	Y	0.53		<0.001	Y	N	65.21		<0.001	Y	0.81		0.37	Y	63.61		<0.001	4	N
A DT<24	0.72		0.006	Y	0.52		<0.001	Y	N	86.86		<0.001	Y	0.84		0.44	Y	85.15		<0.001	4	N

For table key refer to Table 3 legend

- Models using all patients excluding those with Phoenix fail but insufficient data to calculate PSADT n=719
- Models using PSADT calculated with PSA from last pre-Phoenix fail PSA to STI or last available n=794
- Models using PSADT calculated with PSA from last pre-Phoenix fail PSA up to 6 months after Phoenix fail n=794
- Models using PSADT calculated with PSA from last pre-Phoenix fail PSA up to 12 months after Phoenix fail n=794
- Models using PSADT calculated with PSA from last pre-Phoenix fail PSA up to 24 months after Phoenix fail n=794