Metastasis Free Survival is a Strong Surrogate of Overall Survival in Localized Prostate Cancer.

Intermediate Clinical Endpoints of Cancer of the Prostate (ICECaP) Working Group.

Corresponding Author:

Christopher Sweeney, MBBS

Dana Farber Cancer Institute

450 Brookline Ave, D1230

Harvard Medical School

Boston, MA, 02468

Abstract

Background: Adjuvant therapy for intermediate and high-risk localized disease decreases deaths from prostate cancer. Surrogates for overall survival (OS) could expedite the evaluation of new adjuvant therapies.

Methods: By June 2013, 102 completed or ongoing randomized trials were identified and individual patient data was collected from 28 trials with 28,905 patients. Disease-free survival (DFS) and metastasis-free survival (MFS) were determined for 21,140 (from 24 trials) and 12,712 (from 19 trials) patients respectively. We evaluated the surrogacy of DFS and MFS for OS using a 2-stage meta-analytic validation model by determining the correlation of an intermediate clinical endpoint (ICE) with OS and correlation of treatment effects on both the ICE and OS.

Results: The trials enrolled patients from 1987 to 2011. After a median follow-up of 10 years, 45% of 21,140 men and 45% of 12,712 men achieved a DFS and MFS event, respectively. For the DFS and MFS, 61% and 90% of the patients respectively were from radiation trials and 63% and 66% had high-risk disease. At the patient level, the Kendall's tau correlation with OS was 0.85 and 0.91 respectively for DFS and MFS. At the trial level, the R-squared (R^2) was 0.86 (95% CI: 0.78-0.90) and 0.83 (95%CI: 0.71-0.88) from weighted linear regression of 8-year OS rates versus 5 year DFS and MFS rates respectively. The treatment effects (measured by log hazard ratios) for the surrogates and OS were well correlated (R^2: 0.73(0.53-0.82) for DFS and 0.92(0.81-0.95) for MFS).

Conclusions: MFS is a strong surrogate for OS for localized prostate cancer associated with a significant risk of death from prostate cancer.

INTRODUCTION

Each year there are about 1.1 million newly diagnosed cases of prostate cancer with more than 300,000 deaths worldwide[1]. Treatment of intermediate and high risk localized disease with adjuvant systemic therapy is associated with fewer prostate cancer deaths[1-4]. Advances in understanding prostate cancer biology and drug development have resulted in new therapies prolonging the lives of some men with metastatic castration resistant prostate cancer (mCRPC)[5]. Use of these therapies in the adjuvant setting, when micrometastases if present are more sensitive to therapies, may actually eradicate the disease and further decrease the number of men who die of prostate cancer. However, adjuvant prostate cancer clinical trials take longer than a decade to reach the irrefutable endpoint of overall survival (OS). While DFS has proven to be a surrogate for OS and used as a primary endpoint in adjuvant colon cancer trials[6], no intermediate clinical endpoint (ICE) are accepted as a robust surrogate for OS in prostate cancer trials.

An ICE can serve as a good surrogate for OS when there is no curative salvage therapy for relapsed disease and/or substantial risk of dying of the cancer [7-10]. Prior preliminary attempts using single studies to identify ICEs as surrogate for OS in localized prostate cancer have included: time to biochemical failure; PSA-doubling time (PSA-DT); PSA nadir; end of treatment PSA; disease free survival (DFS); and metastasis free survival (MFS)[11]. We hypothesized that DFS and/or MFS may be surrogates for OS as they track more closely with prostate cancer death than a PSA-based ICE[11,12]. A major proportion of patients with intermediate- or high-risk disease localized prostate cancer are cured, and even if they relapse, often die of causes other than prostate cancer. We therefore also investigated the surrogacy of time to disease recurrence (TDR) and time to metastasis (TTM) for disease specific survival (DSS), where non-prostate cancer deaths were not counted as an event.

METHODS:

Search strategy and selection criteria

To enable a meta-analysis of individual patient data (IPD) from localized prostate cancer randomized controlled trials, we conducted a systematic review of studies following the PRISMA guidelines[11]. Eligible trials included randomized, controlled trials for localized disease which were closed to accrual and conducted in Australia, New Zealand, Canada, Europe or USA. Trials with primary endpoints other than efficacy (e.g. safety, toxicity, QOL, feasibility, dosimetry, and patient decision-making) without systematic long-term follow-up were excluded.

At time of project initiation, June 2013, 102 trials were identified as potentially eligible of which 43/102 (42%) had both data suitable for use and a study group that agreed to participate. This resulted in possible IPD from 28,905 patients. For this analysis, IPD was able to be provided for 28/43 (65%) trials with 22,825 patients. Not all trials collected all of the endpoints of interest. Therefore, for the DFS and MFS analysis, 21,140 (from 24/43 (56%) trials) and 12,712 patients (from 19/43 (44%) trials) could be included respectively. Trials which did not document data on these endpoints were excluded. The selection process and reasons for exclusion are shown in Figure Supplemental (S)1.

Statistical analyses

Definition of endpoints

DFS was measured from the date of randomization to date of first evidence of recorded clinical recurrence (local/regional recurrence and/or distant metastases confirmed by imaging or histological evidence) or death from any cause; or censored at the date of last follow-up. MFS had the same definition as DFS but did not include local-regional recurrence. TDR and TTM were defined analogously to DFS and MFS but non-prostate cancer deaths without prior progression were censored or counted as competing risk. OS was measured from the date of randomization to death from any cause, censored at the date of last follow-up in patients who were alive. DSS was defined similarly as OS but non-prostate deaths were censored or

considered as competing risk in sensitivity analyses. Local recurrence and cause of death were based on trial-defined events (see Table S8).

Surrogacy criteria

We evaluated the surrogacy of DFS and MFS with OS using a widely accepted[13] meta-analytic 2-stage validation model where two conditions must hold to claim an ICE is a surrogate for OS[14,15] (See supplemental statistical methods for discussion on choice of model). Condition 1 requires that the ICE and OS be correlated. Condition 2 requires that the treatment effects on both end points be also correlated. The validity of the surrogate is reflected by the strength of the correlations. To be consistent with other surrogacy assessments in oncology, we defined *a priori* a clinically relevant surrogacy of a R-squared (R^2) of 0.7 or higher[11].

Condition 1 was tested at both patient and trial levels. At the patient level, the associations of OS with DFS and MFS were evaluated through a bivariate copula model (see supplemental statistical methods for details) fitted on individual patient data[16]. Kendall's tau (range 0-1) quantified the correlation between the endpoints. At the trial level, we first obtained Kaplan-Meier estimates of 5-year DFS or MFS rate and 8-year OS rate for each treatment arm within each trial. We then performed weighted linear regression (WLR) analyses between trial and arm-specific OS rate at 8 years versus DFS and MFS rate at 5 years. These timepoints were chosen as they are frequently reported in the literature. Regressions were weighted by the inverse variances of the 5-year estimates of the ICE. The R^2 was used to quantify the proportion of variance explained by the regressions.

To test condition 2, we performed Cox regression models to obtain the study-specific treatment effects (i.e. the natural log [hazard ratio (HR)]) on the ICE and OS. We then fit a WLR model between treatment effects on OS versus treatment effects on DFS or MFS. Regressions were weighted by the inverse variances of the natural log [HR]) on the ICE and R^2 was used to

quantify the proportion of variance explained by the regressions. This approach was also applied to the surrogacy analysis of TDR and TTM for DSS (non-prostate cancer deaths were censored).

Subgroup and sensitivity analysis

Given the heterogeneous population and treatment in the localized disease setting, we conducted pre-planned subgroup analyses by (1) types of primary therapy (radical prostatectomy [RP] versus radiation therapy [RT]); (2) within RT-trials: duration of ADT (≤6 or >6 months); (3) patient risk groups defined by NCCN, D'Amico or pathological features. Because a large proportion of TDR and TTM endpoints are censored due to non-prostate cancer deaths, we performed a sensitivity analysis to estimate trial-level correlation between cumulative incidence estimates of TDR/TTM and DSS and between sub-distribution treatment effect hazard ratio (sHR) estimates for TDR/TTM and DSS from competing risk models[17] where non-prostate cancer deaths were considered as the competing risk for each endpoint. Model accuracy was assessed by a leave-one-out-cross validation (Supplemental methods).

Surrogate Threshold Effect (STE)

STE is defined as the minimum treatment effect (HR) on the surrogate necessary to predict a non-zero treatment effect (i.e. HR different from 1) on OS in a future trial[18]. To obtain STE, we constructed the 95% prediction limits for the regression line of treatment effect on OS versus treatment effect on the surrogate, accounting for the mean weights of the current trials. The intersection of the upper 95% prediction limit with the horizontal line (representing a HR of 1 for OS) was defined as STE, corresponding to no treatment effect on OS.

All analyses were performed using SAS Software version 9.4 or later (SAS Institute Inc, Cary, NC) and the R packages (www.r-project.org).

RESULTS:

Trial and patient characteristics

21,140 patients from 24 trials and 12,712 patients from 19 trials had documented data on DFS and MFS analysis respectively (Table S1, S2). Five trials were split according to type of primary therapy or experimental arm resulting in 31 and 21 study units for the DFS and MFS analysis respectively (see supplemental Statistical Methods).

The trials enrolled patients from 1987 to 2011 and median follow-up was 10 years (Range: <0.1~22.7 years). More than 80% of the patients were younger than 75 years old (Table S3). For the DFS and MFS analysis, 61% and 90% of the patients respectively were on radiation trials and 63% and 66% had high-risk disease. The observed 5-year rates for DFS was 76%, 79% for MFS and 84% for OS. Figure 1 shows the Kaplan Meier distributions of the endpoints. The estimated hazard function by years since randomization for each endpoint is shown in Figure 2.

Surrogacy condition 1: correlation between ICE and OS.

At the individual patient level, the correlation with OS was 0.85 (95%CI: 0.85-0.86) and 0.91 (95%CI: 0.91-0.91) respectively for DFS and MFS, as measured by the Kendall's tau from a copula model. When non-prostate cancer deaths were censored, the correlation with DSS was 0.68 (95%CI: 0.67-0.69) for TDR and 0.91 (95%CI: 0.91-0.92) for TTM. The tight correlation between the endpoints is reflected by the tight correlation between trial and arm-specific Kaplan-Meier estimates of OS or DSS at 8 years versus Kaplan-Meier estimates of the surrogates at 5 years (Figure 3). From the WLR, the R^2 was 0.86 (95%CI: 0.78-0.90) and 0.83 (95%CI: 0.71-0.88) between 8-year OS rates versus 5-year DFS and MFS rates respectively. When non-prostate cancer deaths were censored, there was still a high correlation of 8-year DSS rates (R^2: 0.80 (95%CI: 0.70-0.85) with 5-year TDR and 0.86 (95%CI: 0.75-0.90) for 5-year TTM) (Table 1).

Surrogacy condition 2: correlation between treatment effect on ICE and OS

At the trial level, trial-specific treatment effects, measured by the HRs for each endpoint, are shown in the forest plots (Figure S2). The R^2 was 0.73 (95%CI: 0.53-0.82) from the WLR of Log(HR)-OS versus log(HR)-DFS and reduced to 0.63 (95%CI: 0.36-0.75) with non-prostate cancer deaths censored. There was a strong correlation between Log(HR)-OS and log(HR)-MFS across trials (R^2: 0.92[95%CI: 0.81-0.95]), and the high correlation remained when non-prostate cancer deaths were censored (R^2: 0.89[95%CI: 0.72-0.93]) (Figure 4). The estimated WLR equation for each endpoint is provided in Table 1.

Subgroup and sensitivity Analysis

Overall, the results were consistent when the analysis was restricted to the high risk population only, or in subgroup analysis by type of primary therapy and by exposure to ADT within RT-based trials at both patient and trial level (Tables S4, S5 & S6). At the patient level, the Kendall's tau correlation between OS and DFS was 0.91(95%CI: 0.90-0.92) and 0.84(95%CI: 0.83-0.84) in RP- and RT-based trials respectively. At the trial level, the R^2 from the WLR of Log(HR)-OS versus log(HR)-DFS was 0.87(95%CI:0.31-0.93) for RP trials and 0.75(95%CI : 0.48-0.84) for RT trials. For the MFS endpoint, no separate analysis was conducted for RP based trials since 90% of the patients were from radiation trials. The correlation between OS or DSS and each ICE was slightly stronger in those who received >6 months of adjuvant ADT compared to those who received no or short-term neoadjuvant/adjuvant ADT (Table S4).

Results were also consistent in WLR analysis of trial level correlations when non-prostate cancer deaths were treated as competing risk (Table S7), and in leave-one-out cross-validation (Fig S3,S4).

STE and implications for trial designs

The STE on OS was a HR(DFS) of 0.67 and a HR(MFS) of 0.88, indicating that a risk reduction of 33% and 12% respectively would predict a non-zero effect on OS (Figure 4). Additionally, the STE on DSS was a HR(TDR) of 0.49 and a HR(TTM) of 0.74, hence, a larger treatment effect on the TDR would be required to predict a treatment benefit on DSS.

Given the strong correlation between MFS and OS, clinical trials can be designed using MFS as primary endpoint, instead of OS (See Supplementary Material: *Study Designs Using MFS and OS Endpoint*). Historically trials have been designed with an OS hazard ratio ranging from 0.71 to 0.75. These trials have a study duration of 11.5 to 16.2 years with 1000 patients enrolled over 5-years (Figure S5). Clearly, the study durations would be shorter if the same treatment effects were assumed for MFS (Figure S5). The WLR analyses (Table 1, Figure 4) predicts that for OS hazard ratios ranging from 0.71 to 0.75, the corresponding MFS hazard ratios would range from 0.65 to 0.7 (Table S9), so the benefit of using MFS instead of OS could be even greater. However, the surrogate threshold effect, which is a MFS hazard ratio of 0.88, implies that a future trial would require an upper limit of the confidence interval for the estimated HR(MFS) to fall below the STE to predict a significant effect on OS. Hence, depending on the assumed hazard ratios and the number of patients, the duration of the trial may favor choosing MFS or OS as the primary endpoint (Figure 5). MFS would be the preferred primary endpoint for HR(OS) lower than 0.7, while OS would be the preferred primary endpoint for HR(OS) greater than 0.72. For example, a trial with 1,000 patients designed to detect a treatment effect of HR(MFS) of 0.6 would have a total study duration of 7.7 years. The associated predicted HR(OS) is 0.67 and a trial designed to detect this effect would have a total study duration of 8.8 years.

DISCUSSION:

In a cohort of prostate cancer patients with an approximate15% chance of dying of prostate cancer over a 10-year period, DFS and MFS are valid surrogates for OS. As the estimated hazard across times curves (Figure 2) depict, early prostate cancer recurrences are associated with death from prostate cancer before dying from a competing co-morbidity in a patient population, 80% of whom are younger than 75 years old and fit for enrolment on a clinical trial.

The practical output for surrogacy work includes being able to complete trials in a more expeditious manner. The advantage of using a surrogate such as MFS rather than OS is the ability to observe the number of required events earlier, but there is some uncertainty of how well the surrogate predicts the effect on the true endpoint. However, this uncertainty is captured by the STE, which is the minimum treatment effect required on the surrogate to predict a significant treatment effect on the true endpoint. In short, use of MFS can allow an expeditious evaluation of a new therapy if it has a meaningful treatment effect on MFS. Notably, a HR(MFS) of 0.6 has been observed in adjuvant trials of testosterone suppression (TS) plus radiation versus radiation in high risk localized disease and resulted in improvements in overall survival[3,4,19-22]. There are possibly other health economic benefits for preventing the morbidity and adverse effects of treatment associated with a metastatic event[11]. Defining these benefits is part of ongoing work being conducted by the ICECaP working group.

Use of the IPD was critical in conducting this analysis and allowed a side-by-side comparison of DFS and MFS as surrogates for OS. There were more patients and trials suitable for DFS than the MFS analyses as some studies did not record events beyond first clinical progression and are only viable for DFS analysis. As such, systematic follow-up until first distant recurrence is required in future studies to capture MFS events. The lower correlation of 0.7 for DFS versus 0.9 for MFS results in a lower STE for DFS (0.67 vs 0.88) as the prediction intervals for DFS are wider and hence a need for a greater treatment effect. This is presumably due to local

recurrences are possibly indolent and/or cured with salvage therapy. The sensitivity analysis showed the MFS correlation with OS was maintained whether the primary localized therapy was surgery or radiation based and whether adjuvant ADT was used (Table S4); of note, the 5 trials in which MFS was not collected (N=8,428 patients) were all trials of adjuvant hormonal therapy. Notably, early metastatic relapse is associated with death from prostate cancer and the sensitivity analyses have shown this is regardless of receipt of ADT in the adjuvant setting and presumably also for biochemical or metastatic disease. Moreover, the subgroup analysis by duration of ADT could only be done at the IPD level as most trials were designed to compare duration of ADT.

The IPD also provides unique insights into the natural history of prostate cancer. Figure 2 details a constant rate of relapses and late relapses have less impact on OS than relapses before 7 years in this cohort with a median OS of 12.7 yrs. Presumably there is an increase of non-prostate cancer deaths in later years and later relapses have a more indolent course.

The cross-validation, subgroup and sensitivity analysis provide further reassurance that the results are robust. There were some limitations of our study. First, the DFS endpoint incorporated local recurrence as an event defined by the trials with variations in the definition of local recurrence. Second, we could not provide a separate analysis for MFS for surgery based patients given limited numbers of surgery based trials.

While it would be preferable to have an earlier endpoint than 5 year DFS or MFS, this time-point was chosen as it was associated with enough events to allow a robust analysis. Correlation with 10-year OS was thwarted as some of the trials did not have enough follow-up resulting in a smaller number of units and fewer patients at risk and presumably greater impact from other causes of death. As such the 8 year OS rate was more reliable. Additionally, since OS data requires long term follow-up, most of the trials included were commenced before 2005. Our

ongoing work with recently completed trials will investigate the reproducibility of our findings in the era with new therapies prolonging the OS of men with metastatic HSPC and CRPC. However, seeing as these have a modest improvement in OS and do not cure the disease, it is anticipated surrogacy will still persist.

In conclusion, MFS is a strong surrogate for OS in clinically localized prostate cancer in a patient population with approximately 15% chance of dying of prostate cancer over 10 years despite potentially curative local therapy. The surrogacy is independent of primary local interventions and type of adjuvant therapy. The linear regression graphs used to generate the STE can be used to define relative improvements in MFS that are associated with clinically meaningful improvements in OS.

Funding

- Prostate Cancer Foundation (PCF) Challenge Award
- Grants from Industry: Astellas/Medivation, Janssen, Millennium, Sotio, Sanofi

Role of the funding source

This study was designed by CJS, WX and MR in collaboration with the ICECaP Working Group (WG). The DFCI coordinating center had full access to the data and independent WG members oversaw Statistical Analysis Plan development and interpretation of the data. The corresponding author had final responsibility for the decision to submit for publication. The final report was shared with the pharmaceutical companies who provided financial support as investigator initiated grants but had no input on the design or interpretation of the results.

Acknowledgements:

Writing Committee:

- Co-ordinating Center at Dana Farber Cancer Institute (DFCI): Christopher Sweeney – Chair (Lank Center for Genitourinary Oncology, Dana Farber Cancer Institute, Boston, MA, USA), Meredith Regan (Department of Biostatistics and Computational Biology, Dana- Farber Cancer Institute, Boston, MA, US), Wanling Xie (Department of Biostatistics and Computational Biology, Dana- Farber Cancer Institute, Boston, MA, US)
- Members Independent of Coordinating Center overseeing Statistical Analysis Plan: Marc Buyse (International Drug Development Institute, Louvain-la-Neuve, Belgium), Susan Halabi (Biostatistics and Bioinformatics, Duke University, Durham, USA), Philip Kantoff (Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, USA, Prostate Cancer Foundation, Santa Monica, USA), A. Oliver Sartor (Dept. Of Medicine & Urology, Tulane University, New Orleans, USA), Howard Soule (Prostate Cancer Foundation, Santa Monica, USA).

- External Review

  Noel Clarke (Urological oncology, The Christie NHS Foundation Trust, Manchester, Great Britain), Laurence Collette (Headquarters, EORTC, Brussels, Belgium), James Dignam (Dept. of Public Health Science, University of Chicago, Chicago, USA), Karim Fizazi (Department of Cancer Medicine, Institut Gustave Roussy, Villejuif, France), Wendy Parulekar (Canadian Cancer Trials Group, Cancer Research Institute, Queen's University, Kinston, Canada), Howard Sandler (Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, CA, USA), Matthew Sydes (MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology- UCL, London, Great Britain), Bertrand Tombal (Institut de Recherche Clinique, Université Catholique de Louvain, Brussels, Belgium); Scott Williams (Radiation Oncology, Peter MacCallum Cancer Centre, East Melbourne, Australia)

Working Group Members

ICECaP Working Group Members (in alphabetical order)

- Ove Andren, John Armstrong, Donald Berry, Michel Bolla, Marc Buyse, Simon Chowdhury, Noel Clarke, Joseph Chin; Laurence Collette, Matthew Cooperberg, Jim Denham, Savino Mauro Di Stasi, Mario Eisenberger, James Dignam, Karim Fizazi, Boris Freidlin, Martin Gleave, Muriel Habibian, Susan Halabi, Nick James, Jonathan Jarow, Nancy Keating, Philip Kantoff, Gary Kelloff, Laurence Klotz, Himu Lukka, Malcolm Mason, Nicolas Mottet; Andrea Miyahira, Mari Nakabayashi, Wendy Parulekar, Meredith Regan, Howard Sandler, Oliver Sartor, Peter Scardino, Howard Scher, Richard Simon, Jonathan Simons, Eric Small, Howard Soule, Christopher Sweeney, Matthew Sydes, Catherine Tangen, Ian Thompson, Bertrand Tombal, Anders Widmark, Thomas Wiegel, Manfred Wirth; Scott Williams, Eric Yeoh; Wanling Xie, Almudena Zapatero

Other Acknowledgments

- Victoria Wong, Frontier Science & Technology Research Foundation: Computer programming.

REFERENCES:

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87-108.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.
3. Bolla M, Gonzalez D, Warde P, et al. Improved survival in patients with locally advanced prostate cancer treated with radiotherapy and goserelin.[see comment]. New England Journal of Medicine 1997;337:295-300.
4. Denham JW, Steigler A, Lamb DS, et al. Short-term neoadjuvant androgen deprivation and radiotherapy for locally advanced prostate cancer: 10-year data from the TROG 96.01 randomised trial. Lancet Oncol 2011;12:451-9.
5. Attard G, Parker C, Eeles RA, et al. Prostate cancer. Lancet 2016;387:70-82.
6. 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;23:8664-70.
7. Hamdy FC, Donovan JL, Lane JA, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med 2016;375:1415-24.
8. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 2014;370:932-42.
9. Crook JM, O'Callaghan CJ, Duncan G, et al. Intermittent androgen suppression for rising PSA level after radiotherapy. N Engl J Med 2012;367:895-903.
10. Ross AE, Yousefi K, Davicioni E, et al. Utility of Risk Models in Decision Making After Radical Prostatectomy: Lessons from a Natural History Cohort of Intermediate- and High-Risk Men. Eur Urol 2016;69:496-504.
11. ICECaP Working Group. The Development of Intermediate Clinical Endpoints in Cancer of the Prostate (ICECaP). J Natl Cancer Inst 2015;107.
12. Schweizer MT, Zhou XC, Wang H, et al. Metastasis-free survival is associated with overall survival in men with PSA-recurrent prostate cancer treated with deferred androgen deprivation therapy. Ann Oncol 2013;24:2881-6.
13. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. Int J Technol Assess Health Care 2014;30:312-24.
14. Buyse M, Michiels S, Squifflet P, et al. Leukemia-free survival as a surrogate end point for overall survival in the evaluation of maintenance therapy for patients with acute myeloid leukemia in complete remission. Haematologica 2011;96:1106-12.
15. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000;1:49-67.
16. Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate end points in multiple randomized clinical trials with failure time endpoints. Appl Statist 2001;50:405-4222.
17. Fine J, Gray R. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association 1999;94:496-509.
18. Burzykowski T, Buyse M. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. Pharmaceut Statist 2006;5:173–86.
19. Shipley WU, Seiferheld W, Lukka HR, et al. Radiation with or without Antiandrogen Therapy in Recurrent Prostate Cancer. N Engl J Med 2017;376:417-28.
20. Messing EM, Manola J, Sarosdy M, Wilding G, Crawford ED, Trump D. Immediate hormonal therapy compared with observation after radical prostatectomy and pelvic lymphadenectomy in men with node-positive prostate cancer. N Engl J Med 1999;341:1781-8.
21. Messing EM, Manola J, Yao J, et al. Immediate versus deferred androgen deprivation treatment in patients with node-positive prostate cancer after radical prostatectomy and pelvic lymphadenectomy. Lancet Oncology 2006;7:472-9.

22. Bolla M, Van Tienhoven G, Warde P, et al. External irradiation with or without long-term androgen suppression for prostate cancer with high metastatic risk: 10-year results of an EORTC randomised study. Lancet Oncol 2010;11:1066-73.

Table 1 Two-condition surrogacy analysis

True Endpoint (TE)	Intermediate Clinic Endpoint (ICE)				Condition 1 (TE and ICE are correlated)		Condition 2 (Treatment effects on both endpoints are correlated)	
					Correlation at the patient level	Regression of 8-year TE rate versus 5-year ICE rate*** (by trial and arm)	Regression of Log(HR)-TE versus Log(HR)-ICE (by trial)	
		Number of trials	Number of units*	Number of patients	Kendall Tau, (95% CI)	R-squared, 95% CI	R-squared, 95% CI	Regression Equation
OS	DFS	24	31	21,140	0.85 (0.85-0.86)	0.86 (0.78-0.90)	0.73 (0.53-0.82)	$Log(HR)_{OS}$= 0.035 + 0.605 x $Log(HR)_{DFS}$
DSS	TDR	21**	28	20,496**	0.68 (0.67-0.69)	0.80 (0.70-0.85)	0.63 (0.36-0.75)	$Log(HR)_{DSS}$ = 0.027 + 0.809 x $Log(HR)_{TDR}$
OS	MFS	19	21	12,712	0.91 (0.91-0.91)	0.83 (0.71-0.88)	0.92 (0.81-0.95)	$Log(HR)_{OS}$= -0.021 + 0.740 x $Log(HR)_{MFS}$
DSS	TTM	16**	18	12,068**	0.91 (0.91-0.92)	0.86 (0.75-0.90)	0.89 (0.72-0.93)	$Log(HR)_{DSS}$=-0.072 + 0.880 x $Log(HR)_{TTM}$

OS: Overall survival, DSS: Disease specific survival, DFS: Disease free survival, TDR: Time to disease recurrence, MFS: Metastasis free survival, TTM: Time to metastasis
HR: Hazard ratio

*Five trials were split according to type of primary therapy or experimental arm (if ≥2 experimental arms).

**Excluding 3 studies with number of prostate cancer death less than 3

***8-year TE rates and 5-year ICE rates were Kaplan Meier estimates by trial and treatment arm, excluding three studies with median follow-up less than 6 years

Figure Legends

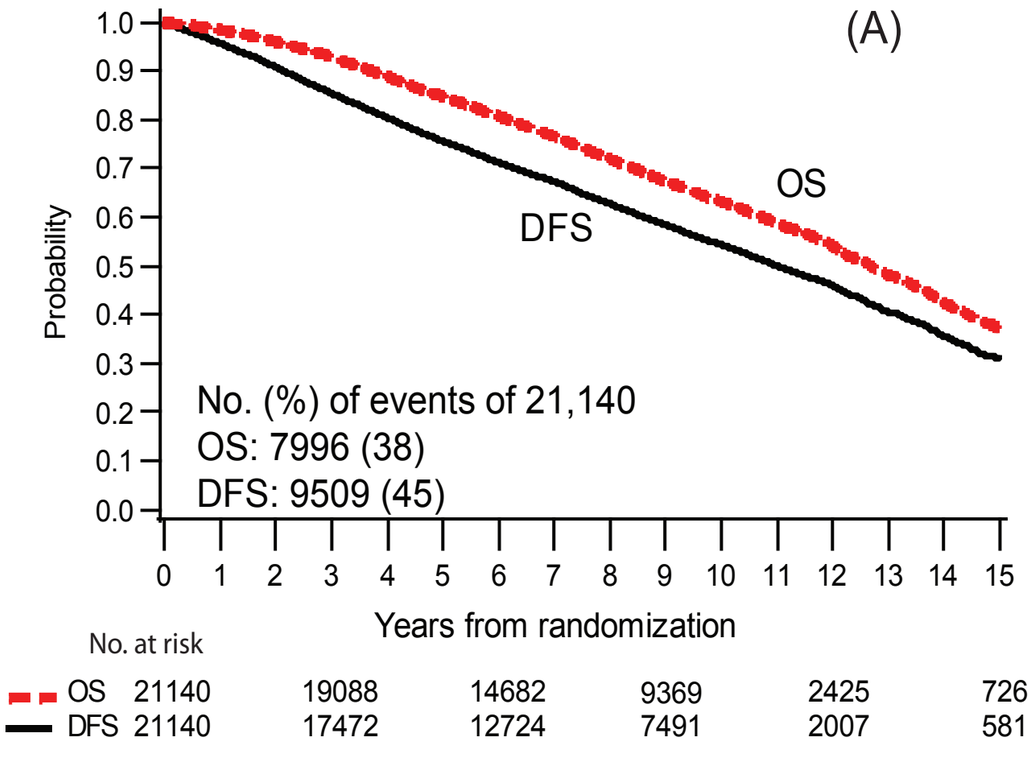
Figure 1 Kaplan Meier estimates of endpoints (A)OS and DFS (B)DSS and TDR (C)OS and MFS (D)DSS and TTM. Median follow-up was 10 years.

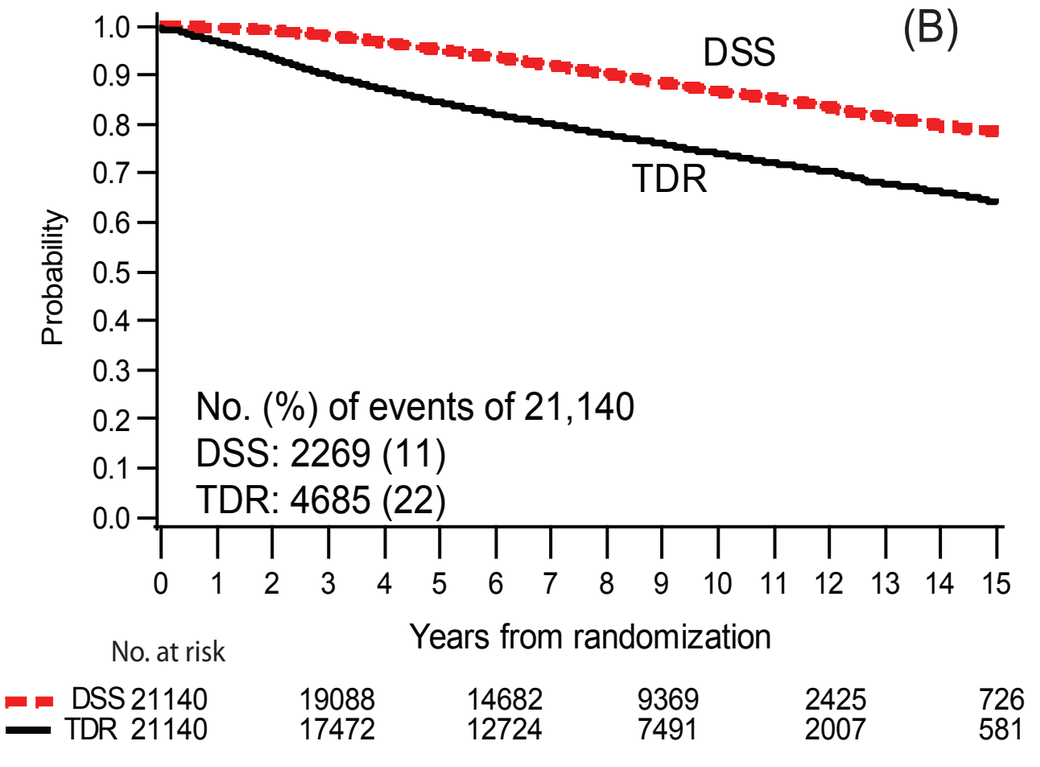
Figure 2 Estimated hazard across times (A)OS and DFS (B)DSS and TDR (C)OS and MFS (D)DSS and TTM

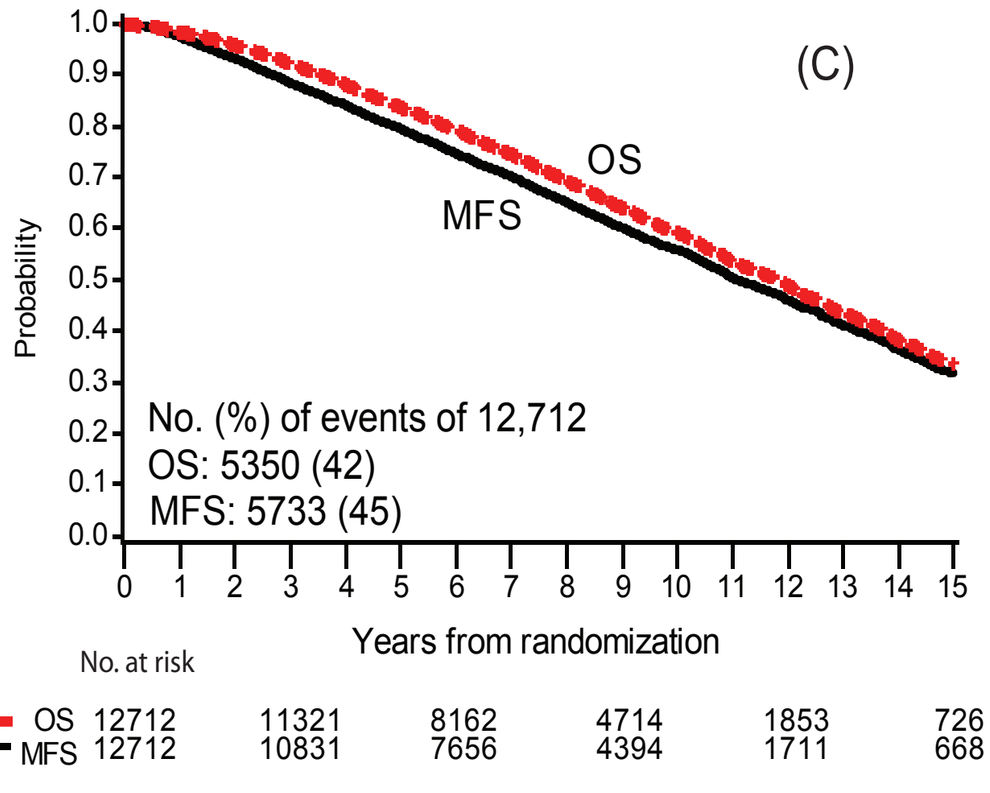
Figure 3 OS or DSS rate at 8 years versus surrogate endpoints at 5 years: (A)8 year OS vs 5 year DFS, (B)8 year DSS vs 5 year TDR, (C) 8 year OS vs 5 year MFS, (D)8 year DSS vs 5 year TTM. All rates were Kaplan Meier estimates by trial and treatment arm. Circle size and regression were weighed by inverse variance of the 5 year rate estimate for the surrogates.

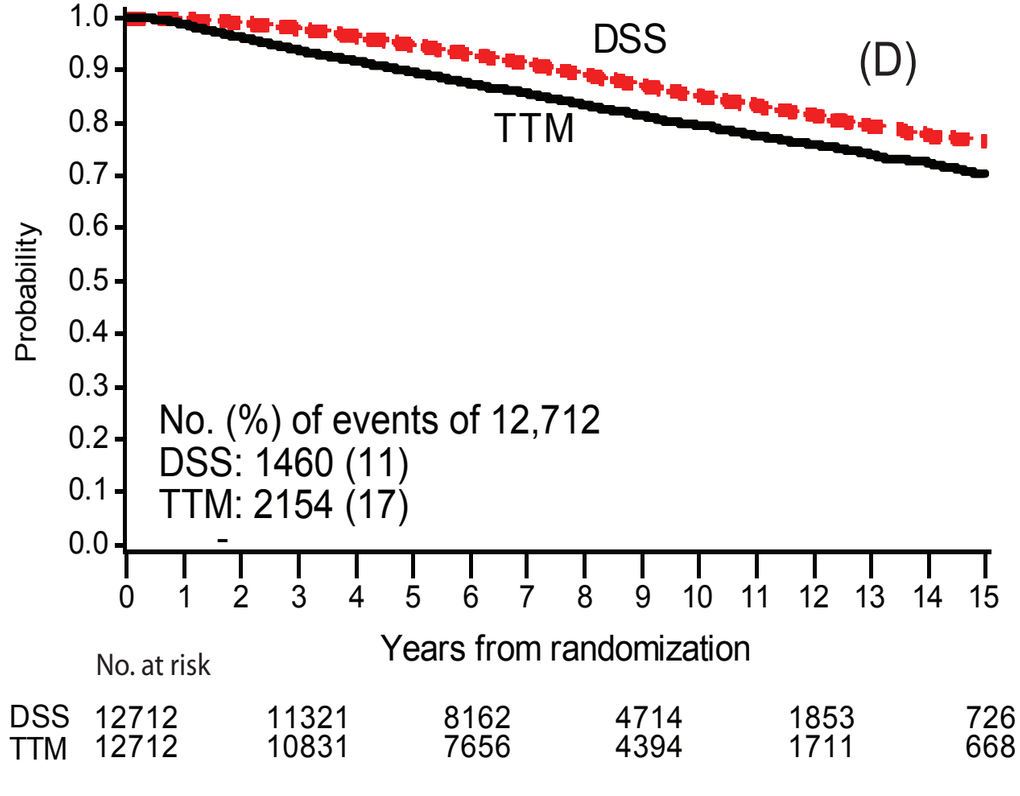
Figure 4 Treatment effects (hazard ratio(HR)) on OS or DSS versus treatment effects on surrogates: (A) OS-HR versus DFS-HR, (B) DSS-HR versus TDR-HR, (C)OS-HR versus MFS-HR, (D)DSS-HR versus TTM-HR. HRs were estimated from Cox regression for each study and values were natural logarithm transformed. Circle size and regression were weighed by inverse variance of Log-HR estimates for the surrogates.

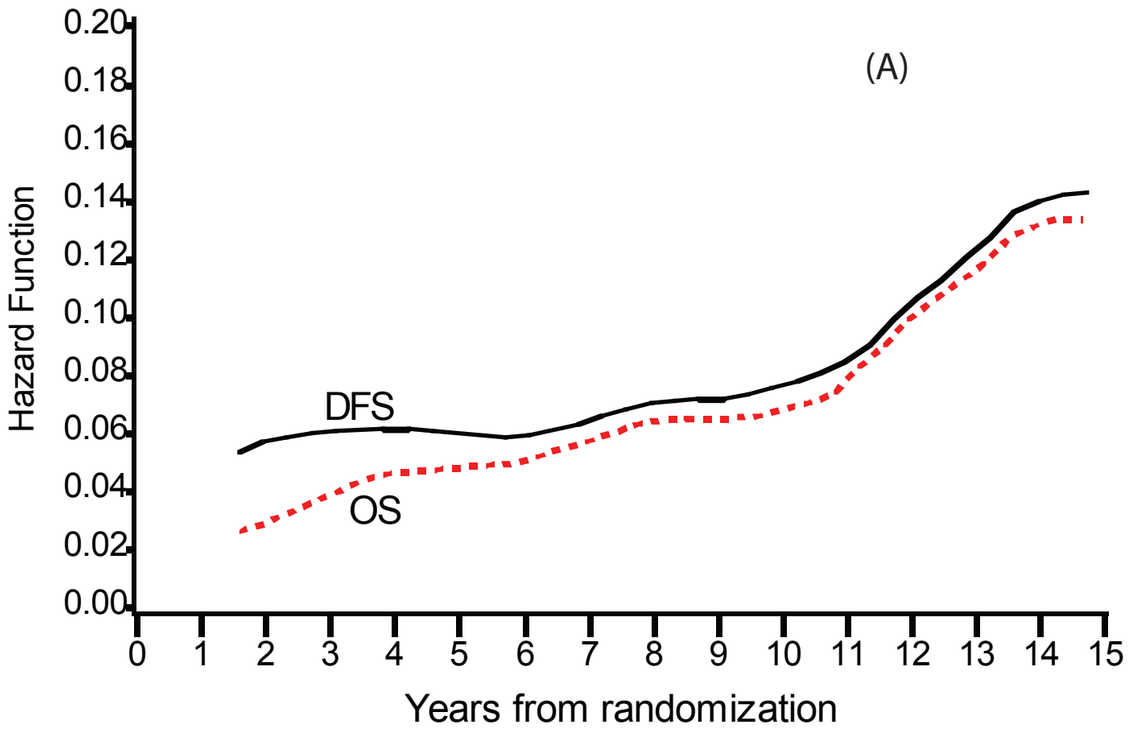
Figure 5: Total study duration required in the study designs using MFS hazard ratios and testing STE (solid line) or using predicted OS hazard ratios from weighted linear regression (dashed line). MFS would be the preferred primary endpoint for HR(OS) lower than 0.70, while OS would be the preferred primary endpoint for HR(OS) greater than 0.72 (gray vertical dashed lines). Design assumptions include (1)5-years MFS and OS rate of 0.79 and 0.84 (hazard=0.04714 and 0.03487 under exponential distribution) respectively, (2)5 years of accrual period, (3)type I error of 0.025 (one-sided) and type II error of 0.20**.**

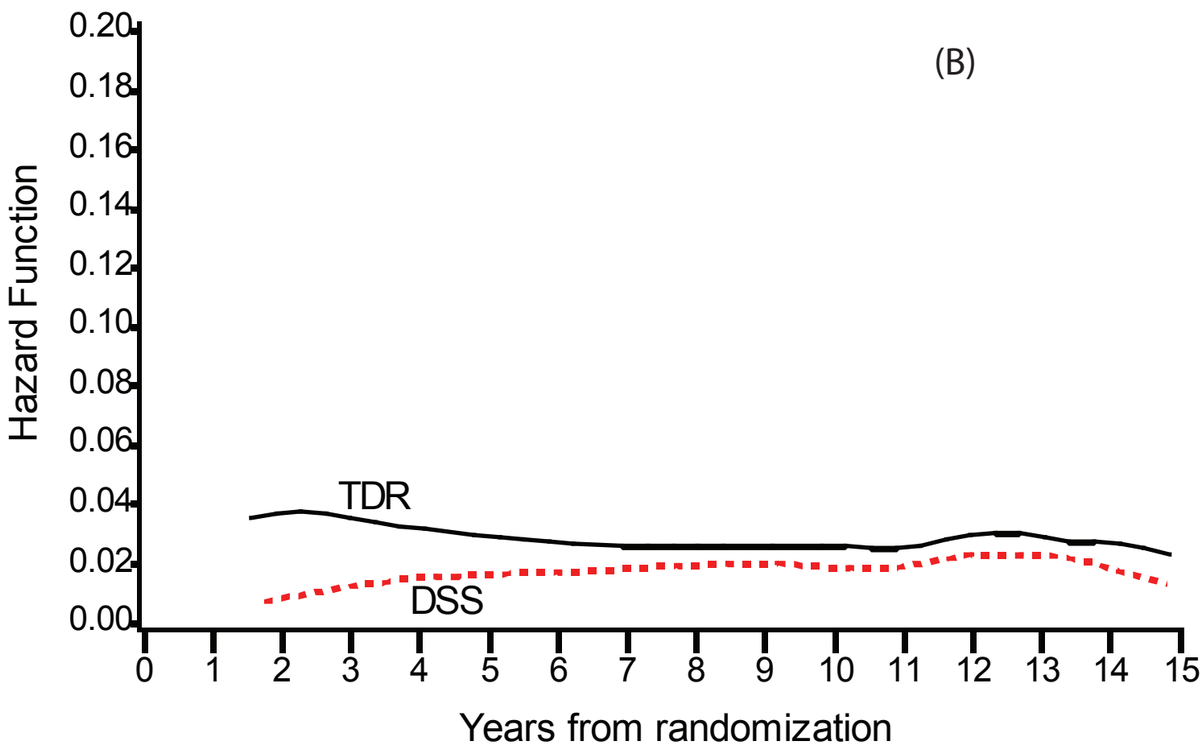

(A)
Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OS
DFS
No. (%) of events of 21,140
OS: 7996 (38)
DFS: 9509 (45)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years from randomization
No. at risk
OS 21140 19088 14682 9369 2425 726
DFS 21140 17472 12724 7491 2007 581

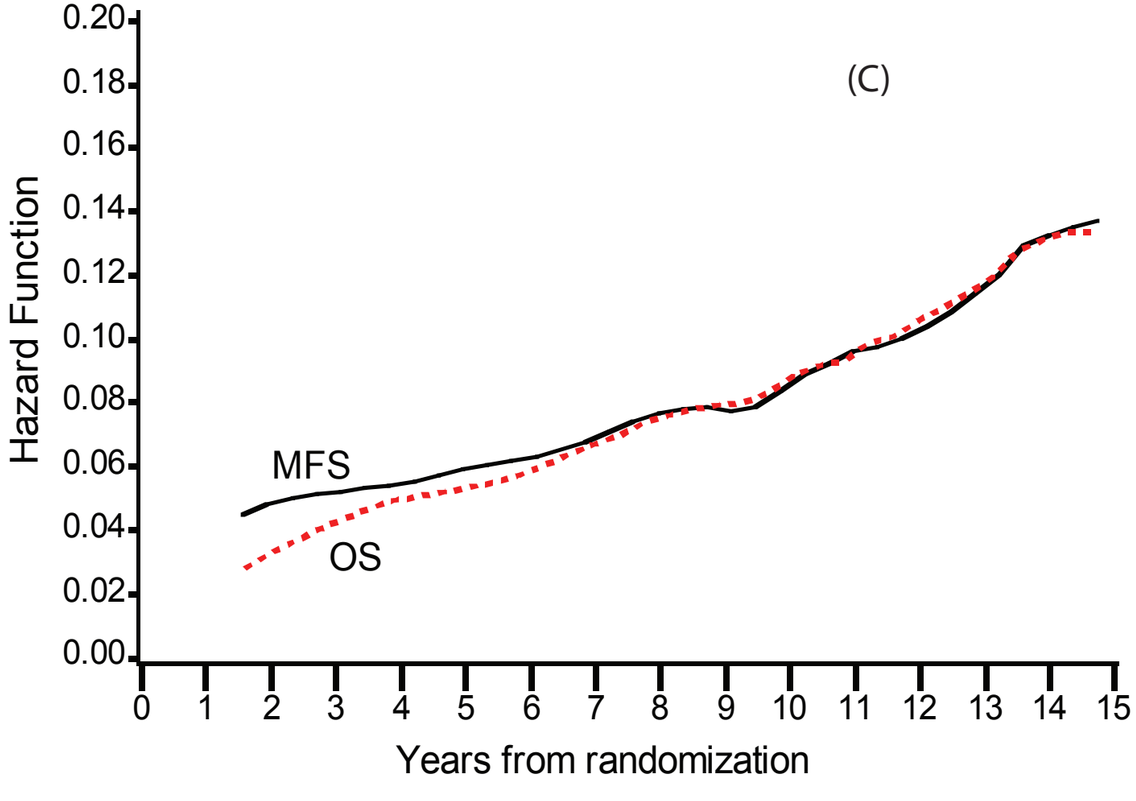

(B)
Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
DSS
TDR
No. (%) of events of 21,140
DSS: 2269 (11)
TDR: 4685 (22)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years from randomization
No. at risk
DSS 21140 19088 14682 9369 2425 726
TDR 21140 17472 12724 7491 2007 581

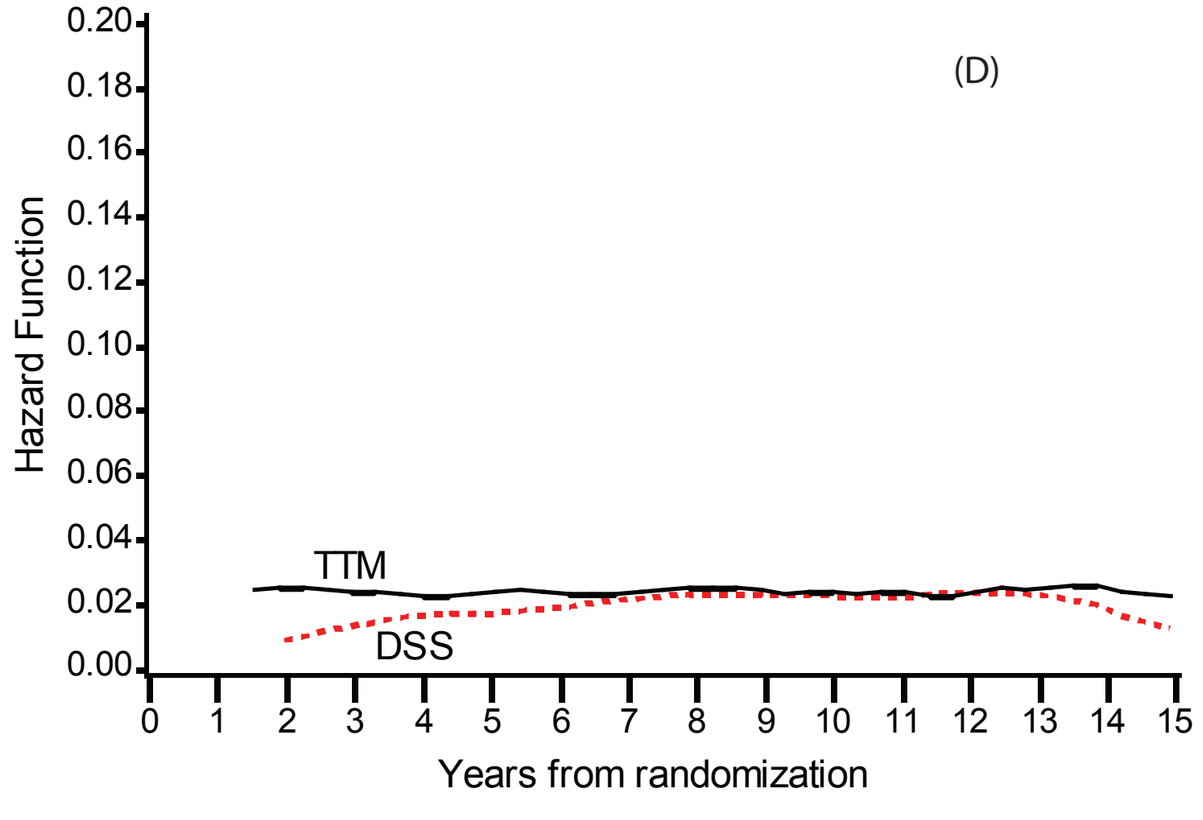

(C)
Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OS
MFS
No. (%) of events of 12,712
OS: 5350 (42)
MFS: 5733 (45)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years from randomization
No. at risk
OS 12712 11321 8162 4714 1853 726
MFS 12712 10831 7656 4394 1711 668


(D)
Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
DSS
TTM
No. (%) of events of 12,712
DSS: 1460 (11)
TTM: 2154 (17)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years from randomization
No. at risk
DSS 12712 11321 8162 4714 1853 726
TTM 12712 10831 7656 4394 1711 668

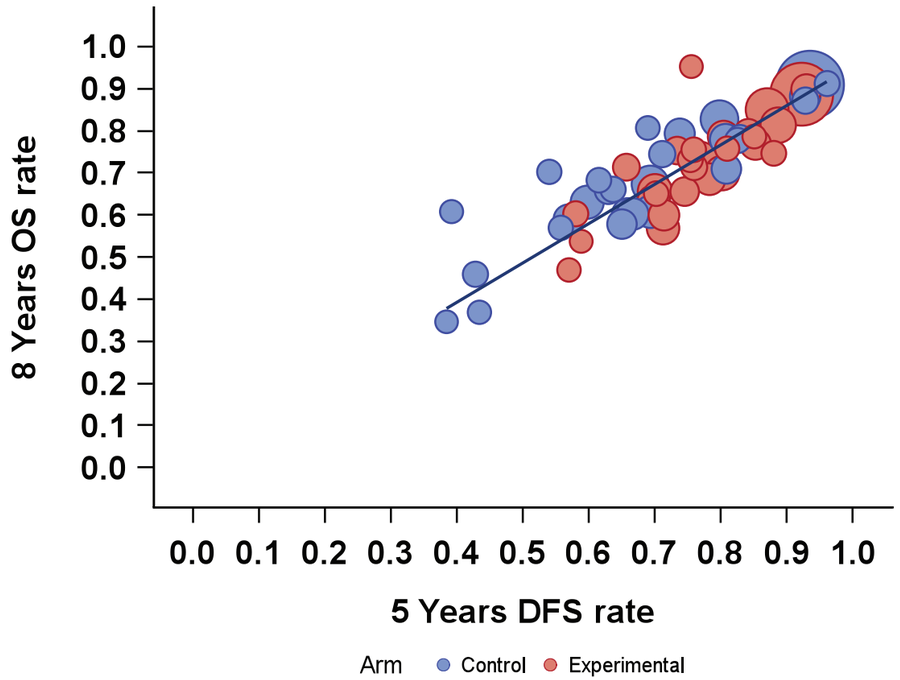

(A)
Hazard Function
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
DFS
OS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years from randomization


(B)
Hazard Function
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
TDR
DSS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years from randomization

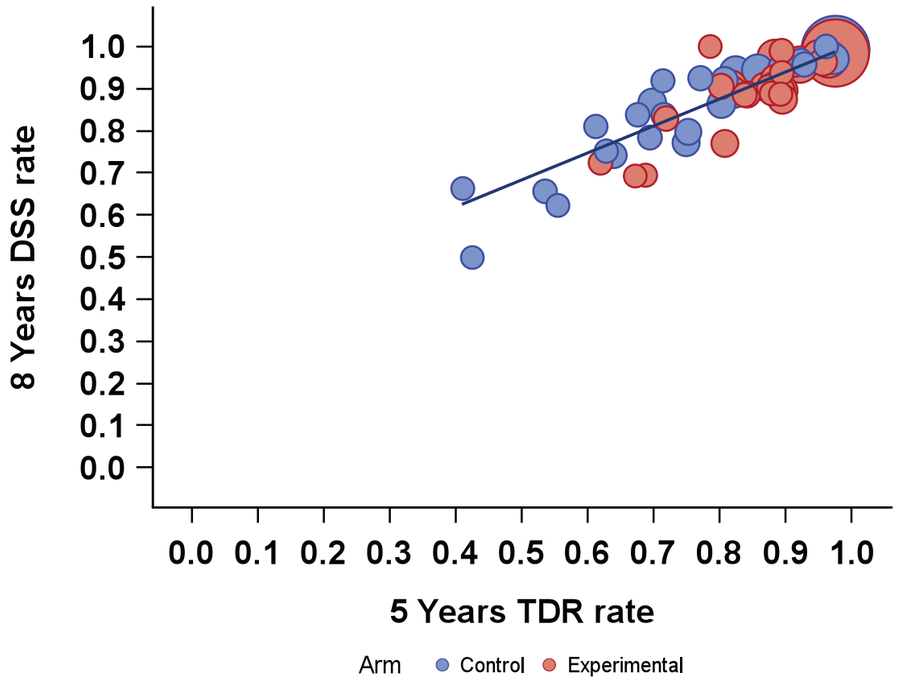

(C)
Hazard Function
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
MFS
OS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years from randomization


(D)
Hazard Function
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
TTM
DSS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years from randomization

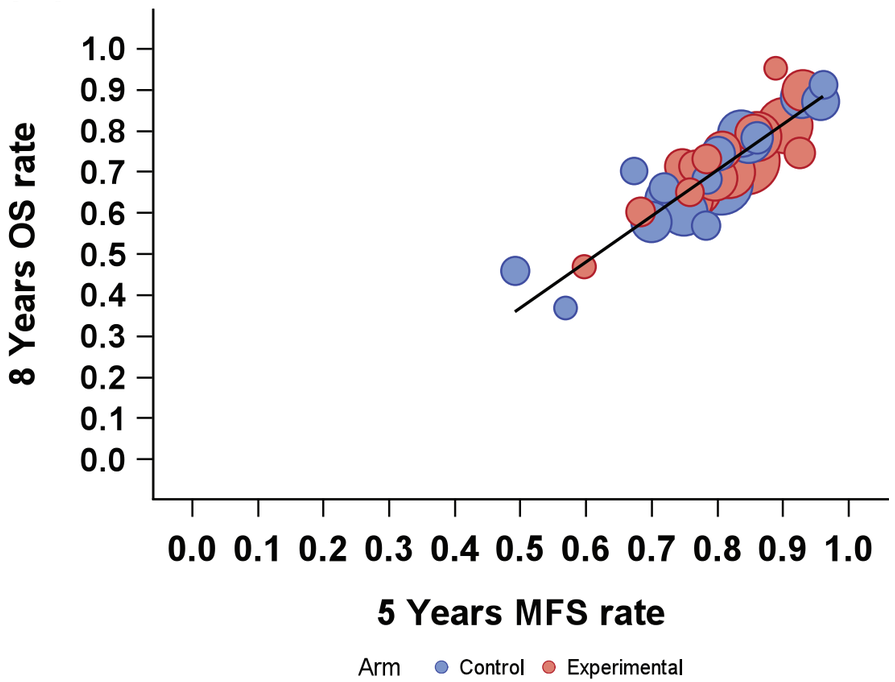
(A)


8 Years OS rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
5 Years DFS rate
Arm
Control
Experimental

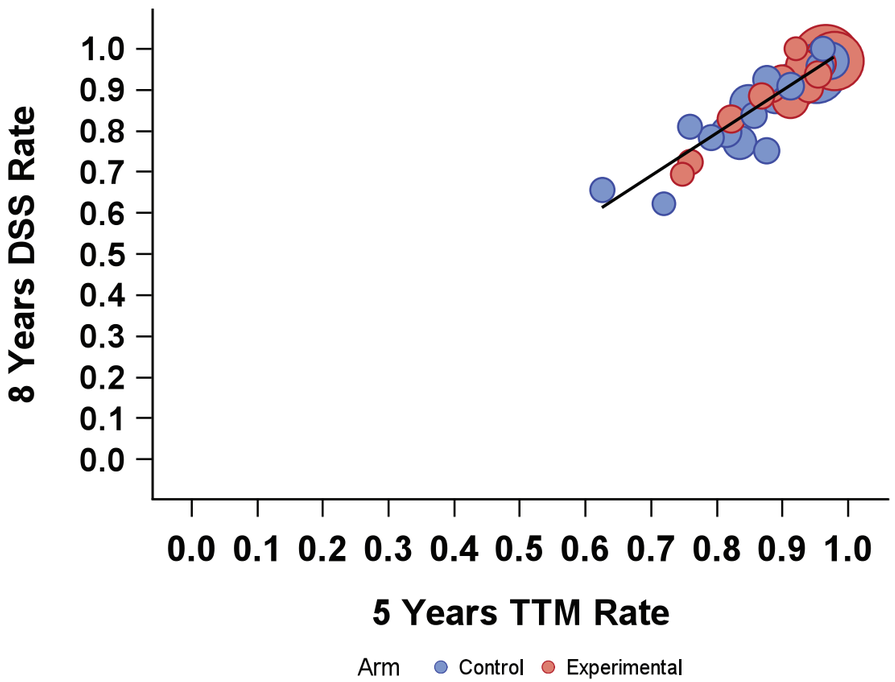
(B)


8 Years DSS rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
5 Years TDR rate
Arm
Control
Experimental

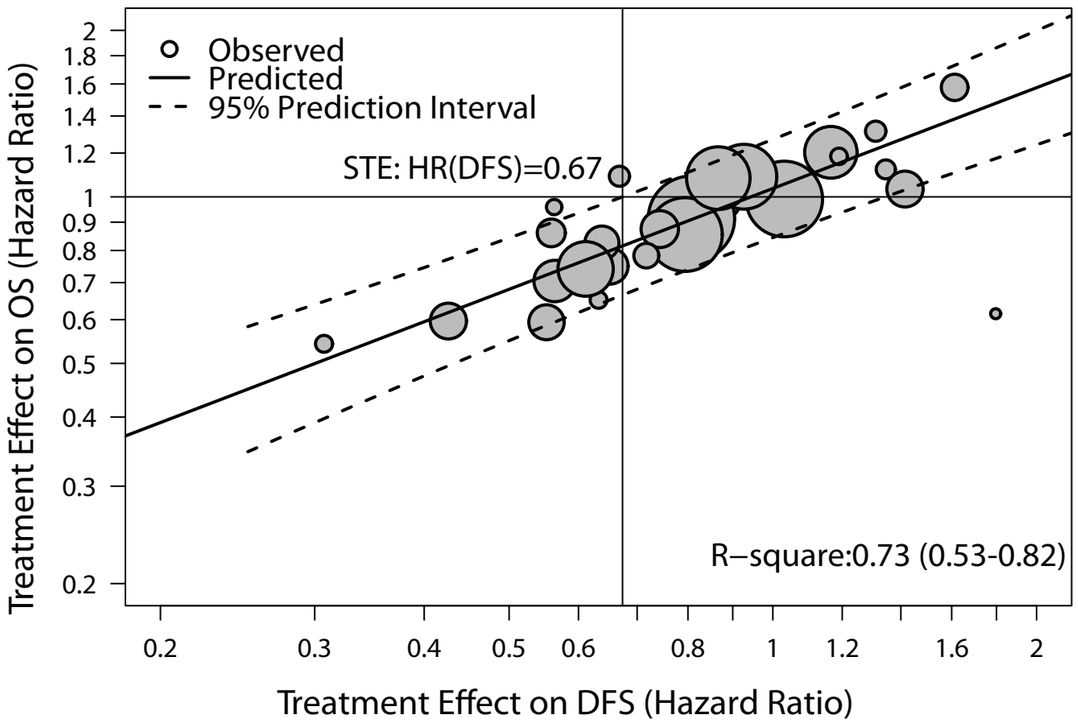
(C)


8 Years OS rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
5 Years MFS rate
Arm
Control
Experimental

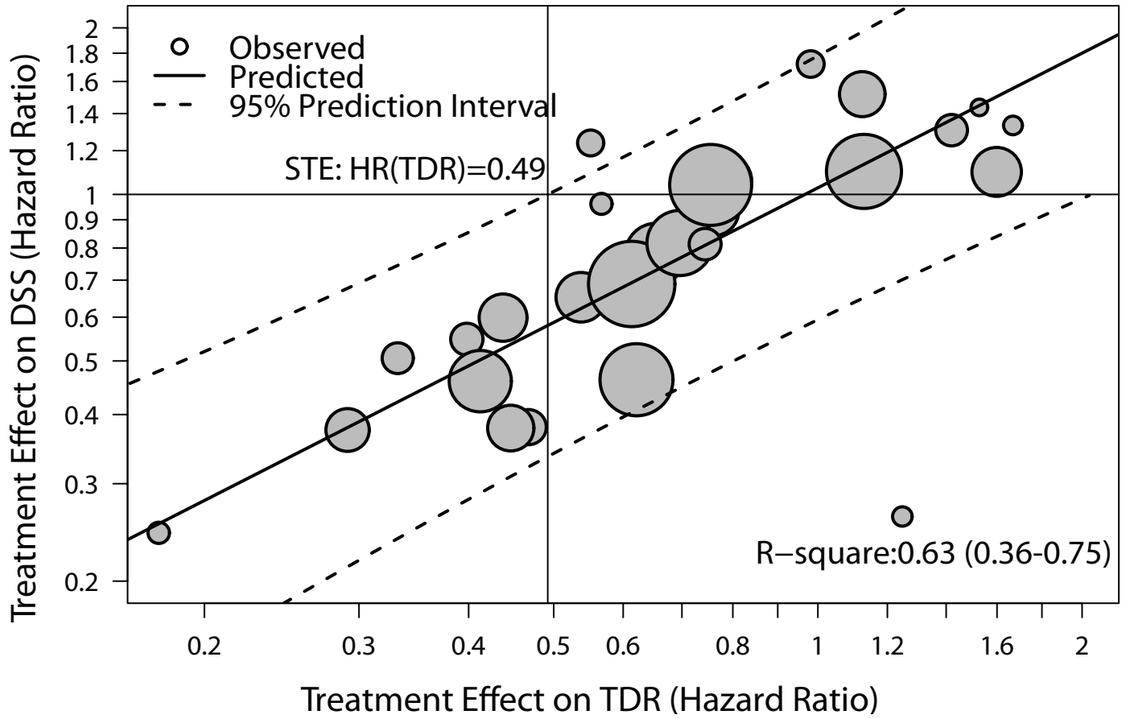
(D)


8 Years DSS Rate
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
5 Years TTM Rate
Arm
Control
Experimental

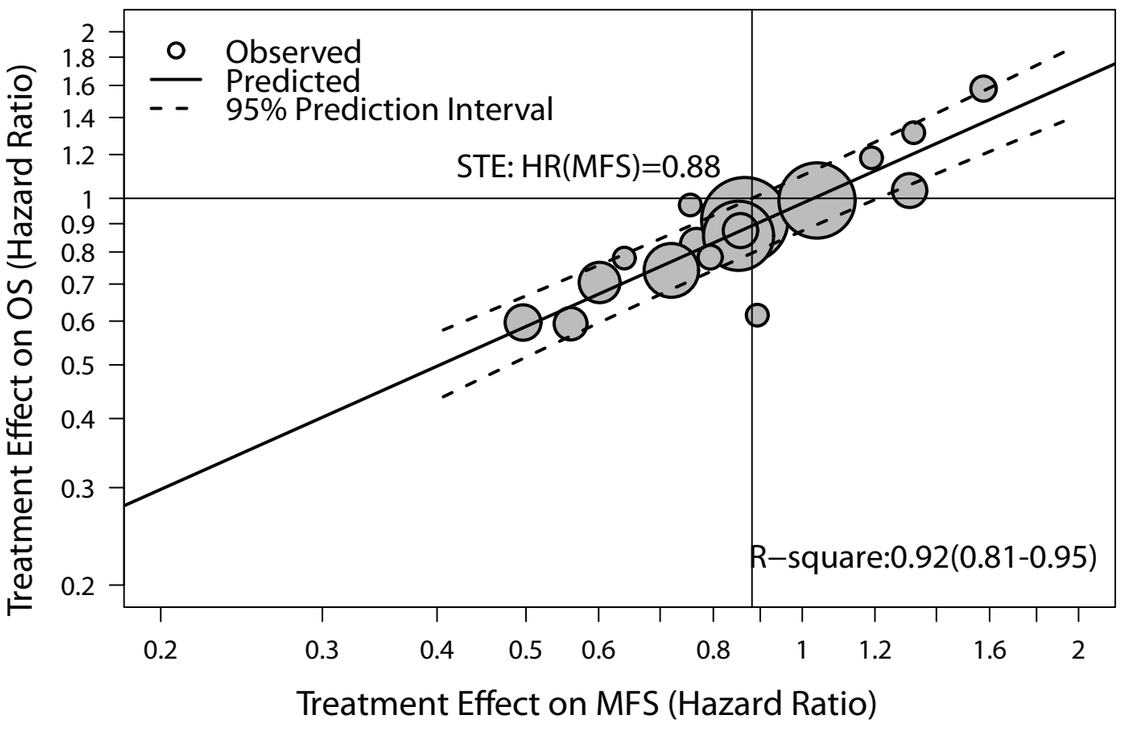
(A)


Observed
Predicted
95% Prediction Interval
STE: HR(DFS)=0.67
R−square:0.73 (0.53-0.82)
Treatment Effect on OS (Hazard Ratio)
Treatment Effect on DFS (Hazard Ratio)
2
1.8
1.6
1.4
1.2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.2
0.3
0.4
0.5
0.6
0.8
1
1.2
1.6
2

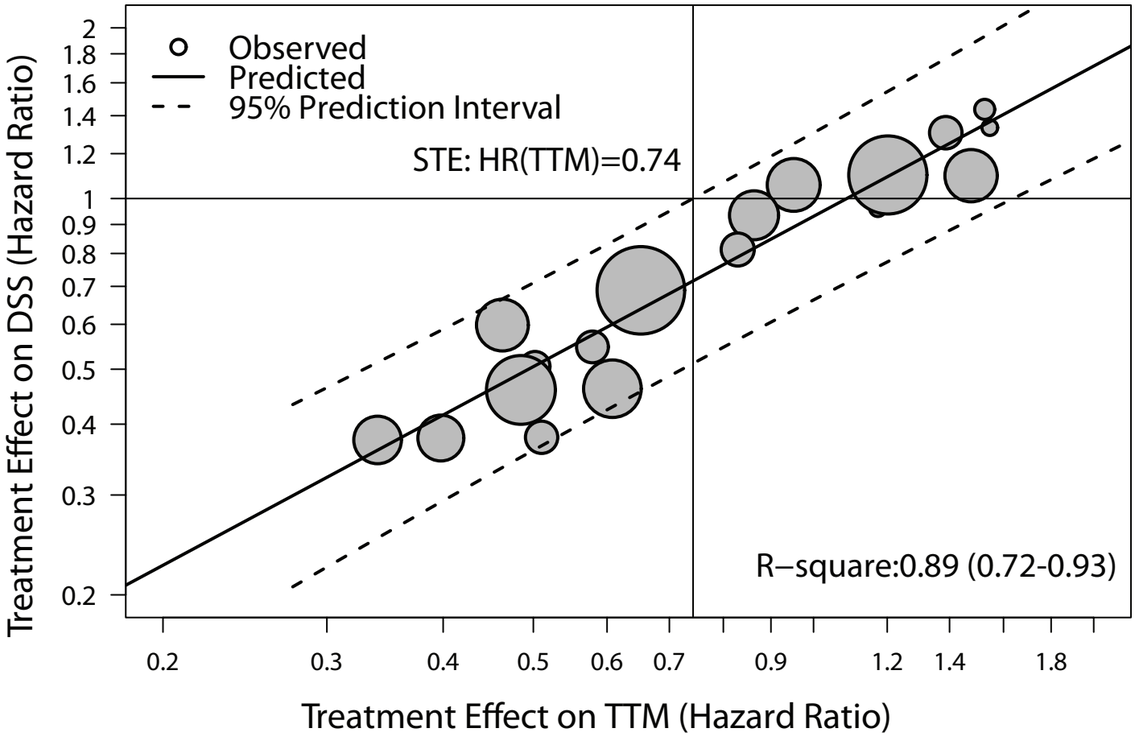
(B)

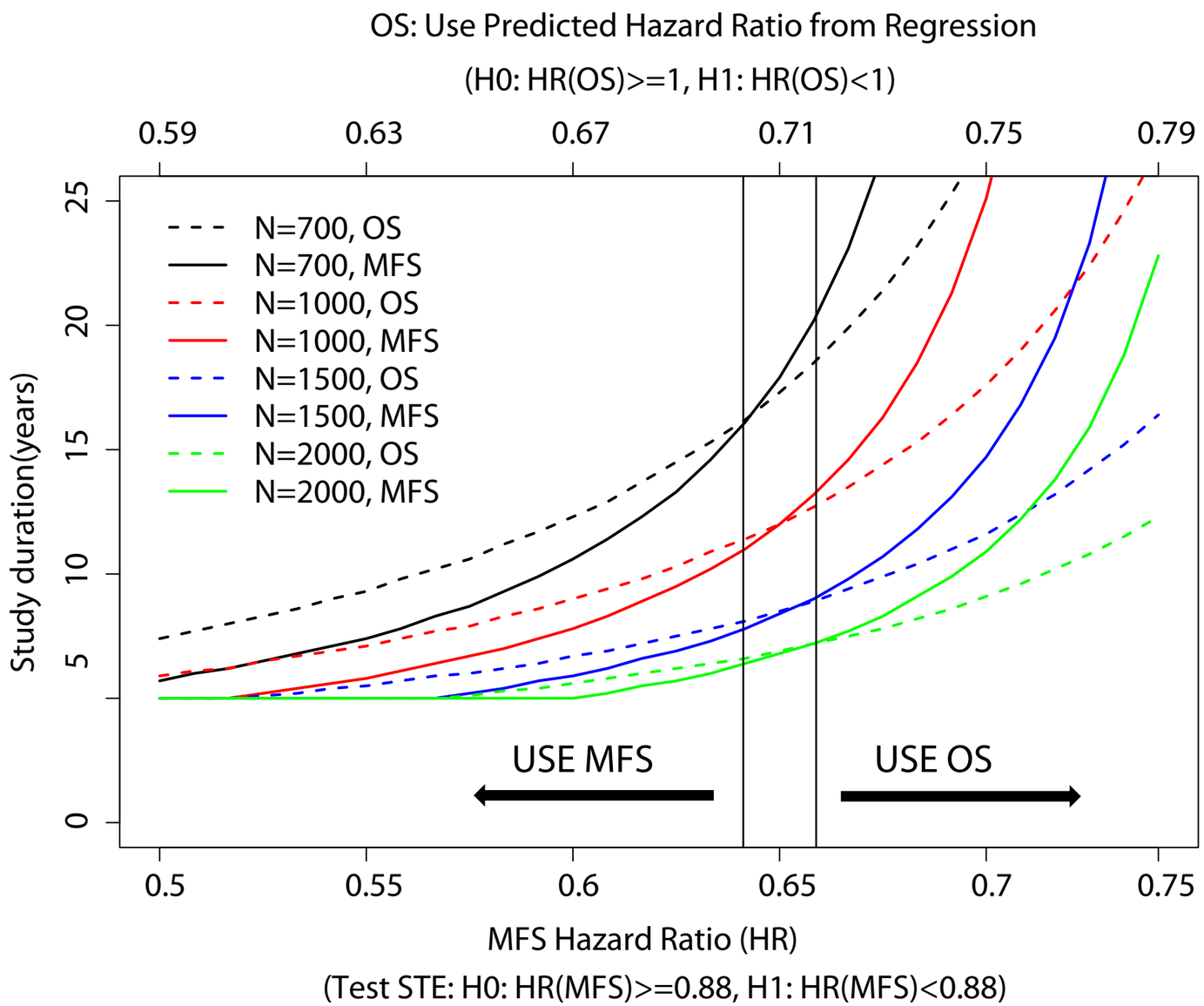

Observed
Predicted
95% Prediction Interval
STE: HR(TDR)=0.49
R−square:0.63 (0.36-0.75)
Treatment Effect on DSS (Hazard Ratio)
Treatment Effect on TDR (Hazard Ratio)
2
1.8
1.6
1.4
1.2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.2
0.3
0.4
0.5
0.6
0.8
1
1.2
1.6
2

(C)


Observed
Predicted
95% Prediction Interval
STE: HR(MFS)=0.88
R−square:0.92(0.81-0.95)
Treatment Effect on OS (Hazard Ratio)
Treatment Effect on MFS (Hazard Ratio)
2
1.8
1.6
1.4
1.2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.2
0.3
0.4
0.5
0.6
0.8
1
1.2
1.6
2

(D)


Observed
Predicted
95% Prediction Interval
STE: HR(TTM)=0.74
R−square:0.89 (0.72-0.93)
Treatment Effect on DSS (Hazard Ratio)
Treatment Effect on TTM (Hazard Ratio)
2
1.8
1.6
1.4
1.2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.2
0.3
0.4
0.5
0.6
0.7
0.9
1.2
1.4
1.8


OS: Use Predicted Hazard Ratio from Regression
(H0: HR(OS)>=1, H1: HR(OS)<1)
0.59
0.63
0.67
0.71
0.75
0.79
N=700, OS
N=700, MFS
N=1000, OS
N=1000, MFS
N=1500, OS
N=1500, MFS
N=2000, OS
N=2000, MFS
Study duration(years)
0
5
10
15
20
25
USE MFS
USE OS
0.5
0.55
0.6
0.65
0.7
0.75
MFS Hazard Ratio (HR)
(Test STE: H0: HR(MFS)>=0.88, H1: HR(MFS)<0.88)

Supplemental Material

Contents

Statistical methods

Trials and Study Population

24 and 19 trials with documented data on DFS and MFS were included respectively (see Figure S1 flow chart). One trial (TROG 96.01) was split into two study units given two experimental arms. Four trials (EPC 23:N=3292, EPC 24: N=3603, EPC 25: N=1218, and MRCPR04:N=508) included both surgery and radiation based patients and/or patients without primary therapy. We split each of these trials according to type of primary therapy as all other trials are primary therapy specific. Therefore, for the DFS analysis, the total number of units at the trial level was 31 with the split of TROG 96.01, three EPC studies and MRCPR04. For MFS analysis, the total number of units at the trial level was 21 with the split of TROG 96.01 and MRCPR04.

At the patient level, we used the ITT population, which includes all patients who were randomized to the study treatments.

Choice of methodology

Choice of methodology for establishing surrogacy has been discussed in our previous JNCI paper with a full Statistical Analysis Plan (SAP) published online (*https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djv261*). Please refer to SAP section 4.3 about discussing the strength and limitations of existing surrogacy evaluation approaches.

The meta-analytic approach we have used is widely accepted as a gold standard to evaluate surrogacy when data are available from several clinical trials [Ciani et al 20014]. Historically, the Prentice criteria have frequently been used when data were available from a single trial. [Buyse et al 2016]. Although Prentice's "full capture" criterion is very appealing conceptually, it requires techniques of causal inference (and the accompanying assumption of no unmeasured confounders) to be properly implemented. We acknowledge that the association measures presented here do not guarantee causation, but given the number of trials included and the strength of the associations, we think these results provide compelling evidence of surrogacy.

Ciani, O., Davis, S., Tappenden, P., Cantrell, A., Garside, R., Stein, K., Saad, E., Buyse, M., Taylor, R.S. (2014). Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers.*International Journal of Technology Assessment in Health Care* **30**, 1-13.
Buyse, M., Molenberghs, G., Paoletti, X., Oba, K., Alonso, A., Van der Elst, W. and Burzykowski, T. (2016). Statistical evaluation of surrogate endpoints with examples from cancer trials. *Biometrical Journal* **58**,104-32.

Copula Models

To estimate the association at the individual level between the distribution of OS and the surrogates, we fit bivariate copula models on individual patient data (Burzykowski et al, 2001). The Weibull distribution was assumed to evaluate the effect of treatment on the marginal distribution of each endpoint. Clayton, Hougaard and Plackett's copula models were considered. The Plackett's copula was chosen for the DFS endpoint and Clayton's copula was chosen for the MFS endpoint as they provided the best model fitness based on the AIC criteria (Reference: Fang et al. 2014 Comparison of Two Methods to Check Copula Fitting, IAENG International Journal of Applied Mathematics, 44:1). Patient level correlation was quantified by Kendall's Tau (range 0-1) estimated from the copula.

Weighted Linear Regression (WLR) Analyses

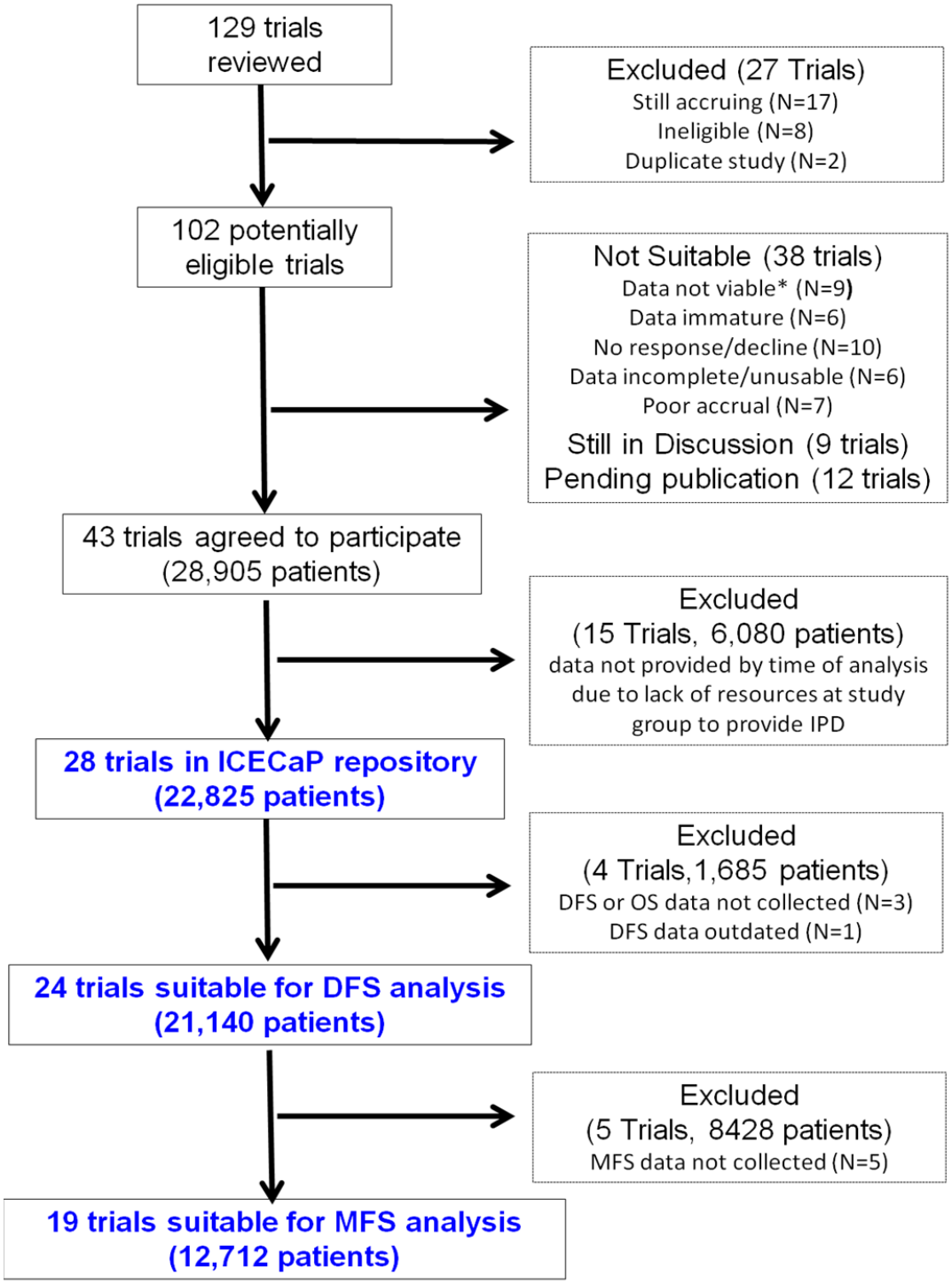
At the trial level, we performed WLR analysis between Kaplan-Meier estimates of OS at 8 years versus DFS and MFS at 5 years. Regressions were weighted by the inverse variances of the 5 year estimates of the surrogate. Likewise, we perform WLR analysis between treatment effects on OS versus treatment effects on DFS or MFS. Regressions were weighted by the inverse variances of the natural log (HR)) on the surrogate. Other weighting methods such as by number of events of the endpoints yielded similar results (data not shown).

For trial-level surrogacy analysis, the correlation between treatment effects on OS and DFS/MFS was also estimated using an error-in-variables regression model on the estimated treatment effects on the surrogate and on OS (Burzykowski et al, 2001). Such a model appropriately accounts for estimation errors in the treatment effects. However, due to the frequent convergence issue, such a model could not be fitted. Hence, we reported the trial level correlation based on the WLR analysis.

Leave-one-out cross validation

Model accuracy was assessed by a leave-one-out-cross validation (Supplemental methods). Each trial was left out once and the WLR (i.e. treatment effect on OS or DSS versus treatment effect on the ICE) was rebuilt on the remaining n-1 trials. This model was then applied to the left-out trial to obtain the predicted treatment effect (log [HR]) on OS or DSS, along with 95% prediction intervals (accounted for the weight of the left-out trial). R^2 was also calculated from the remaining n-1 trials model to evaluate the impact of a single trial on correlation between treatment effects on endpoints.

Figure S1: Flowchart of section and participation of randomized clinical trials for localized prostate cancer in ICECaP



*pre-PSA era, old databases, data inaccessible
ICECaP: Intermediate Clinical Endpoints in Cancer of the Prostate, OS: Overall survival, DFS: Disease free survival, MFS: Metastasis free survival. IPD: individual patient data

Table S1 Comparing trial characteristics between studies with and without individual patient data (IPD) in ICECaP database

	Trials without IPD in ICECaP (15 trials)		**Trials with IPD in ICECap (28 trials)**	
		%	N	%
No. randomized per trial, median (IQR)	300	138-677	439	270-1088
Follow-up, median(IQR), year	7.6	5.7-10.0	8.1	6.0-10.0
Study region				
US	6	40	9	32
Canada	3	20	8	29
UK	4	27	6	21
Europe	2	13	13	46
ANZ	.	.	5	18
Type of treatment				
RT-based	10	67	17	61
RP-based	1	7	6	21
RP & RT	4	27	4	14
Other	.	.	1	4
Year of first enrollment				
1980-1989	5	33	4	14
1990-1994	3	20	3	11
1995-1999	3	20	17	61
2000-2004	2	13	2	7
2005-2009	.	.	2	7
Unknown	2	13	.	.
Year of last enrollment				
1980-1989	2	13	.	.
1990-1994	4	27	1	4
1995-1999	4	27	10	36
2000-2004	1	7	13	46
2005-2009	1	7	3	11
2010-present	.	.	1	4
Unknown	3	20	.	.
OS as the primary endpoint	13	87	20	71
Endpoints include				
OS	14	93	28	100
PCSM	8	53	17	61
TTM/MFS	9	60	22	79
PSA progression	10	67	23	82

Table S2 Trial and patient characteristics for DFS and MFS analysis

	DFS analysis (N=21,140)		MFS analysis (N=12,712)	
	N	**%**	**N**	**%**
Year of Randomization				
1987-1994	2165	10	2067	16
1995-1999	13982	66	5742	45
2000-2004	3820	18	3730	29
2005-2011	1173	6	1173	9
Type of treatment				
Comparing primary therapies	201	1	201	2
RP +/- Adjuvant RT or ADT	5518	26	966	8
RT dose	1366	6	1149	9
RT +/- ADT	11619	55	10249	81
No primary therapy +/- ADT/other	2436	12	147	1
Age at randomization				
64 or younger	6043	29	3223	25
65-74	11857	56	7310	58
75 or older	3226	15	2174	17
Unknown	14	0.07	5	0.04
NCCN risk group				
Low	2092	10	1094	9
Low/Intermediate*	702	3	100	1
Intermediate	5898	28	3497	28
High	11502	54	7586	60
Unknown	946	4	435	3
High risk**				
No	7225	34	4158	33
Yes	13324	63	8394	66
Unknown	591	3	160	1

*T2 disease but T2 subtype is not determined.

**defined as high risk if patient had one of these features: high risk by NCCN or D'Amico criteria, high risk by pathological criteria (pathological Gleason>7, or seminal vesicles involvement, or ≥pT3b stage) or Pathology N1 (ECOG3886 only).

Table S3 Trials included for DFS (N=24) and MFS (N=19) analysis. Trials are ordered by type of local therapy.

Study	Stage (Clinical |Pathological)	Year Enrolled	Type of Local therapy	Control Arm	Experimental Arm	Total N	Median Follow-up, years	MFS Analysis
RP & RT/other								
EPC Trial23	T1b-4	1995-1998	RP/RT	placebo	Casodex	3292	10.9	No
EPC Trial24	T1b-4	1995-1998	RP/RT/None	placebo	Casodex	3603	10.4	No
EPC Trial25	T1b-4	1995-1998	RP/RT/None	placebo	Casodex	1218	11.1	No
Cryo vs RT Ontario	T2c-3b	1999-2002	RT/Cryoablation	NADT 6mo + RT	NADT 6mo + Cryoablation	64	10.7	Yes
RP vs RT DiStasi et al	T1-2	1997-2001	RP/RT	RP	RT	137	13.1	Yes
RP-based trials								
ECOG3886	T1b-2|pN1	1988-1993	RP	Observation (Deferred ADT)	Immediate continuous ADT	98	11.8	No
GermanARO9602	T1-3|pT3N0	1997-2004	RP	Observation	Adjuvant RT	307	9.7	Yes
SWOG8794	T1-4|pT3N0M0	1988-1997	RP	Observation	Adjuvant RT	431	14.0	Yes
TAX3501	T1-4	2005-2007	RP	Observation (Deferred)	Immediate 18mo Leuprolide +/- 6cycle Taxotere	228	3.4	Yes
RT-based trials: comparing RT doses								
Australian Study Yeoh et al	T1-2	1996-2003	RT	RT 64Gy (32 fractions/6.5wk)	RT 55Gy (20 fractions /4 wks)	217	9.3	No
GETUG06	T1b-3a	1999-2002	RT	RT 70 Gy	RT 80 Gy	306	5.1	Yes
MRCRT01	T1b-3a	1998-2002	RT	3-6mo NADT + RT 64 Gy	3-6mo NADT + RT 74 Gy	843	10.0	Yes
RT-based trials: comparing duration of ADT								

Study	Stage (Clinical |Pathological)	Year Enrolled	Type of Local therapy	Control Arm	Experimental Arm	Total N	Median Follow-up, years	MFS Analysis
EORTC22863	T1-2 WHO-G3 or T3-4	1987-1995	RT	RT	RT+ AADT 3yr	415	9.4	Yes
EORTC22961	T1c-4	1997-2001	RT	NADT 6mo + RT	NADT 6mo + RT + AADT 2.5yr	970	6.0	Yes
EORTC22991	T1b-2a	2001-2008	RT	RT	RT + AADT 6mo	819	7.2	Yes
French study Mottet et al	T3-4	2000-2003	RT/None	ADT 3yr	ADT 3yr + RT	264	8.4	Yes
GICOR-DART01/05	T1c-3b	2005-2010	RT	NADT 4mo + RT	NADT 4mo + RT + AADT 2yr	352	4.5	Yes
ICORG9701	T1-4	1997-2001	RT	NADT 4mo + RT	NADT 8mo + RT	276	12.0	Yes
NCIC/MRC-PR3	T2-4	1995-2005	RT/None	ADT lifelong	ADT lifelong + RT	1205	8.0	Yes
RTOG9202	T2c-4	1992-1995	RT	NADT 4mo + RT	NADT 4mo + RT + AADT 2yr	1520	19.6	Yes
RTOG9408	T1b-2b	1994-2001	RT	RT	NADT 4mo + RT	1979	9.9	Yes
RTOG9413	T1c-4	1995-1999	RT	RT + AADT 4mo	NADT 4mo + RT	1270	17.1	Yes
TROG9601	T2b-4	1996-2000	RT	RT	NADT 3/6 mo + RT	818	11.4	Yes
RT-based trials: other								
MRCPR04	T2-4	1994-1997	RT/None	Placebo	Clodronate 5yr	508	8.7	Yes

RP: Radical prostatectomy, RT: Radiation therapy, ADT: Androgen deprivation therapy, NADT: Neoadjuvant ADT, AADT: Adjuvant ADT
Mo: Month yr: year

References of included trials

RP & RT/other

1. (EPC23/24/25) Iversen P, McLeod DG, See WA, Morris T, Armstrong J, Wirth MP; Casodex Early Prostate Cancer Trialists' Group. Antiandrogen monotherapy in patients with localized or locally advanced prostate cancer: final results from the bicalutamide Early Prostate Cancer programme at a median follow-up of 9.7 years. BJU Int. 2010. 105(8):1074-81.
2. Chin JL, Al-Zahrani AA, Autran-Gomez AM, Williams AK, Bauman G. Extended followup oncologic outcome of randomized trial between cryoablation and external beam therapy for locally advanced prostate cancer (T2c-T3b). J Urol. 2012. 188(4):1170-5.
3. S. M. Di Stasi, A. Giannantoni, M. Valenti, L. Storti, F. Attisani, T. Palloni, E. A. Jannini, M. Bibas and G. Zampa. Multicenter, randomized, phase III trial comparing radical retropubic prostatectomy with conventional external beam radiotherapy for localized prostate cancer: An interim report. JCO, 2006 ASCO Annual Meeting Proceedings (Post-Meeting Edition).Vol 24, No 18S (June 20 Supplement), 2006: 4607

RP-based trials

4. Messing EM, Manola J, Yao J, Kiernan M, Crawford D, Wilding G, di'SantAgnese PA, Trump D; Eastern Cooperative Oncology Group study EST 3886.. Immediate versus deferred androgen deprivation treatment in patients with node-positive prostate cancer after radical prostatectomy and pelvic lymphadenectomy. Lancet Oncol. 2006 Jun;7(6):472-9.
5. Wiegel T, Bartkowiak D, Bottke D, Bronner C, Steiner U, Siegmann A, Golz R, Störkel S, Willich N, Semjonow A, Stöckle M, Rübe C, Rebmann U, Kälble T,
   Feldmann HJ, Wirth M, Hofmann R, Engenhart-Cabillic R, Hinke A, Hinkelbein W, Miller K. Adjuvant radiotherapy versus wait-and-see after radical prostatectomy: 10-year follow-up of the ARO 96-02/AUO AP 09/95 trial. Eur Urol. 2014. Aug;66(2):243-50.
6. Thompson IM, Tangen CM, Paradelo J, Lucia MS, Miller G, Troyer D, Messing E, Forman J, Chin J, Swanson G, Canby-Hagino E, Crawford ED. Adjuvant radiotherapy for pathological T3N0M0 prostate cancer significantly reduces risk of metastases and improves survival: long-term followup of a randomized clinical trial. J Urol. 2009 Mar;181(3):956-62.
7. Schweizer MT, Huang P, Kattan MW, Kibel AS, de Wit R, Sternberg CN, Epstein JI, Eisenberger MA. Adjuvant leuprolide with or without docetaxel in patients with high-risk prostate cancer after radical prostatectomy (TAX-3501): important lessons for future trials. Cancer. 2013 Oct 15;119(20):3610-8.

RT-based trials: comparing RT doses

8. Yeoh EE, Botten RJ, Butters J, Di Matteo AC, Holloway RH, Fowler J. Hypofractionated versus conventionally fractionated radiotherapy for prostate
   carcinoma: final results of phase III randomized trial. Int J Radiat Oncol Biol Phys. 2011 Dec 1;81(5):1271-8.
9. Beckendorf V, Guerif S, Le Prisé E, Cosset JM, Bougnoux A, Chauvet B, Salem N,Chapet O, Bourdain S, Bachaud JM, Maingon P, Hannoun-Levi JM, Malissard L, Simon JM, Pommier P, Hay M, Dubray B, Lagrange JL, Luporsi E, Bey P. 70 Gy versus 80 Gy in localized prostate cancer: 5-year results of GETUG 06 randomized trial. Int J Radiat Oncol Biol Phys. 2011 Jul 15;80(4):1056-63.
10. Dearnaley DP, Jovic G, Syndikus I, Khoo V, Cowan RA, Graham JD, Aird EG,Bottomley D, Huddart RA, Jose CC, Matthews JH, Millar JL, Murphy C, Russell JM, Scrase CD, Parmar MK, Sydes MR. Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: long-term results from the MRC RT01 randomised controlled trial. Lancet Oncol. 2014 Apr;15(4):464-73.

RT-based trials: comparing duration of ADT

11. Bolla M, Van Tienhoven G, Warde P, Dubois JB, Mirimanoff RO, Storme G, Bernier J, Kuten A, Sternberg C, Billiet I, Torecilla JL, Pfeffer R, Cutajar CL, Van der Kwast T, Collette L. External irradiation with or without long-term androgen suppression for prostate cancer with high metastatic risk: 10-year results of an EORTC randomised study. Lancet Oncol. 2010 Nov;11(11):1066-73.

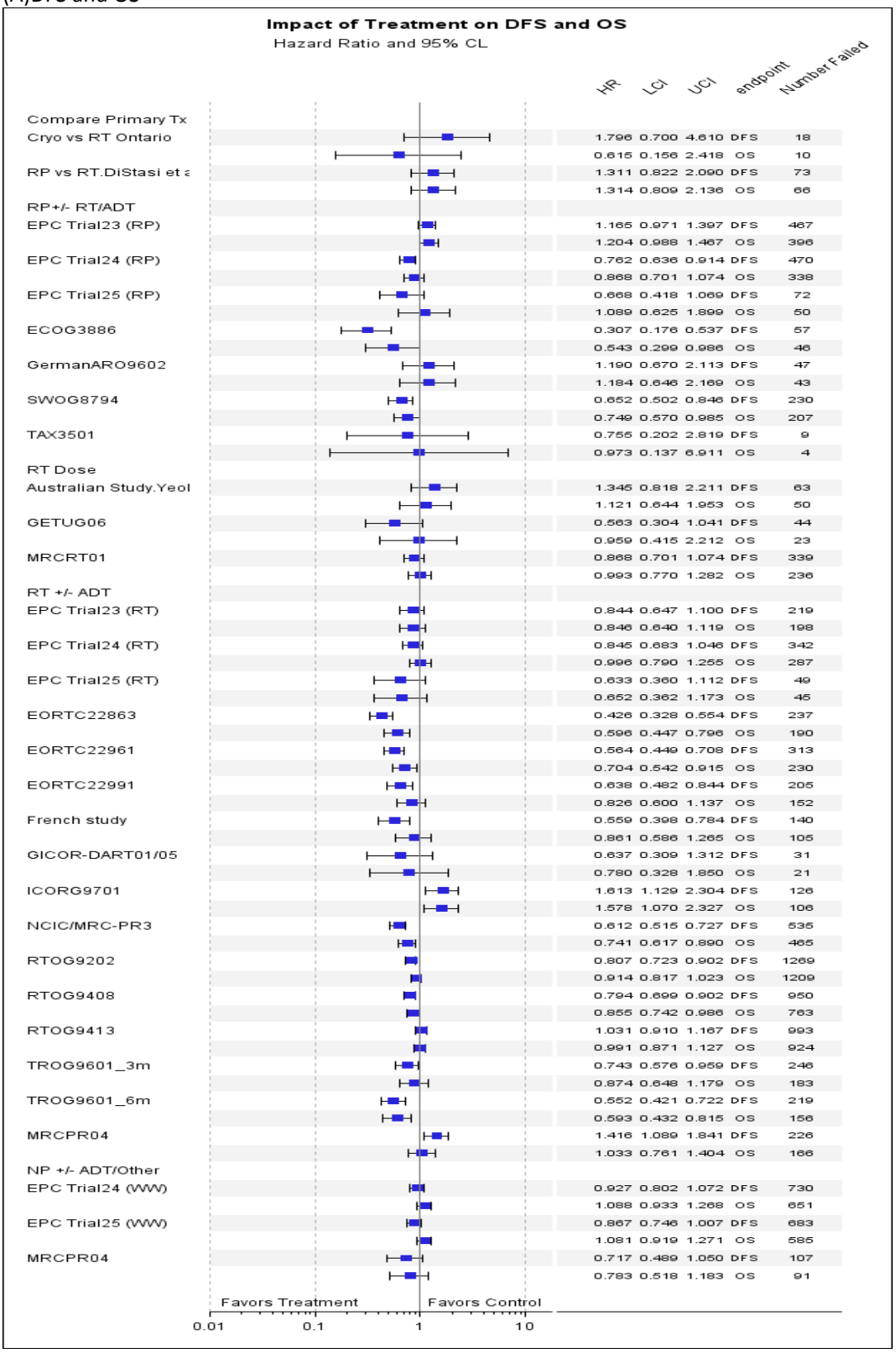
12. Bolla M, de Reijke TM, Van Tienhoven G, Van den Bergh AC, Oddens J, Poortmans PM, Gez E, Kil P, Akdas A, Soete G, Kariakine O, van der Steen-Banasik EM, Musat E, Piérart M, Mauer ME, Collette L; EORTC Radiation Oncology Group and Genito-Urinary Tract Cancer Group. Duration of androgen suppression in the treatment of prostate cancer. N Engl J Med. 2009 Jun 11;360(24):2516-27.
13. Bolla M, Maingon P, Carrie C, Villa S, Kitsios P, Poortmans PM, Sundar S, van der Steen-Banasik EM, Armstrong J, Bosset JF, Herrera FG, Pieters B, Slot A, Bahl A, Ben-Yosef R, Boehmer D, Scrase C, Renard L, Shash E, Coens C, van den Bergh AC, Collette L. Short Androgen Suppression and Radiation Dose Escalation for Intermediate- and High-Risk Localized Prostate Cancer: Results of EORTC Trial 22991. J Clin Oncol. 2016 May 20;34(15):1748-56.
14. Mottet N, Peneau M, Mazeron JJ, Molinie V, Richaud P. Addition of radiotherapy to long-term androgen deprivation in locally advanced prostate cancer: an open randomised phase 3 trial. Eur Urol. 2012 Aug;62(2):213-9.
15. Zapatero A, Guerrero A, Maldonado X, Alvarez A, Gonzalez San Segundo C, Cabeza Rodríguez MA, Macias V, Pedro Olive A, Casas F, Boladeras A, de Vidales CM, Vazquez de la Torre ML, Villà S, Perez de la Haza A, Calvo FA. High-dose radiotherapy with short-term or long-term androgen deprivation in localized prostate cancer (DART01/05 GICOR): a randomised, controlled, phase 3 trial. Lancet Oncol. 2015 Mar;16(3):320-7.
16. Armstrong JG, Gillham CM, Dunne MT, Fitzpatrick DA, Finn MA, Cannon ME, Taylor JC, O'Shea CM, Buckney SJ, Thirion PG. A randomized trial (Irish clinical oncology research group 97-01) comparing short versus protracted neoadjuvant hormonal therapy before radiotherapy for localized prostate cancer. Int J Radiat Oncol Biol Phys. 2011 Sep 1;81(1):35-45.
17. Warde P, Mason M, Ding K, Kirkbride P, Brundage M, Cowan R, Gospodarowicz M, Sanders K, Kostashuk E, Swanson G, Barber J, Hiltz A, Parmar MK, Sathya J,Anderson J, Hayter C, Hetherington J, Sydes MR, Parulekar W; NCIC CTG PR.3/MRC UK PR07 investigators.. Combined androgen deprivation therapy and radiation therapy for locally advanced prostate cancer: a randomised, phase 3 trial. Lancet. 2011 Dec 17;378(9809):2104-11.
18. Horwitz EM, Bae K, Hanks GE, Porter A, Grignon DJ, Brereton HD, Venkatesan V, Lawton CA, Rosenthal SA, Sandler HM, Shipley WU. 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 May 20;26(15):2497-504.
19. Jones CU, Hunt D, McGowan DG, Amin MB, Chetner MP, Bruner DW, Leibenhaut MH,Husain SM, Rotman M, Souhami L, Sandler HM, Shipley WU. Radiotherapy and short-term androgen deprivation for localized prostate cancer. N Engl J Med. 2011. Jul 14;365(2):107-18.
20. Lawton CA, DeSilvio M, Roach M 3rd, Uhl V, Kirsch R, Seider M, Rotman M, Jones C, Asbell S, Valicenti R, Hahn S, Thomas CR Jr. An update of the phase III trial comparing whole pelvic to prostate only radiotherapy and neoadjuvant to adjuvant total androgen suppression: updated analysis of RTOG 94-13, with emphasis on unexpected hormone/radiation interactions. Int J Radiat Oncol Biol Phys. 2007. 1;69(3):646-55.
21. Denham JW, Steigler A, Lamb DS, Joseph D, Turner S, Matthews J, Atkinson C,North J, Christie D, Spry NA, Tai KH, Wynne C, D'Este C. Short-term neoadjuvant androgen deprivation and radiotherapy for locally advanced prostate cancer:10-year data from the TROG 96.01 randomised trial. Lancet Oncol. 2011.May;12(5):451-9.

RT-based trials: other

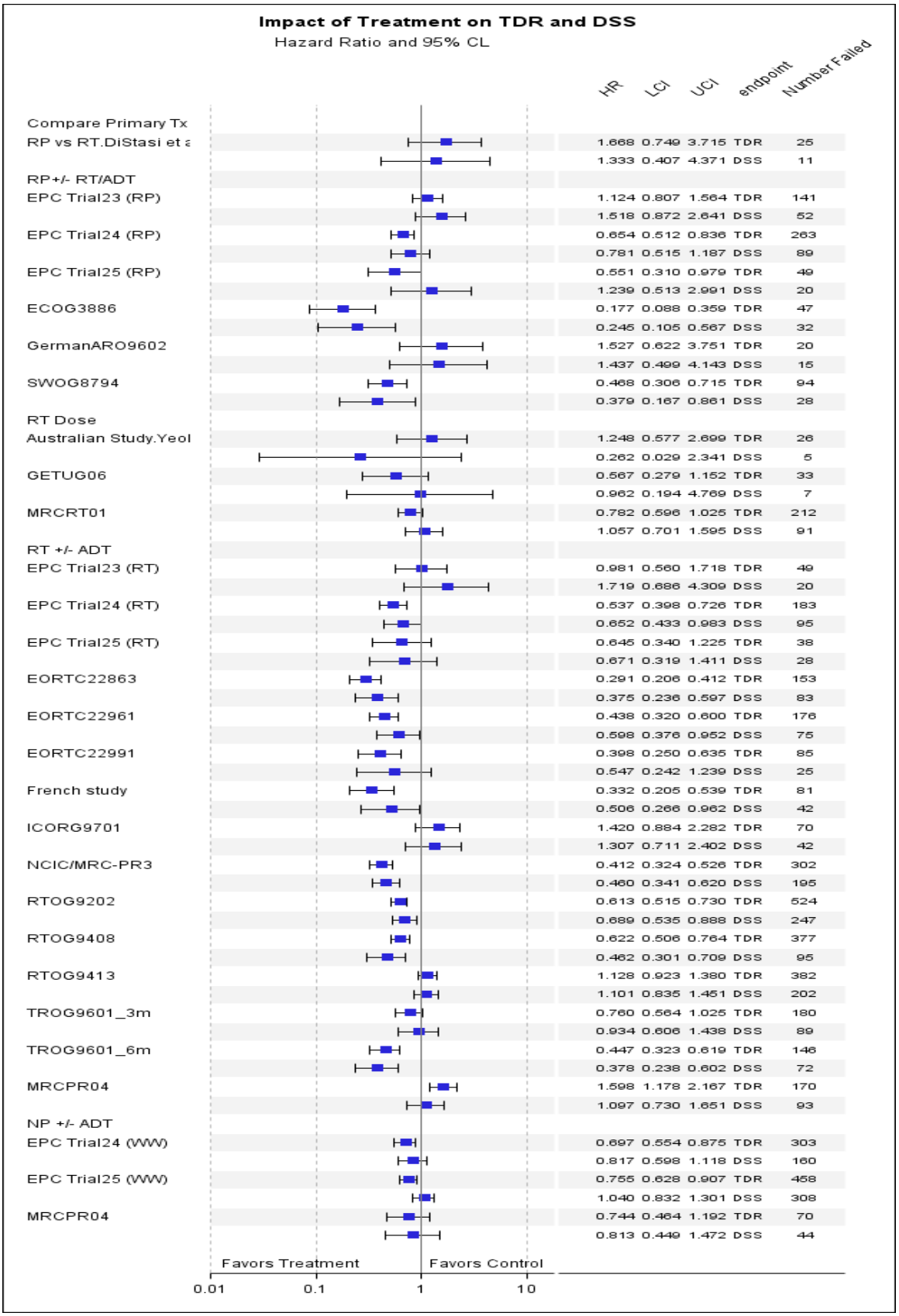
22. Dearnaley DP, Mason MD, Parmar MK, Sanders K, Sydes MR. Adjuvant therapy with oral sodium clodronate in locally advanced and metastatic prostate cancer: long-term overall survival results from the MRC PR04 and PR05 randomised controlled trials. Lancet Oncol. 2009 Sep;10(9):872-6.

Figure S2 Forest plots of study specific treatment effects (HR) on endpoints. Trials are ordered by type of therapy: (1) Compare primary therapy, (2) RP+/- adjuvant RT/ADT, (3)RT-based: comparing RT dose, (4) RT-based: comparing ADT duration, (5)No primary therapy(NP) +/- ADT/other

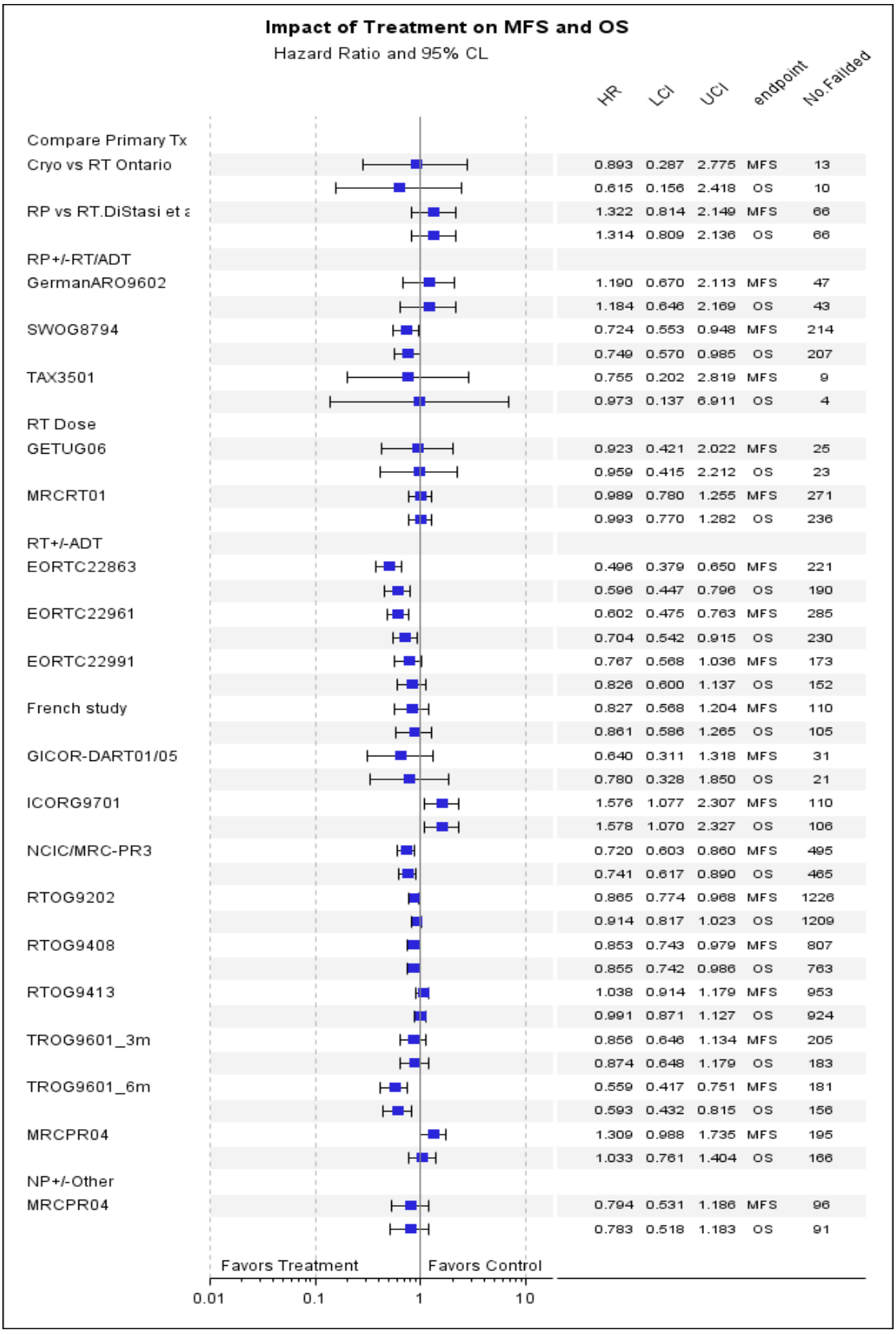
(A)DFS and OS



(B)TDR and DSS


Impact of Treatment on TDR and DSS
Hazard Ratio and 95% CL
HR LCI UCI endpoint Number Failed
Compare Primary Tx
RP vs RT.DiStasi et a 1.668 0.749 3.715 TDR 25
1.333 0.407 4.371 DSS 11
RP+/- RT/ADT
EPC Trial23 (RP) 1.124 0.807 1.564 TDR 141
1.518 0.872 2.641 DSS 52
EPC Trial24 (RP) 0.654 0.512 0.836 TDR 263
0.781 0.515 1.187 DSS 89
EPC Trial25 (RP) 0.551 0.310 0.979 TDR 49
1.239 0.513 2.991 DSS 20
ECOG3886 0.177 0.088 0.359 TDR 47
0.245 0.105 0.567 DSS 32
GermanARO9602 1.527 0.622 3.751 TDR 20
1.437 0.499 4.143 DSS 15
SWOG8794 0.468 0.306 0.715 TDR 94
0.379 0.167 0.861 DSS 28
RT Dose
Australian Study.Yeol 1.248 0.577 2.699 TDR 26
0.262 0.029 2.341 DSS 5
GETUG06 0.567 0.279 1.152 TDR 33
0.962 0.194 4.769 DSS 7
MRCRT01 0.782 0.596 1.025 TDR 212
1.057 0.701 1.595 DSS 91
RT +/- ADT
EPC Trial23 (RT) 0.981 0.560 1.718 TDR 49
1.719 0.686 4.309 DSS 20
EPC Trial24 (RT) 0.537 0.398 0.726 TDR 183
0.652 0.433 0.983 DSS 95
EPC Trial25 (RT) 0.645 0.340 1.225 TDR 38
0.671 0.319 1.411 DSS 28
EORTC22863 0.291 0.206 0.412 TDR 153
0.375 0.236 0.597 DSS 83
EORTC22961 0.438 0.320 0.600 TDR 176
0.598 0.376 0.952 DSS 75
EORTC22991 0.398 0.250 0.635 TDR 85
0.547 0.242 1.239 DSS 25
French study 0.332 0.205 0.539 TDR 81
0.506 0.266 0.962 DSS 42
ICORG9701 1.420 0.884 2.282 TDR 70
1.307 0.711 2.402 DSS 42
NCIC/MRC-PR3 0.412 0.324 0.526 TDR 302
0.460 0.341 0.620 DSS 195
RTOG9202 0.613 0.515 0.730 TDR 524
0.689 0.535 0.888 DSS 247
RTOG9408 0.622 0.506 0.764 TDR 377
0.462 0.301 0.709 DSS 95
RTOG9413 1.128 0.923 1.380 TDR 382
1.101 0.835 1.451 DSS 202
TROG9601_3m 0.760 0.564 1.025 TDR 180
0.934 0.606 1.438 DSS 89
TROG9601_6m 0.447 0.323 0.619 TDR 146
0.378 0.238 0.602 DSS 72
MRCPR04 1.598 1.178 2.167 TDR 170
1.097 0.730 1.651 DSS 93
NP +/- ADT
EPC Trial24 (WW) 0.697 0.554 0.875 TDR 303
0.817 0.598 1.118 DSS 160
EPC Trial25 (WW) 0.755 0.628 0.907 TDR 458
1.040 0.832 1.301 DSS 308
MRCPR04 0.744 0.464 1.192 TDR 70
0.813 0.449 1.472 DSS 44
Favors Treatment
Favors Control
0.01
0.1
1
10

(C)MFS and OS


Impact of Treatment on MFS and OS
Hazard Ratio and 95% CL
HR LCI UCI endpoint No.Falided
Compare Primary Tx
Cryo vs RT Ontario 0.893 0.287 2.775 MFS 13
0.615 0.156 2.418 OS 10
RP vs RT.DiStasi et a 1.322 0.814 2.149 MFS 66
1.314 0.809 2.136 OS 66
RP+/-RT/ADT
GermanARO9602 1.190 0.670 2.113 MFS 47
1.184 0.646 2.169 OS 43
SWOG8794 0.724 0.553 0.948 MFS 214
0.749 0.570 0.985 OS 207
TAX3501 0.755 0.202 2.819 MFS 9
0.973 0.137 6.911 OS 4
RT Dose
GETUG06 0.923 0.421 2.022 MFS 25
0.959 0.415 2.212 OS 23
MRCRT01 0.989 0.780 1.255 MFS 271
0.993 0.770 1.282 OS 236
RT+/-ADT
EORTC22863 0.496 0.379 0.650 MFS 221
0.596 0.447 0.796 OS 190
EORTC22961 0.602 0.475 0.763 MFS 285
0.704 0.542 0.915 OS 230
EORTC22991 0.767 0.568 1.036 MFS 173
0.826 0.600 1.137 OS 152
French study 0.827 0.568 1.204 MFS 110
0.861 0.586 1.265 OS 105
GICOR-DART01/05 0.640 0.311 1.318 MFS 31
0.780 0.328 1.850 OS 21
ICORG9701 1.576 1.077 2.307 MFS 110
1.578 1.070 2.327 OS 106
NCIC/MRC-PR3 0.720 0.603 0.860 MFS 495
0.741 0.617 0.890 OS 465
RTOG9202 0.865 0.774 0.968 MFS 1226
0.914 0.817 1.023 OS 1209
RTOG9408 0.853 0.743 0.979 MFS 807
0.855 0.742 0.986 OS 763
RTOG9413 1.038 0.914 1.179 MFS 953
0.991 0.871 1.127 OS 924
TROG9601_3m 0.856 0.646 1.134 MFS 205
0.874 0.648 1.179 OS 183
TROG9601_6m 0.559 0.417 0.751 MFS 181
0.593 0.432 0.815 OS 156
MRCPR04 1.309 0.988 1.735 MFS 195
1.033 0.761 1.404 OS 166
NP+/-Other
MRCPR04 0.794 0.531 1.186 MFS 96
0.783 0.518 1.183 OS 91
Favors Treatment
Favors Control
0.01
0.1
1
10

(D)TTM and DSS

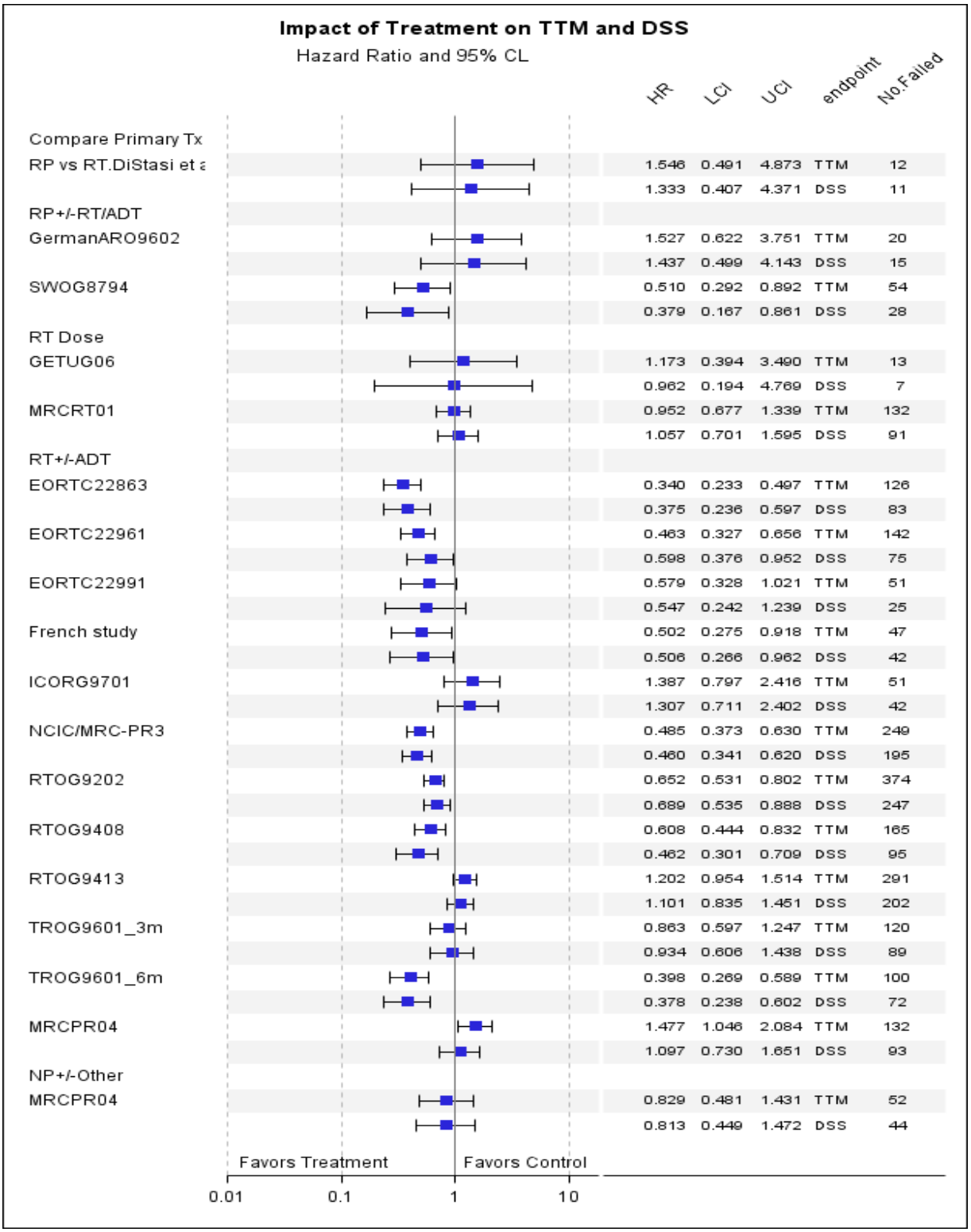

Impact of Treatment on TTM and DSS
Hazard Ratio and 95% CL
HR LCI UCI endpoint No.Failed
Compare Primary Tx
RP vs RT.DiStasi et a 1.546 0.491 4.873 TTM 12
1.333 0.407 4.371 DSS 11
RP+/-RT/ADT
GermanARO9602 1.527 0.622 3.751 TTM 20
1.437 0.499 4.143 DSS 15
SWOG8794 0.510 0.292 0.892 TTM 54
0.379 0.167 0.861 DSS 28
RT Dose
GETUG06 1.173 0.394 3.490 TTM 13
0.962 0.194 4.769 DSS 7
MRCRT01 0.952 0.677 1.339 TTM 132
1.057 0.701 1.595 DSS 91
RT+/-ADT
EORTC22863 0.340 0.233 0.497 TTM 126
0.375 0.236 0.597 DSS 83
EORTC22961 0.463 0.327 0.656 TTM 142
0.598 0.376 0.952 DSS 75
EORTC22991 0.579 0.328 1.021 TTM 51
0.547 0.242 1.239 DSS 25
French study 0.502 0.275 0.918 TTM 47
0.506 0.266 0.962 DSS 42
ICORG9701 1.387 0.797 2.416 TTM 51
1.307 0.711 2.402 DSS 42
NCIC/MRC-PR3 0.485 0.373 0.630 TTM 249
0.460 0.341 0.620 DSS 195
RTOG9202 0.652 0.531 0.802 TTM 374
0.689 0.535 0.888 DSS 247
RTOG9408 0.608 0.444 0.832 TTM 165
0.462 0.301 0.709 DSS 95
RTOG9413 1.202 0.954 1.514 TTM 291
1.101 0.835 1.451 DSS 202
TROG9601_3m 0.863 0.597 1.247 TTM 120
0.934 0.606 1.438 DSS 89
TROG9601_6m 0.398 0.269 0.589 TTM 100
0.378 0.238 0.602 DSS 72
MRCPR04 1.477 1.046 2.084 TTM 132
1.097 0.730 1.651 DSS 93
NP+/-Other
MRCPR04 0.829 0.481 1.431 TTM 52
0.813 0.449 1.472 DSS 44
Favors Treatment
Favors Control
0.01
0.1
1
10

Table S4 Subgroup analysis of correlation between the surrogates and true endpoints at the patient level

	Correlation between DFS and OS		Correlation between TDR and DSS		Correlation between MFS and OS		Correlation between TTM and DSS	
	No. of patients (trials)	Kendall's Tau (95%CI)	No. of patients (trials)	Kendall's Tau (95%CI)	No. of patients (trials)	Kendall's Tau (95%CI)	No. of patients (trials)	Kendall's Tau (95%CI)
All patients	21,140 (31)	0.85 (0.85-0.86)	20,496(28)	0.68 (0.67-0.69)	12,712(21)	0.91 (0.91-0.91)	12,068(18)	0.91 (0.91-0.92)
By type of therapy								
RP based	5,518 (7)	0.91(0.90-0.92)	5,290(6)	0.75 (0.73-0.78)				
RT based*	13,186(21)	0.84(0.83-0.84)	12,770(19)	0.66 (0.65- 0.67)	11,599(17)	0.91 (0.90-0.91)	11,183(15)	0.92 (0.91-0.92)
No local therapy	2,436(3)	0.86(0.85-0.87)	2,436(3)	0.69 (0.67-0.71)				
Within RT-based trials								
No use of ADT**	4,131(12)	0.76(0.75-0.78)	4,131(12)	0.64(0.61-0.66)	2,544(8)	0.91 (0.90- 0.92)	2,544(8)	0.90(0.89-0.92)
Use of ADT	9,055(15)	0.87(0.86-0.87)	8,639(13)	0.67(0.65-0.69)	9,055(15)	0.91 (0.91 0.91)	8,639(13)	0.92(0.92-0.93)
Duration of ADT								
≤6months	5,674(11)	0.84(0.83-0.85)	5,434(9)	0.63 (0.60-0.66)	5,674(11)	0.90 (0.89-0.90)	5,434(9)	0.91(0.90-0.92)
>6months	3,381(8)	0.90(0.89-0.91)	3,205(7)	0.75 (0.73-0.77)	3,381(8)	0.93 (0.93-0.94)	3,205(7)	0.94(0.94-0.95)
High risk patients only	13,324(31)	0.81(0.80-0.82)	12,824(28)	0.69(0.68- 0.71)	8,394(21)	0.88 (0.88-0.89)	7,894(18)	0.90(0.89-0.91)

OS: Overall survival, DSS: Disease specific survival, DFS: Disease free survival, TDR: Time to disease recurrence, MFS: Metastasis free survival, TTM: Time to metastasis
RP: Radical prostatectomy, RT: Radiation therapy, ADT: Androgen deprivation therapy

Table S5 Subgroup analysis of correlation between 8-year OS/DSS rate and 5-year surrogate rate

	Correlation between 5-year DFS and 8-year OS (by trial and arm)		Correlation between 5-year TDR and 8-year DSS (by trial and arm)		Correlation between 5-year MFS and 8-year OS (by trial and arm)		Correlation between 5-year TTM and 8-year DSS (by trial and arm)	
	No. of units	R-squared (95%CI)	No. of units	R-squared (95%CI)	No. of units	R-squared (95%CI)	No. of units	R-squared (95%CI)
All patients	56	0.86 (0.78- 0.90)	56	0.80 (0.70-0.85)	36*	0.83 (0.71-0.88)	36*	0.86 (0.75-0.90)
By type of therapy								
RP based	13	0.86 (0.58-0.92)	13	0.81 (0.46-0.89)	-	-	-	-
RT based*	37	0.68 (0.48-0.78)	37	0.71 (0.52-0.80)	29	0.77 (0.57-0.84)	29	0.87 (0.74-0.91)
High risk patients only	56	0.85 (0.77- 0.89)	56	0.84 (0.75-0.88)	36	0.70 (0.50-0.79)	36	0.70 (0.50-0.79)

OS: Overall survival, DSS: Disease specific survival, DFS: Disease free survival, TDR: Time to disease recurrence, MFS: Metastasis free survival, TTM: Time to metastasis
RP: Radical prostatectomy, RT: Radiation therapy, CI: Confidence interval

Table S6 Subgroup analysis of correlation between treatment effect on OS/DSS and treatment effect on surrogate endpoints

	Correlation between Log-HR (OS) vs Log-HR (DFS) (by trial)		Correlation between Log-HR (DSS) vs Log-HR (TDR) (by trial)		Correlation between Log-HR (OS) vs Log-HR (MFS) (by trial)		Correlation between Log-HR (DSS) vs Log-HR (TTM) (by trial)	
	No. of units	R-squared (95%CI)	No. of units	R-squared (95%CI)	No. of units	R-squared (95%CI)	No. of units	R-squared (95%CI)
All patients	31	0.73(0.53-0.82)	28	0.63(0.36-0.75)	21	0.92(0.81-0.95)	18	0.89 (0.72-0.93)
By type of therapy								
RP based	7	0.87(0.31-0.93)	6	0.79(0.04-0.89)				
RT based*	21	0.75(0.48-0.84)	19	0.63(0.27-0.77)	17	0.92(0.78-0.95)	15	0.89 (0.70-0.94)
High risk patients only	31	0.78(0.59-0.85)	28	0.82(0.66-0.88)	21	0.92(0.81-0.95)	18	0.92 (0.79-0.95)

OS: Overall survival, DSS: Disease specific survival, DFS: Disease free survival, TDR: Time to disease recurrence, MFS: Metastasis free survival, TTM: Time to metastasis
RP: Radical prostatectomy, RT: Radiation therapy, HR: Hazard ratio, CI: Confidence interval

Table S7 Correlation between endpoints at the trial level from competing risk models. Event rates and hazard ratios were estimated from competing risk models where non-prostate cancer death was considered as a competing risk

		Cumulative incidence of 8-year PCSM versus 5-year TDR/TTM (by trial and arm)		Treatment effect SHR(PCSM) versus SHR(TDR/TTM) (by trial)		
True endpoint	Surrogate endpoint	No. of units	R-squared, 95% CI	No. of units	R-squared, 95% CI	Regression equation
DSS	TDR	56*	0.79 (0.68-0.84)	28**	0.63 (0.36-0.75)	$Log(SHR)_{PCSM} = 0.025 + 0.799 \times Log(SHR)_{TDR}$
DSS	TTM	36*	0.86 (0.75-0.90)	18**	0.89 (0.73-0.93)	$Log(SHR)_{PCSM} = 0\text{-}.063 + 0.883 \times Log(SHR)_{TTM}$

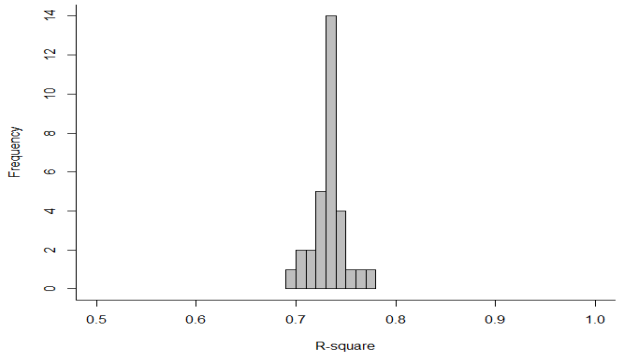
*Excluding 3 studies with median follow-up less than 6 years; 28 trials with a total of 56 arms were included for DFS/TDR analysis and 18 trials with a total of 36 arms were included for MFS/TTM analysis.

**Excluded 3 trials with number of prostate cancer death less than 3.

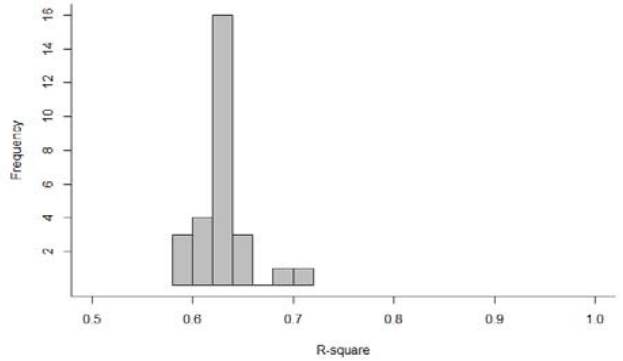
PCSM: Prostate cancer specific mortality, TDR: Time to disease recurrence, TTM: Time to metastasis, SHR: Sub-distribution hazard ratio, CI: Confidence interval

Figure S3 Leave-one-out-cross validation: R-squared between HR(OS/DSS) and HR(surrogates)

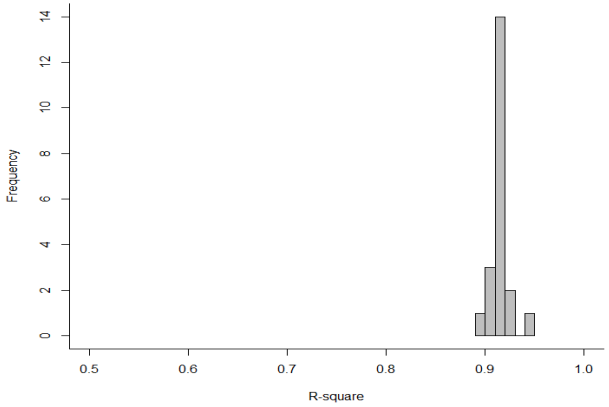
(A) R-squared between Log-HR(OS) and log-HR(DFS)
Median (range): 0.73(0.69-0.77)



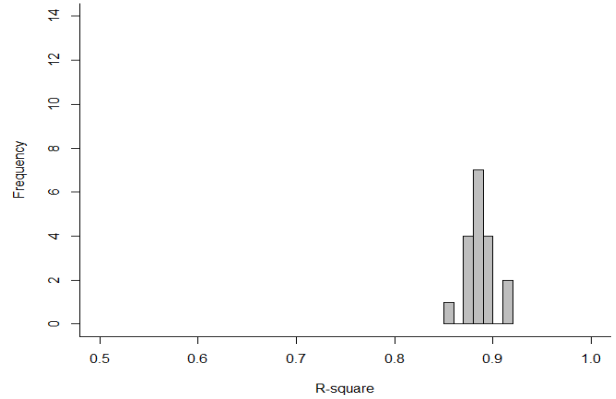
(B) R-squared between Log-HR(DSS) and log-HR(TDR)
Median (range): 0.63(0.59,0.70)



(C) R-squared between Log-HR(OS) and log-HR(MFS)
Median (range): 0.92 (0.90, 0.94)



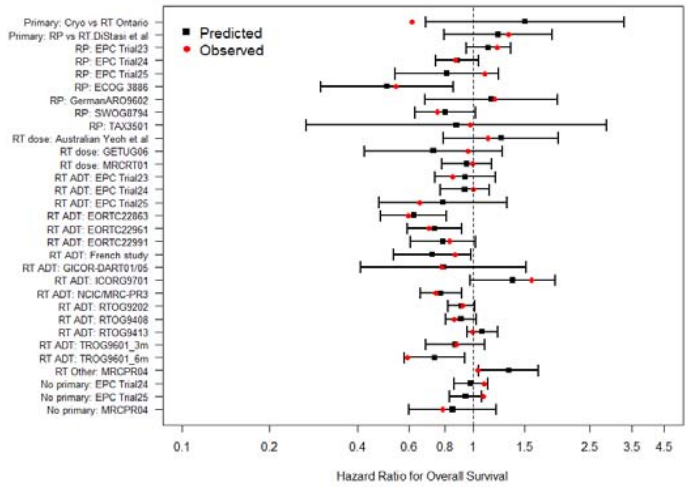
(D) R-squared between Log-HR(DSS) and log-HR(TTM)
Median (range): 0.89 (0.85, 0.92)



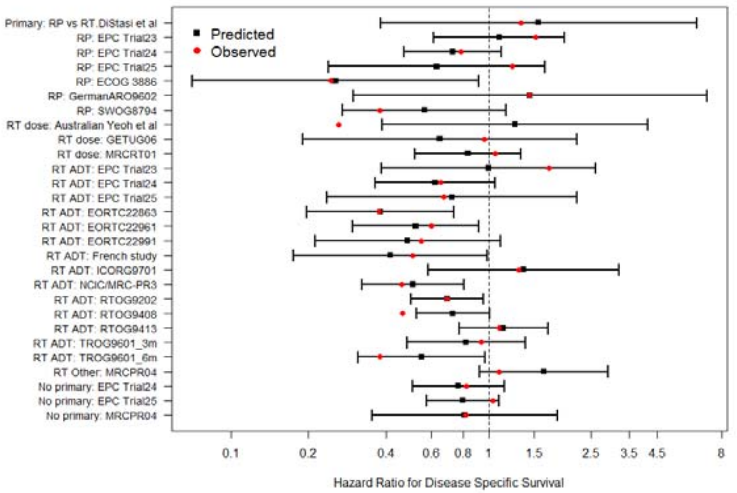
OS: Overall survival, DSS: Disease specific survival, DFS: Disease free survival, TDR: Time to disease recurrence, MFS: Metastasis free survival, TTM: Time to metastasis
HR: Hazard ratio

Figure S4 Leave-one-out-cross-validation: observed versus predicted treatment effects on OS/DSS. Black squares correspond to predicted hazard ratios (HR) on OS or DSS using the observed HR on surrogates of that particular trial, based on the regression model built on all the other (i.e. the remaining n-1) trials; red circles correspond to the observed HR on OS or DSS of that particular trial; horizontal lines correspond to 95% prediction intervals.

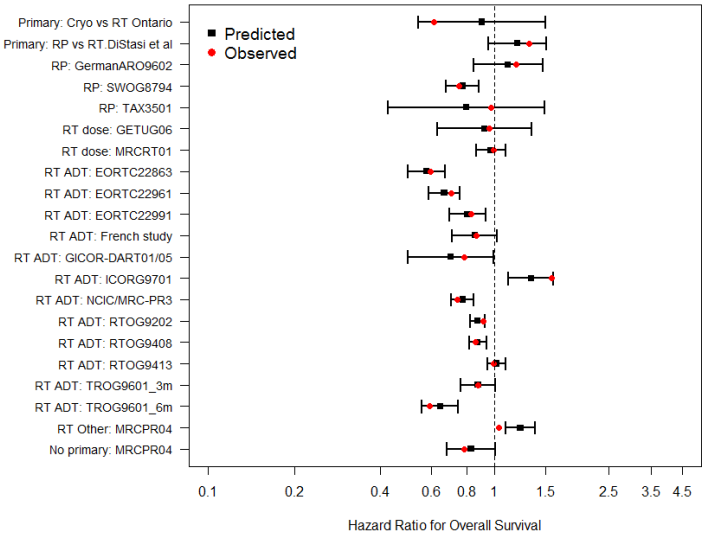
(A) Prediction for HR(OS) based on HR(DFS)



(B) Prediction for HR(DSS) based on HR(TDR)



(C) Prediction for HR(OS) based on HR(MFS)



(D) Prediction for HR(DSS) based on HR(TTM)

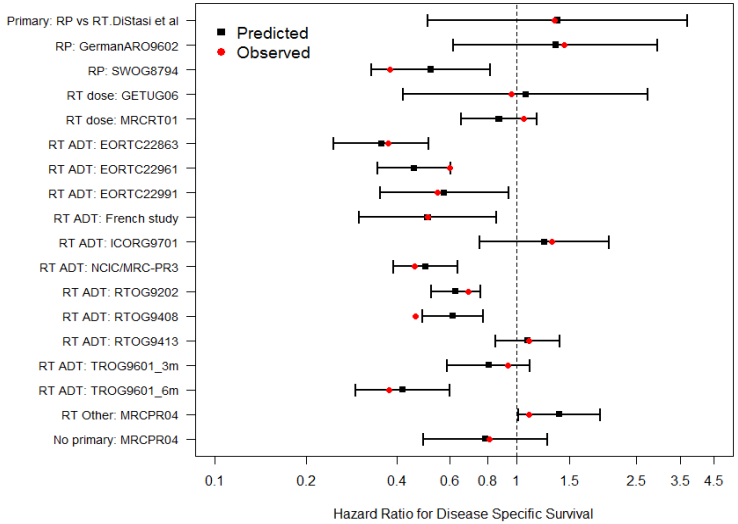


Table S8 Definition of Local progression and frequency of bone scan or image by study

Study	Local progression definition	Frequency of bone scan or image
Trials with both RP, RT or other		
EPC Trial23/24/25	Local progression was not defined but was included as part of objective (clinical) progression. Any of the following will be sufficient for clinical progression: (A)objective progression by bone scan, CT, MRI, biopsy etc (B)local or symptomatic progression: (i)ureteric obstruction either by primary tumor or pelvic nodal disease (ii)lymphedema of lower extremities due to pelvic model involvement. (iii)recurrent vesical obstruction, bleeding or pain due to growth of primary tumor	Not included for MFS analysis
Cryo vs RT Ontario	Clinical local recurrence is clinical evidence based on histological evidence.	Bone scan as clinically indicated
RP vs RT.DiStasi et al	Not reported	Not reported
RP-based trials		
ECOG3886	Not reported	Not included for MFS analysis
GermanARO9602	Not reported	Not reported
SWOG8794	biopsy-proven local recurrence	As clinically indicated
TAX3501	Not reported	Bone scans and CT scans were repeated once a year until disease progression and then every 6 months and or as clinically indicated.
RT-based: comparing RT doses		
Australian Study Yeoh et al	Local progression was not defined but was included as part of clinical relapse. The sites of clinical relapse were local, pelvic nodal, and bony metastatic based on histopathologic and radiologic (bone and CT abdominal scan) reassessment.	Not included for MFS analysis
GETUG06	Local relapse was based on digital examination, only four patients with positive biopsy findings	Not reported

Study	Local progression definition	Frequency of bone scan or image
MRCRT01	Local control was defined as time to clinically assessed failure; failure proven only by biopsy was excluded from this endpoint because only 304 of 843 (36%) patients consented to research biopsies planned at 2 years.	Bone scans, CT, or MRI were done as clinically indicated
RT-based: comparing duration of ADT		
EORTC22863	Local failure was defined as an increase of more than 50 percent in the product of the two maximal perpendicular diameters of the primary lesion as measured digitally, by CT or transabdominal ultrasonography; in case of doubt, biopsy was highly recommended. Local progression was defined as the recurrence of a palpable tumor after initial regression.	Chest X-ray, bone scan and C.T. scan of liver, retroperitoneum, abdomen and pelvis are performed annually.
EORTC22961	Local progression is assessed by the following symptoms :(A)Palpable enlargement of an existing abnormality or regrowth of a previously regressed prostate gland must be considered as a disease progression or recurrence when there is a 25% or greater increase in the product of the two largest diameters of the prostate, and must be documented by a positive biopsy to be considered as failure or relapse. (B) The development of an obstructed ureter constitutes evidence of progression. (C)Urethral obstruction or bleeding necessitating a trans-urethral resection constitutes evidence of progression only if the resected tissues demonstrate viable malignancy.	Chest X-ray, Technetium bone scan, CT scan of pelvis and abdomen are not mandatory every year, but are required, should there be a clinical and/or biochemical (PSA) suspicion of progression. CT scan or MRI should be performed when the interpretation of bone scan is difficult.

Study	Local progression definition	Frequency of bone scan or image
EORTC22991	Local progression is assessed by the following symptoms: - Palpable enlargement of an existing abnormality or regrowth of a previously regressed prostate lobe must be considered as a disease progression or recurrence, if there is a 25% or more increase in the size of the existing abnormality or of the involved prostate lobe, and documented by a positive biopsy. - Urethral obstruction or bleeding necessitating a trans-urethral resection constitutes evidence of progression only if the resected tissues demonstrate viable malignancy	Imaging studies (Bone Scan,CT-of abdomen +pelvis, MRI or Chest X-ray) will be done in case of suspicion of clinical and/or biochemical progression.
French study Mottet et al	Localregional progression was defined as >50% increase in prostate volume compared with the lowest value by ultrasound, the appearance of a new palpable prostate lesion in the event of previous complete clinical normalization, and identification of new regional lymph nodes by CT scan.	Ultrasound was recommended at 6 month, 1, 3, and 5 yr. CT and bone scans were systematically performed in case of clinical or biologic progression
GICOR-DART01/05	Not reported	Imaging (abdominal-pelvic CT and bone scan) was repeated in cases in which clinical or biochemical progression was suspected.
ICORG9701	Not reported	Bone scan Annually
NCIC/MRC-PR3	Local progression was defined as either ureteral obstruction or progressive disease accompanied by a biopsy sample showing tumor.	Not reported
RTOG9202 & RTOG9408	The time to local progression will be measured from the date of first treatment to the date of documented local progression as determined by clinical exam. (Note, because an endpoint in this study is tumor clearance and local control, a biopsy of the prostate will be obtained at 24 months following completion of radiation therapy.)	Bone scan as indicated. A bone scan will be performed on any patient who presents with complaints of bone pain that cannot be attributed to any inter-current disease. Discretionary plain films may be needed to evaluate lesions seen on bone scan to confirm the diagnosis of metastatic disease.

Study	Local progression definition	Frequency of bone scan or image

RTOG9413	Date of randomization to the date of local progression defined as any of the below:
	(A)Tumor progression
	(B)Positive repeat biopsy ≥ 2 years after treatment
	(C)If tumor never cleared or only had partial response then
	If tumor regrowth > 50% then
	o If < 2 years from randomization then failure date = date of tumor size measurement
	o If ≥ 2 years from randomization then failure date = day 1 (persistence)
	If tumor regrowth ≤ 50% then
	o If ≥ 2 years from randomization then failure date = day 1 (persistence)
		As clinically indicated, e.g. rising psa or bone pain. A bone scan will be performed on any patient who presents with complaints of bone pain that can not be attributed to any intercurrent disease. Discretionary plain films may be needed to evaluate lesions seen on bone scan to confirm the diagnosis of metastatic disease.

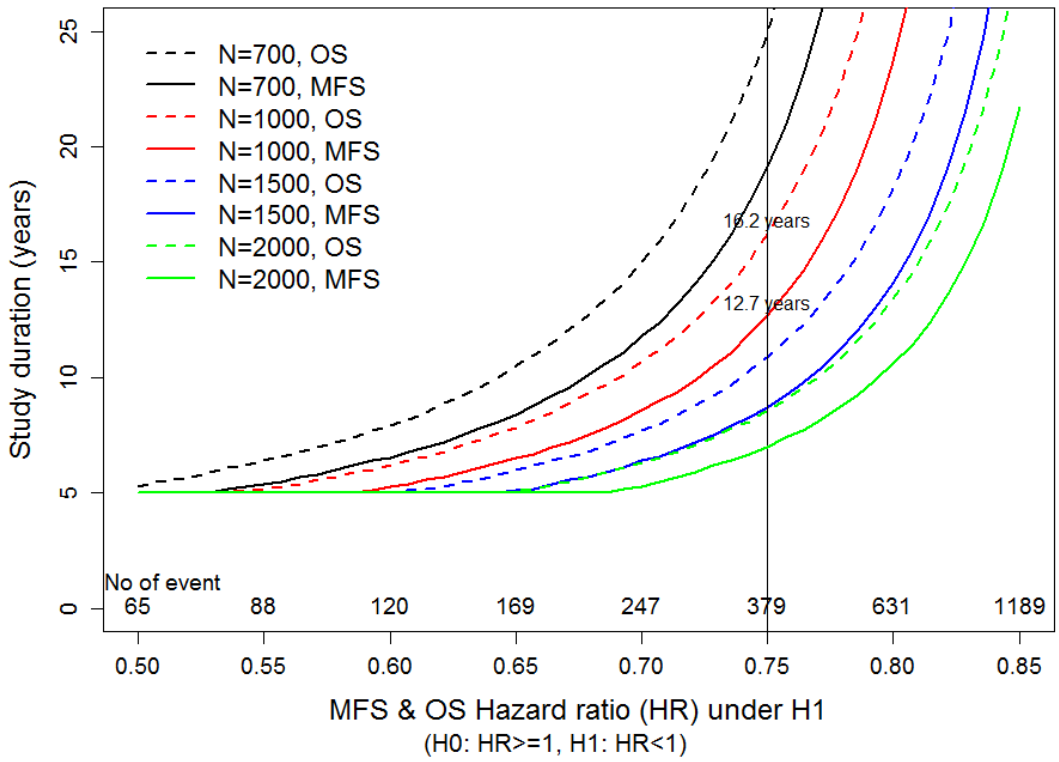
TROG9601	Local progression is defined as occurring at the time of first evidence of palpable malignant induration or to confirmatory biopsy or trans-urethral resection specimen histopathology if the procedure has occurred at least two years after radiotherapy.	Investigations—including biopsy, CT scan, chest radiograph, and isotope bone scan—were mandated if symptoms suggested a need, or if PSA reached 20 µg/L without signs of recurrence.
RT-based trials: other		
MRCPR04	Not reported	Not reported

Supplementary Material: Study Designs Using MFS and OS Endpoint

Design 1: Design a trial using MFS or OS endpoint, assuming that MFS and OS are independent endpoint

Design assumptions:
Randomization: 1:1
5-years MFS rate: 0.79 (hazard=0.04714 under exponential distribution)
5-year OS rate: 0.84 (hazard=0.03487 under exponential distribution)
Accrual period: 5 years
Type I error: 0.025 (one-sided) or 0.05 (two-sided)
Type II error: 0.20
Null Hypothesis (H0): HR(MFS)≥1 , HR(OS)≥1
Alternative hypothesis (H1): HR(MFS) < 1, HR(OS) < 1,

Figure S5 Total study duration required to observe the number of events under H1 according to various hazard ratios (H1) and sample sizes by using MFS or OS endpoint



From the ICECaP data, the observed 5-year MFS was 0.79 versus 5-year OS 0.84. Historically trials have been designed with a proposed treatment effect of HR(OS) 0.70 to 0.75, which have a study duration of 10.7 to 16.2 years with 1000 patients enrolled over 5-years. If we assume the treatment effect on OS is the same on MFS, i.e. HR(MFS) 0.75 the study duration would be 12.7 years.

Design 2: Design a trial using MFS endpoint, with validation that MFS is correlated with OS and applying the surrogate threshold effect (STE) based on ICECaP data (Figure 5 in main paper).

Design assumptions:
Randomization: 1:1
5-years MFS rate: 0.79 (hazard=0.04714 under exponential distribution)
5-year OS rate: 0.84 (hazard=0.03487 under exponential distribution)
Accrual period: 5 years
Type I error: 0.025 (one-sided) or 0.05 (two-sided)
Type II error: 0.20
Null Hypothesis (H0): HR(MFS**)≥0.88** , HR(OS)≥1
Alternative hypothesis (H1): HR(MFS)**< 0.88**, HR(OS) < 1,

From the weighted linear regression (WLR) analyses (Table 1 in main paper), there is not a 1:1 relationship between treatment effect on MFS and OS. Moreover, the STE on OS was a HR(MFS) of 0.88, which implies that a future trial would require an upper limit of the confidence interval for the estimated HR(MFS) to fall below the STE in order to predict a significant non-zero effect on OS. This is equivalent to planning a trial to test the hypothesis of HR(MFS) greater than, or equal to, the STE (0.88).

From the WLR, we can calculate the predicted HR(OS) and prediction intervals based on HR(MFS). We compare the study designs using HR(MFS) (test STE: H0: HR(MFS) ≥0.88, H1: HR(MFS)<0.88 to account for the uncertainty in the use of the surrogate) versus the designs using the predicted HR(OS) (test: H0: HR(OS) ≥1, H1: HR(OS)<1).

Table S9 The required number of events for the study designs using HR(MFS) and test the STE (0.88) versus designs using the associated predicted HR(OS) from WLR

Design using HR(MFS): H0: HR(MFS)≥0.88, H1: HR(MFS)<0.88						
HR(MFS) (under H1)	0.50	0.55	0.60	0.65	0.70	0.75
Required MFS events (under H1)	98	142	214	342	600	1229
Design using predicted HR(OS): H0: HR(OS)≥1, H1: HR(OS)<1						
Predicted HR(OS) & prediction interval*	0.59 (0.48-0.71)	0.63 (0.54-0.74)	0.67 (0.59-0.77)	0.71 (0.64-0.79)	0.75 (0.69-0.82)	0.79 (0.75-0.84)
Required OS events based on predicted HR(OS)	113	147	196	268	379	565

*Based on the WLR equation $Log(HR)_{OS}$= -0.021 + 0.740 x $Log(HR)_{MFS.}$ Prediction intervals were constructed with the weights equalling to the inverse variance of $Log(HR)_{MFS}$ (i.e. = No. of MFS events/4).

FigureS6 Total study duration required in the study designs using MFS hazard ratios and test STE (solid lines) or using predicted OS hazard ratios from WLR (dashed lines)

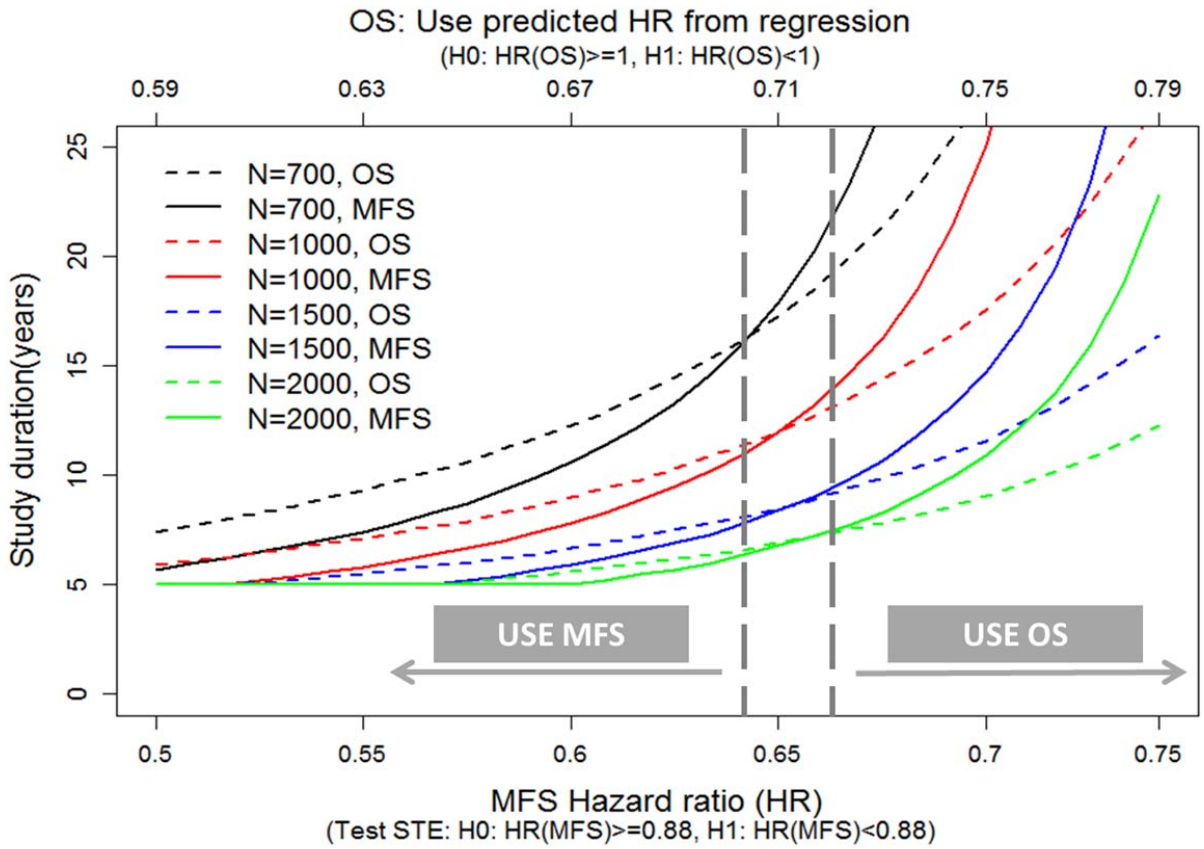


Figure S5 illustrates various situations and shows that use of MFS may reduce study duration even when the STE is duly taken into account. Specifically, MFS would be the preferred primary endpoint for HR(OS) lower than 0.7, while OS would be the preferred primary endpoint for HR(OS) greater than 0.72 (vertical gray dashed lines). For example, a trial with 1,000 patients designed to detect a treatment effect of HR(MFS) of 0.6 would have a total study duration of 7.7 years. The predicted HR(OS) is 0.67 and a trial designed to detect this effect would have a total study duration of 8.8 years. In short, use of MFS will allow an expeditious evaluation of a new therapy if it has a meaningful treatment effect on MFS.

Table S10 Kaplan Meier estimate of endpoints based on ICECaP data

(A)MFS analysis dataset (n=12,712 from 19 trials)

		All patients			High risk patients only		
Endpoint	**year**	**Event-Free rate**	**LCL**	**UCL**	**Event-Free rate**	**LCL**	**UCL**
OS	5	0.84	0.83	0.84	0.81	0.80	0.82
OS	8	0.70	0.69	0.70	0.66	0.65	0.67
OS	10	0.59	0.58	0.60	0.56	0.54	0.57
DSS	5	0.95	0.94	0.95	0.93	0.92	0.94
DSS	8	0.89	0.89	0.90	0.86	0.85	0.87
DSS	10	0.85	0.84	0.86	0.81	0.80	0.82
MFS	5	0.79	0.79	0.80	0.76	0.75	0.77
MFS	8	0.65	0.64	0.66	0.61	0.60	0.62
MFS	10	0.56	0.55	0.57	0.51	0.50	0.53
TTM	5	0.90	0.89	0.90	0.86	0.86	0.87
TTM	8	0.83	0.83	0.84	0.79	0.78	0.80
TTM	10	0.80	0.79	0.80	0.74	0.73	0.75

(B) DFS analysis dataset (N=21,140 from 24 trials)

		All patients			High risk patients only		
Endpoint	year	Event-Free rate	LCL	UCL	Event-Free rate	LCL	UCL
OS	5	0.85	0.84	0.85	0.83	0.82	0.84
OS	8	0.72	0.72	0.73	0.69	0.68	0.70
OS	10	0.63	0.63	0.64	0.60	0.59	0.61
DSS	5	0.95	0.95	0.96	0.93	0.93	0.94
DSS	8	0.91	0.90	0.91	0.87	0.86	0.88
DSS	10	0.87	0.86	0.87	0.82	0.82	0.83
DFS	5	0.76	0.75	0.76	0.72	0.71	0.73
DFS	8	0.63	0.62	0.64	0.59	0.58	0.60
DFS	10	0.55	0.54	0.55	0.51	0.50	0.52
TDR	5	0.85	0.84	0.85	0.81	0.80	0.81
TDR	8	0.78	0.77	0.79	0.73	0.72	0.74
TDR	10	0.74	0.74	0.75	0.69	0.68	0.70