Artificial Intelligence Predictive Model for Hormone Therapy Use in Prostate Cancer

Daniel E. Spratt, M.D.[1], **Siyi Tang, Ph.D.**[2,3], **Yilun Sun, Ph.D.**[1,4], **Huei-Chung Huang, M.A.**[3], **Emmalyn Chen, Ph.D.**[3], **Osama Mohamad, M.D., Ph.D.**[5], **Andrew J. Armstrong, M.D.**[6], **Jonathan D. Tward, M.D., Ph.D.**[7], **Paul L. Nguyen, M.D.**[8], **Joshua M. Lang, M.D., M.Sc.**[9], **Jingbin Zhang, M.Sc.**[3], **Akinori Mitani, M.D., Ph.D.**[3], **Jeffry P. Simko, M.D., Ph.D.**[5], **Sandy DeVries, M.A.**[10], **Douwe van der Wal, M.Sc.**[3], **Hans Pinckaers, M.D., M.Sc.**[3], **Jedidiah M. Monson, M.D.**[11], **Holly A. Campbell, M.D.**[12], **James Wallace, M.D.**[13], **Michelle J. Ferguson, M.D.**[14], **Jean-Paul Bahary, M.D.**[15], **Edward M. Schaeffer, M.D., Ph.D.**[16], **Howard M. Sandler, M.D.**[17], **Phuoc T. Tran, M.D., Ph.D.**[18], **Joseph P. Rodgers, M.S.**[19,20], **Andre Esteva, Ph.D.**[3], **Rikiya Yamashita, M.D., Ph.D.**[3], **Felix Y. Feng, M.D.**[5] on behalf of NRG Prostate Cancer AI Consortium*

[1]Department of Radiation Oncology, University Hospitals Seidman Cancer Center, Case Western Reserve University, Cleveland

[2]Department of Electrical Engineering, Stanford University, Stanford, CA

[3]Artera, Inc., Los Altos, CA

[4]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland

[5]Department of Radiation Oncology, University of California, San Francisco, San Francisco

[6]Duke Cancer Institute Center for Prostate and Urologic Cancer, Division of Medical Oncology, Department of Medicine, Duke University, Durham, NC

[7]Department of Radiation Oncology, Huntsman Cancer Institute, University of Utah, Salt Lake City

[8]Department of Radiation Oncology, Dana-Farber/Brigham Cancer Center, Boston

[9]Division of Hematology/Medical Oncology, University of Wisconsin, Madison, WI

[10]NRG Oncology Biospecimen Bank, University of California, San Francisco, San Francisco

[11]Department of Radiation Oncology, Saint Agnes Medical Center, Fresno, CA

[12]Department of Radiation Oncology, Saint John Regional Hospital, Saint John, NB, Canada

[13]University of Chicago Medicine Medical Group, Chicago

[14]Department of Radiation Oncology, Allan Blair Cancer Centre, Regina, SK, Canada

[15]Department of Radiation Oncology, Centre Hospitalier de l'Universite de Montreal, Montreal

Dr. Feng can be contacted at felix.feng@ucsf.edu or at Box 3110, Room 450, 1450 3rd Street, San Francisco, CA 94158.

Drs. Spratt, Tang, and Sun contributed equally as co-first authors. Drs. Yamashita and Feng contributed equally as co-senior authors.

ClinicalTrials.gov numbers NCT00767286, NCT00002597, NCT00769548, NCT00005044, and NCT00033631.)

[16]Department of Urology, Northwestern University Feinberg School of Medicine, Chicago

[17]Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles

[18]Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore

[19]Statistics and Data Management Center, NRG Oncology, Philadelphia

[20]Statistics and Data Management Center, American College of Radiology, Philadelphia

Abstract

BACKGROUND—Androgen deprivation therapy (ADT) with radiotherapy can benefit patients with localized prostate cancer. However, ADT can negatively impact quality of life, and there remain no validated predictive models to guide its use.

METHODS—We used digital pathology images from pretreatment prostate tissue and clinical data from 5727 patients enrolled in five phase 3 randomized trials, in which treatment was radiotherapy with or without ADT, as our data source to develop and validate an artificial intelligence (AI)–derived predictive patient-specific model that would determine which patients would develop the primary end point of distant metastasis. The model used baseline data to provide a binary output that a given patient will likely benefit from ADT or not. After the model was locked, validation was performed using data from NRG Oncology/Radiation Therapy Oncology Group (RTOG) 9408 (n=1594), a trial that randomly assigned men to radiotherapy plus or minus 4 months of ADT. Fine–Gray regression and restricted mean survival times were used to assess the interaction between treatment and the predictive model and within predictive model–positive, i.e., benefited from ADT, and –negative subgroup treatment effects.

RESULTS—Overall, in the NRG/RTOG 9408 validation cohort (14.9 years of median follow-up), ADT significantly improved time to distant metastasis. Of these enrolled patients, 543 (34%) were model positive, and ADT significantly reduced the risk of distant metastasis compared with radiotherapy alone. Of 1051 patients who were model negative, ADT did not provide benefit.

CONCLUSIONS—Our AI-based predictive model was able to identify patients with a predominantly intermediate risk for prostate cancer likely to benefit from short-term ADT.

Introduction

Radiotherapy is a common form of treatment administered with curative intent for localized prostate cancer. Trials conducted since the 1980s consistently demonstrate an improvement in oncologic outcomes when androgen deprivation therapy (ADT) is added to radiotherapy.[1–5] However, ADT has well-documented toxicity, including hot flashes, declines in libido and erectile function, loss of muscle mass, increase in body fat, osteoporosis, and potentially deleterious effects on cardiac and brain health.[6]

Although consistent oncologic benefits of ADT have been demonstrated, the majority of men with localized prostate cancer treated with radiotherapy alone without ADT never develop distant metastasis.[1,5,7–10] Unfortunately, there remains no validated way to identify which men specifically derive benefit from ADT with radiotherapy; current guidelines recommend the use of ADT on the basis of prognostic National Comprehensive Cancer

Network (NCCN) risk groups or other methods of prognostication.[11] Gleason grading has modest prognostic ability, and a number of tissue-based gene expression, serum, and imaging biomarkers have also been developed to help determine which men may benefit from ADT. Although some of these markers have demonstrated prognostic value,[12] none have been shown to function as predictive biomarkers for ADT use with randomized trial validation. An easy and reliable method to guide the individualized use of ADT with radiotherapy for men with localized prostate cancer would be of value to such patients.

Digital pathology has been used for years as a method to archive, visualize, and share histopathology images.[13] More recently, there has been growing interest in leveraging artificial intelligence (AI) to assist in the diagnosis and grading of prostate cancer.[14–16] Fundamentally, these efforts restrict AI to predict human interpretable and defined features (i.e., Gleason score). In a recent study, a multimodal artificial intelligence (MMAI) system leveraging digital histopathology and clinical data from five NRG Oncology phase 3 clinical trials, termed the MMAI Prostate Prognostic Model, was used to develop and validate prognostic models that consistently outperformed NCCN risk groups to determine which men with localized prostate cancer would benefit from ADT.[17] In this study, we extend this approach by adapting the MMAI Prostate Prognostic Model to develop and test a predictive model on the basis of "deep learning" that has the potential to be used to identify which patients would benefit from ADT.

In this report, we used extant data from four NRG Oncology North American phase 3 randomized trials (i.e., NRG 9202, 9413, 9910, and 0126) with long-term follow-up data, including pathology images. Data from these trials were acquired and digitized, and they were used to train a predictive AI model for the identification of men with localized prostate cancer who were likely to derive a differential benefit from the addition of ADT to radiotherapy. This predictive model for differential benefit from ADT was then validated using data from NRG Oncology/Radiation Therapy Oncology Group (RTOG) 9408, a clinical trial that randomly assigned men to treatment with radiotherapy plus or minus 4 months of ADT; this trial consisted mostly of men with intermediate-risk prostate cancer, defined as a Gleason score of 7 or a Gleason score of 6 or less with a prostate-specific antigen (PSA) of 10 to 20 ng/ml or clinical stage T2b and not high risk (Supplementary Appendix).[1,7–10]

Methods

ANCILLARY PROJECT AND TRIAL DETAILS

NRG Oncology randomized phase 3 trials conducted in men with localized nonmetastatic prostate cancer that enrolled at least a subset of patients with intermediate risk for disease, included treatment with radiotherapy alone or with ADT, had long-term follow-up defined as a median follow-up greater than 8 years, and had stored histopathology slides in the NRG Oncology Biospecimen Bank were eligible for inclusion. Trials testing the use of chemotherapy were excluded. Data from five prospective phase 3 randomized trials (NRG/RTOG 9202, 9413, 9910, 0126, and 9408) were identified and used for the development and validation of a predictive model for the use of ADT in patients with localized prostate cancer.[1,7–10] NRG/RTOG 9408 was used as the validation cohort in this study because

it represents one of the largest phase 3 clinical trials evaluating patients who received radiotherapy with or without 4 months of ADT. All image data from the remaining trials were used for the image feature extraction model, and full image, clinical, and outcome data from NRG/RTOG 9910 and 0126 were used for downstream predictive model development.

Details of the eligibility criteria, including the case definitions for intermediate- and high-risk disease, for each trial and the development and validation cohorts can be found in Tables S1 and S2. Briefly, NRG/RTOG 9202 enrolled men with intermediate and high risk prostate cancer; patients were randomly assigned to radiotherapy with 4 versus 28 months of ADT. NRG/RTOG 9413 enrolled men with intermediate and high risk prostate cancer and was a 2×2 factorial trial with randomizations to 4 months of ADT sequencing and use of pelvic nodal radiotherapy. In NRG/RTOG 9910, men with intermediate risk prostate cancer were randomly assigned to radiotherapy with 16 weeks of ADT or with 36 weeks of ADT. In NRG/RTOG 0126, patients with intermediate risk prostate cancer were randomly assigned to lower versus higher doses of radiotherapy without ADT. In NRG/RTOG 9408, men with low, intermediate, or high risk prostate cancer were randomly assigned to radiotherapy with or without 4 months of ADT. Trials that included the use of ADT consisted of combined androgen blockade with a luteinizing hormone-releasing hormone (LHRH) agonist and an anti-androgen. Short-term ADT was defined as 4 months of ADT (and the 36 weeks of ADT in RTOG 9910 given no difference in outcomes), and long-term ADT was solely used in the experimental group of NRG/RTOG 9202 (28 months).

OBJECTIVE AND END POINTS

The primary objective was to develop and validate an AI-based predictive model that could identify a differential benefit from the addition of short-term ADT to radiotherapy in localized prostate cancer. The primary end point used in the model training and validation was the time to distant metastasis measured from the time of randomization until the development of distant metastasis or last follow-up. The secondary objective was to evaluate the predictive model on a secondary end point: prostate cancer–specific mortality (defined in the present study as death in the setting of distant metastasis). Metastasis-free survival (MFS; distant metastasis or death from any cause) and overall survival were evaluated as exploratory end points.

HISTOPATHOLOGY IMAGE ACQUISITION

Unannotated hematoxylin and eosin–stained histopathology slides in patients with localized prostate cancer from the NRG Oncology Biospecimen Bank were independently digitized without access to clinical outcomes data. The slides were digitized using a Leica Biosystems Aperio AT2 digital pathology scanner at a 20× magnification level.

IMAGE FEATURE EXTRACTION MODEL DEVELOPMENT

The first component of model development was image feature extraction, which was trained on images only to recognize defining tissue features and did not evaluate any clinical variables or outcomes. For each patient, the tissues across all available digital slides were divided into 256×256-pixel patches. A Resnet-50 feature extraction model was trained on image patches using self-supervised learning.[18] We used the Momentum Contrast Version

2 (MoVo-v2) training protocol without access to any clinical or outcome data.[19] Over 2.5 million tissue patches across the four trials (NRG/RTOG 9202, 9413, 9910, and 0126) were fed through the model 200 times to train this model.

DOWNSTREAM MULTIMODAL PREDICTIVE MODEL DEVELOPMENT

The second component of model development was downstream multimodal predictive model development, which evaluated the association between all features — clinical and image — with clinical outcomes and included patients from NRG/RTOG 9910 and 0126. Because the other two trials (NRG/RTOG 9202 and 9413) included predominantly men at high risk, these two were excluded from downstream predictive model development to ensure that the development set had a similar patient population as the target population for the predictive model (i.e., intermediate-risk prostate cancer). Both NRG/RTOG 9910 and 0126 were included in downstream multimodal predictive model development because each contributed to one treatment type of interest (radiotherapy plus short-term ADT vs. radiotherapy only, respectively) (Supplementary Appendix). Then, the model development cohort was further stratified by treatment type and randomly split into training (60%) and tuning (40%) sets for model training and hyperparameter tuning, respectively.[20,21] Clinical data, image data, and treatment types were used as inputs to a multimodal predictive model architecture (Fig. S1A). The treatment type was used only for model development; treatment type was not required for model score generation on the locked model. The image and clinical data were preprocessed as specified in the Supplementary Appendix.

The multimodal predictive model optimized the difference in the magnitude of ADT benefit, outputting a continuous score "delta" (Fig. S1A). The 67th percentile of the delta scores in the development set was selected as the cutoff threshold because it maximized the difference between predictive model subgroup treatment effects in the tuning set and would result in reasonably sized predictive model subgroups for clinical utility. Patients with a delta score greater than the cutoff were classified as predictive model positive, and those below the cutoff were classified as predictive model negative (Fig. S1B). Model development was performed using the Python programming language (Python Language Reference; version 3.8.12, Python Software Foundation). After the model was locked, it was provided to independent biostatisticians (H.-C.H. and J.Z.) to perform clinical validation of the model in NRG/RTOG 9408.

STATISTICAL ANALYSIS

The NRG/RTOG 9408 validation cohort characteristics by predictive model status (positive or negative) were reported and compared using the chi-square test or Fisher's exact test in the presence of low cell counts for categorical variables and Wilcoxon's rank-sum test for continuous variables. Time to event was analyzed using the cumulative incidence function; for distant metastasis and prostate cancer–specific mortality, death without the corresponding event was treated as a competing risk. Fine and Gray regression was also performed to estimate the subdistribution hazard ratio and 95% confidence interval (CI) for the short-term ADT treatment effect for distant metastasis and prostate cancer–specific mortality.[22] A test for predictive model–treatment interaction was performed to evaluate this predictive model. Treatment effects of the predictive model–positive and –negative

subgroups were similarly assessed as the overall validation cohort to measure the relative treatment effect between groups. Fifteen-year restricted mean survival times were reported to provide alternative estimates given that nonproportional hazards were observed.[1]

Exploratory subgroup analyses were performed where the primary analysis was reanalyzed within patients at NCCN low and intermediate risk. Because of stage and Gleason score migration, patients at low risk from NRG/RTOG 9408 are more similar to contemporary patients at intermediate risk and were included in the subgroup analyses. Statistical analyses were performed using R, version 3.5.1 (R Foundation for Statistical Computing, Vienna). No multiplicity adjustments for the secondary and exploratory end points were defined. Therefore, only point estimates and 95% CIs are provided. The CIs have not been adjusted for multiple comparisons and should not be used to infer definitive treatment effects. Differences in percentages may not add up because of rounding.

Results

PATIENT AND MODEL CHARACTERISTICS

Of the 7752 eligible patients enrolled in the five phase 3 randomized trials, 6020 (77.7%) patients had available slides at the NRG Biospecimen Bank. Of these patients, 5727 (95.1%) had available pretreatment prostate slides. Pretreatment slides were not available for 285 patients, and 8 patients had insufficient tissue. Additionally, 39 patients with transurethral resection of the prostate samples were further excluded from the validation cohort (NRG/RTOG 9408). Details regarding the representativeness of the trial patients are provided in Table S3.[23]

The development cohort for the downstream predictive model for differential benefit from ADT had 2024 patients with a median follow-up of 10.6 years; 1050 (52%) patients received radiotherapy alone, and 974 (48%) patients received radiotherapy with short-term ADT (Tables S2 and S4). The median PSA was 9 ng/ml (interquartile range, 6 to 13), 87% had an intermediate risk of disease, and the median age was 71 years (interquartile range, 65 to 74). The final locked model was composed primarily of histopathology features (Gleason score and imaging features), contributing to more than 86% of model prediction (Fig. S2). Although histopathology features provide a large contribution, the MMAI architecture utilizes deep learning and also captures interaction effects, with the model benefiting from learning of all features.

The validation set (NRG/RTOG 9408) consisted of 1594 patients with a median follow-up of 14.9 years, with the groups reasonably balanced in size (radiotherapy alone, 806 patients; radiotherapy plus short-term ADT, 788 patients) (Fig. 1 and Table 1). The median PSA was 8 ng/ml (interquartile range, 6 to 12), 56% had an intermediate risk of disease, and the median age was 71 years (interquartile range, 66 to 74). To evaluate the representativeness of the overall trial cohort, baseline characteristics between trial groups, evaluable cohorts, and original eligible cohorts for the NRG/RTOG 9408 trial are outlined in Table 1. In the validation set, 543 patients (34%) were classified as predictive model positive (predicted to benefit most from short-term ADT), and 1051 patients (66%) were predictive model negative (predicted to derive lesser or no benefit from short-term ADT). Baseline

characteristics were generally well matched between patients who were predictive model positive and negative except for Gleason score; 24% of patients who were predictive model positive versus 30% of patients who were predictive model negative had a Gleason score of 7 (Table S5).

SHORT-TERM ADT PREDICTIVE MODEL

In the overall validation cohort, the short-term ADT group had a 15-year distant metastasis estimate of 5.9% (95% CI, 4.2 to 7.6%) compared with the 15-year distant metastasis estimate in the radiotherapy alone group of 9.8% (95% CI, 7.6 to 11.9%; subdistribution hazard ratio, 0.64; 95% CI, 0.45 to 0.90; P=0.01) (Fig. 2A). Applying the locked AI-derived model to the validation set, patients identified as predictive model positive with the addition of short-term ADT had a 15-year distant metastasis estimate of 4.0% (95% CI, 1.5 to 6.4%) compared with radiotherapy alone, with a 15-year distant metastasis estimate of 14.4% (95% CI, 10.0 to 18.8%; subdistribution hazard ratio, 0.34; 95% CI, 0.19 to 0.63; P<0.001) (Fig. 2A). In contrast, for the patients identified as predictive model negative, two treatment groups had 15-year distant metastasis estimates of 6.9% (95% CI, 4.6 to 9.2%) and 7.4% (95% CI, 5.0 to 9.7%), respectively (subdistribution hazard ratio, 0.92; 95% CI, 0.59 to 1.43; P=0.71) (Fig. 2A). The interactions between treatment and predictive model for time to distant metastasis are shown in Figure 3. The absolute benefit of short-term ADT, measured as the difference in distant metastasis between treatment groups at 15 years after randomization, was 10.5 percentage points (95% CI, 5.4 to 15.5%; i.e., 4.0 vs. 14.4% event estimates) (Figs. 2A and 3) in patients who were predictive model positive. In contrast, in patients with predictive model– negative disease, there was a 0.5–percentage point (95% CI, −2.8 to 3.7%; 6.9 vs. 7.4%) reduction in 15-year distant metastasis risk from the addition of ADT. Similarly, the short-term ADT benefit on distant metastasis measured by the restricted mean survival times at 15 years was 0.8 years (95% CI, 0.3 to 1.3) in patients who were predictive model positive and 0.1 years (95% CI, −0.1 to 0.4) in patients who were predictive model negative.

The secondary end point of prostate cancer–specific mortality was also assessed (Figs. 2B and 3). In the overall validation cohort, the short-term ADT group had a 15-year event estimate of 4.4% (95% CI, 2.8 to 5.9%), whereas the radiotherapy alone group had a 15-year event estimate of 8.6% (95% CI, 6.6 to 10.7%; subdistribution hazard ratio, 0.52; 95% CI, 0.35 to 0.78) (Fig. 2B). Patients who were predictive model positive had 15-year prostate cancer–specific mortality estimates of 2.6% (95% CI, 0.5 to 4.6%) if randomly assigned to additional short-term ADT and 12.7% (95% CI, 8.5 to 17.0%) if randomly assigned to radiotherapy only (subdistribution hazard ratio, 0.28; 95% CI, 0.14 to 0.57). In contrast, for patients who were predictive model negative, 15-year event estimates were 5.3% (95% CI, 3.2 to 7.4%) for additional ADT and 6.5% (95% CI, 4.3 to 8.7%) for radiotherapy alone (subdistribution hazard ratio, 0.74; 95% CI, 0.45 to 1.22) (Fig. 2B). Absolute differences in prostate cancer–specific mortality risks at 15 years were 10.2 percentage points (event estimates: 2.6 vs. 12.7%) versus 1.2 percentage points (event estimates, 5.3 vs. 6.5%) in predictive model–positive and –negative subgroups, respectively. The short-term ADT benefits on prostate cancer–specific mortality restricted mean survival times at 15 years were 0.7 years (95% CI, 0.3 to 1.1) in patients who were predictive model positive and

0.2 years (95% CI, −0.1 to 0.4) in patients who were predictive model negative (Fig. 3). On exploratory subset analysis, when restricting the analyses to solely patients at low and intermediate risk for disease, the results remained similar (Fig. S3).

We did not observe differential treatment benefits between predictive model subgroups on the exploratory end points MFS and overall survival (P interaction=0.31 and 0.23, respectively) (Fig. S4). The predictive model effects on distant metastasis and prostate cancer–specific mortality were evaluated within each treatment group (Table S6). For distant metastasis, within the radiotherapy alone group, the predictive model–positive versus – negative subgroup subdistribution hazard ratio was 1.93 (95% CI, 1.24 to 2.98), whereas within the radiotherapy plus short-term ADT group, the predictive model subdistribution hazard ratio was 0.72 (95% CI, 0.39 to 1.34); similar results were found for prostate cancer–specific mortality as well.

Discussion

The current standard of care for men with intermediate-risk (specifically, unfavorable intermediate-risk) localized prostate cancer treated with radiotherapy is the addition of short-term ADT. Despite the improvement in outcomes in all-comers, the majority of men will not develop distant metastasis with radiotherapy alone, and many will experience side effects from ADT. Unfortunately, there are no validated predictive models to guide ADT use or duration in these men. Herein, we report our results using novel deep learning methodology and leveraging image data from over 5000 patients in five phase 3 randomized trials with long-term follow-up to create and validate a predictive model to guide ADT use with radiotherapy in men with localized prostate cancer.

As a patient's prognosis worsens (i.e., going from NCCN low to high risk), the recommendations to add ADT to radiotherapy strengthen. This is despite evidence that NCCN risk groups are not predictive of ADT benefit.[5] To this point, we demonstrate that among patients with positive and negative AI model predictions, the baseline PSA, T stage, and NCCN risk group distribution were similar; there were small differences in the Gleason score. These results confirm that historical categorization of tumor aggressiveness alone is insufficient to determine which patients derive differential relative benefit from ADT.

A concern with any model is the possibility of overfitting and failure to validate. This cannot be overstated, and independent validation remains necessary to prove the performance of a model. In the specific case of predictive models, which aim to identify those patients who derive greater or lesser relative benefit, this almost always should be performed within the context of a randomized trial of the treatment of interest to avoid confounding and bias between groups. Herein, we intentionally selected NRG/RTOG 9408, because it remains the largest published trial of radiotherapy with or without short-term ADT with very long-term follow-up. Although there was clear benefit of ADT in unselected patients in this trial, the majority of patients enrolled had no demonstrable benefit. Our results indicate that over 60% of the patients at intermediate risk enrolled in NRG/RTOG 9408 did not derive benefit from ADT.

The primary end point of time to distant metastasis was specifically selected to train the short-term ADT predictive model. Other end points, such as biochemical recurrence, MFS, and overall survival, all have clinical relevance, but in the context of localized prostate cancer model development, they have notable limitations. ADT inhibits PSA production, and thus, ADT is expected to delay biochemical recurrence irrespective of subgroup. Furthermore, the majority of biochemical recurrence events do not result in metastasis or death.[24] Therefore, it is a suboptimal end point for model training to determine intrinsic tumor-specific benefit from ADT. MFS and overall survival are important end points for determining the net effect of a given therapy and are the gold standard for clinical trial design because they also capture death from competing causes. However, they are suboptimal end points for development of prostate cancer–specific predictive models for localized disease. This is because 78% of deaths in the validation cohort were not from prostate cancer, and only 12% of events in the MFS end point were from metastatic events. Thus, the strongest prediction models for MFS and overall survival would be driven by variables associated with death from nonprostate cancer causes (i.e., comorbid conditions). Importantly, despite the model being trained for distant metastasis, it showed a clear differential impact of ADT by predictive model status for prostate cancer–specific mortality, a cancer-driven end point.

As with any model, generalizability is critical. Concerns have been raised from AI models derived from a limited number of centers and in cohorts with limited diversity. Because of the limitations of the available data, we were unable to fully account for the potential confounding effect of factors impacting various aspects of health (e.g., socioeconomic status). Fortunately, NRG/RTOG enrolls patients from over 500 centers across primarily the United States and Canada from academic, community, and Veterans Affairs centers, and 20% of the 1594 patients in the validation cohort were Black; this is higher than the proportion of Black men (15.6%) given a diagnosis of localized prostate cancer in the United States.[25] This important real-world diversity strengthens the generalizability of our findings. However, this study was underpowered to further assess the predictive performance of the model for Black men, and future studies are needed for evaluation.

The study was limited. Similar to other prognostic and predictive models in active clinical use, our short-term ADT predictive model was not developed and validated as part of a de novo prospective model dedicated trial. This approach is supported by Simon et al.,[26] and use of a randomized trial of radiotherapy with or without ADT strengthens the credibility and level of evidence of our work. During the era of conduct and follow-up of this trial, there was effectively no use of advanced molecular imaging. Grade migration because of changes in the Gleason grading system may also have impacted patient stratification into NCCN risk groups. However, any potential biases introduced by this are likely random and impact both trial groups, and the raw histopathology imagery would not be impacted by changes in definitions of grading over time. Information on other prognostic clinicopathologic variables, such as the percentage of Gleason pattern 4 or the percentage of positive biopsy cores, was not available. Thus, alternative risk classification schemas for exploratory analyses were not performed.[27,28]

Author Manuscript

Conclusions

We have developed and independently validated in a completed phase 3 randomized trial an AI-based predictive model to guide ADT use with radiotherapy in localized prostate cancer using a novel multimodal digital pathology AI-derived platform; details on accessing this predictive model are in the Supplementary Appendix. Using this predictive model, we showed from the trial data that the majority of patients at intermediate risk did not benefit from ADT treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Disclosures

References

1. Jones CU, Pugh SL, Sandler HM, et al. Adding short-term androgen deprivation therapy to radiation therapy in men with localized prostate cancer: long-term update of the NRG/ RTOG 9408 randomized clinical trial [published correction appears in Int J Radiat Oncol Biol Phys 2023;115:265]. Int J Radiat Oncol Biol Phys 2022;112:294–303. DOI: 10.1016/ j.ijrobp.2021.08.031. [PubMed: 34481017]

2. Pilepich MV, Winter K, Lawton CA, et al. Androgen suppression adjuvant to definitive radiotherapy in prostate carcinoma — long-term results of phase III RTOG 85–31. Int J Radiat Oncol Biol Phys 2005;61:1285–1290. DOI: 10.1016/j.ijrobp.2004.08.047. [PubMed: 15817329]

3. D'Amico AV, Chen M-H, Renshaw A, Loffredo M, Kantoff PW. Long-term follow-up of a randomized trial of radiation with or without androgen deprivation therapy for localized prostate cancer. JAMA 2015;314:1291–1293. DOI: 10.1001/jama.2015.8577. [PubMed: 26393854]

4. Bolla M, Neven A, Maingon P, et al. Short androgen suppression and radiation dose escalation in prostate cancer: 12-year results of EORTC Trial 22991 in patients with localized intermediate-risk disease. J Clin Oncol 2021;39:3022–3033. DOI: 10.1200/JCO.21.00855. [PubMed: 34310202]

5. Kishan AU, Sun Y, Hartman H, et al. Androgen deprivation therapy use and duration with definitive radiotherapy for localised prostate cancer: an individual patient data meta-analysis [published correction appears in Lancet Oncol 2022;23:e319]. Lancet Oncol 2022;23: 304–316. DOI: 10.1016/ S1470-2045(21)00705-1. [PubMed: 35051385]

6. Nguyen PL, Alibhai SMH, Basaria S, et al. Adverse effects of androgen deprivation therapy and strategies to mitigate them. Eur Urol 2015;67:825–836. DOI: 10.1016/j.eururo.2014.07.010. [PubMed: 25097095]

7. 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;26:2497–2504. DOI: 10.1200/JCO.2007.14.9021. [PubMed: 18413638]

8. Roach M III, DeSilvio M, Lawton C, et al. Phase III trial comparing whole-pelvic versus prostate-only radiotherapy and neoadjuvant versus adjuvant combined androgen suppression: Radiation Therapy Oncology Group 9413. J Clin Oncol 2003;21:1904–1911. DOI: 10.1200/JCO.2003.05.004. [PubMed: 12743142]

9. Pisansky TM, Hunt D, Gomella LG, et al. Duration of androgen suppression before radiotherapy for localized prostate cancer: radiation therapy oncology group randomized clinical trial 9910. J Clin Oncol 2015;33:332–339. DOI: 10.1200/JCO.2014.58.0662. [PubMed: 25534388]

10. Michalski JM, Moughan J, Purdy J, et al. Effect of standard vs dose-escalated radiation therapy for patients with intermediate-risk prostate cancer: the NRG Oncology RTOG 0126 randomized clinical trial. JAMA Oncol 2018;4:e180039. DOI: 10.1001/jamaoncol.2018.0039. [PubMed: 29543933]

11. Schaeffer E, Srinivas S, Antonarakis ES, et al. NCCN Guidelines Insights: Prostate Cancer, Version 1.2021. J Natl Compr Canc Netw 2021;19:134–143. DOI: 10.6004/jnccn.2021.0008. [PubMed: 33545689]

12. Spratt DE, Zhang J, Santiago-Jiménez M, et al. Development and validation of a novel integrated clinical-genomic risk group classification for localized prostate cancer. J Clin Oncol 2018;36:581–590. DOI: 10.1200/JCO.2017.74.2940. [PubMed: 29185869]

13. Gutman DA, Khalilia M, Lee S, et al. The digital slide archive: a software platform for management, integration, and analysis of histology for cancer research. Cancer Res 2017;77:e75–e78. DOI: 10.1158/0008-5472.CAN-17-0629. [PubMed: 29092945]

14. Tolkach Y, Dohmgörgen T, Toma M, Kristiansen G. High-accuracy prostate cancer pathology using deep learning. Nat Mach Intell 2020;2:411–418. 10.1038/s42256-020-0200-7.

15. Nagpal K, Foote D, Tan F, et al. Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. JAMA Oncol 2020;6:1372–1380. DOI: 10.1001/jamaoncol.2020.2485. [PubMed: 32701148]

16. Pantanowitz L, Quiroga-Garza GM, Bien L, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. Lancet Digit Health 2020;2:e407–e416. DOI: 10.1016/S2589-7500(20)30159-X. [PubMed: 33328045]

17. Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials [published correction appears in NPJ Digit Med 2023;6:27]. NPJ Digit Med 2022;5:71. DOI: 10.1038/s41746-022-00613-w. [PubMed: 35676445]

18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Poster presented at 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, June 27–30, 2016.

19. Chen X, Fan H, Girshick R, He K. Improved baselines with momentum contrastive learning. March 9, 2020 (http://arxiv.org/abs/2003.04297). preprint.

20. Hutter F, Kotthoff L, Vanschoren J, eds. Automated machine learning: methods, systems, challenges. Cham, Switzerland: Springer, 2019. DOI: 10.1007/978-3-030-05318-5.

21. Claesen M, De Moor B. Hyperparameter search in machine learning. April 6, 2015 (https://arxiv.org/abs/1502.02127). preprint.

22. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;94:496–509. DOI: 10.1080/01621459.1999.10474144.

23. Rubin E, ed. Striving for diversity in research studies. N Engl J Med 2021;385:1429–1430. DOI: 10.1056/NEJMe2114651. [PubMed: 34516052]

24. Jones CU, Hunt D, McGowan DG, et al. Radiotherapy and short-term androgen deprivation for localized prostate cancer. N Engl J Med 2011;365:107–118. DOI: 10.1056/NEJMoa1012348. [PubMed: 21751904]

25. National Cancer Institute. Cancer stat facts: prostate cancer. SEER. April 10, 2023 (https://seer.cancer.gov/statfacts/html/prost.html).

26. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009;101:1446–1452. DOI: 10.1093/jnci/djp335. [PubMed: 19815849]

27. Cooperberg MR, Pasta DJ, Elkin EP, et al. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. J Urol 2005;173:1938–1942. DOI: 10.1097/01.ju.0000158155.33890.e7. [PubMed: 15879786]

28. Zumsteg ZS, Spratt DE, Pei I, et al. A new risk classification system for therapeutic decision making with intermediate-risk prostate cancer patients undergoing dose-escalated external-beam radiation therapy. Eur Urol 2013;64:895–902. DOI: 10.1016/j.eururo.2013.03.033. [PubMed: 23541457]
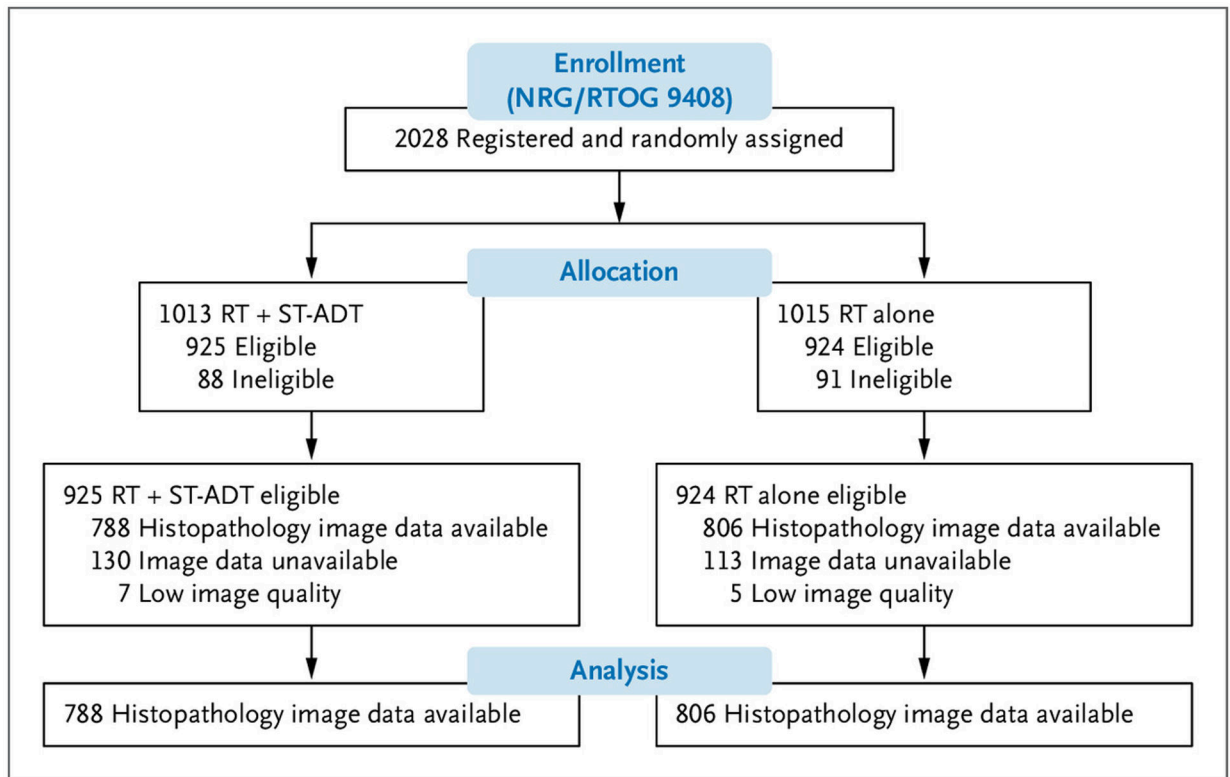
Figure 1.
Consolidated Standards of Reporting Trials Flow Diagram for NRG/RTOG 9408 (Validation Set).

RT denotes radiotherapy; and ST-ADT, short-term androgen deprivation therapy.
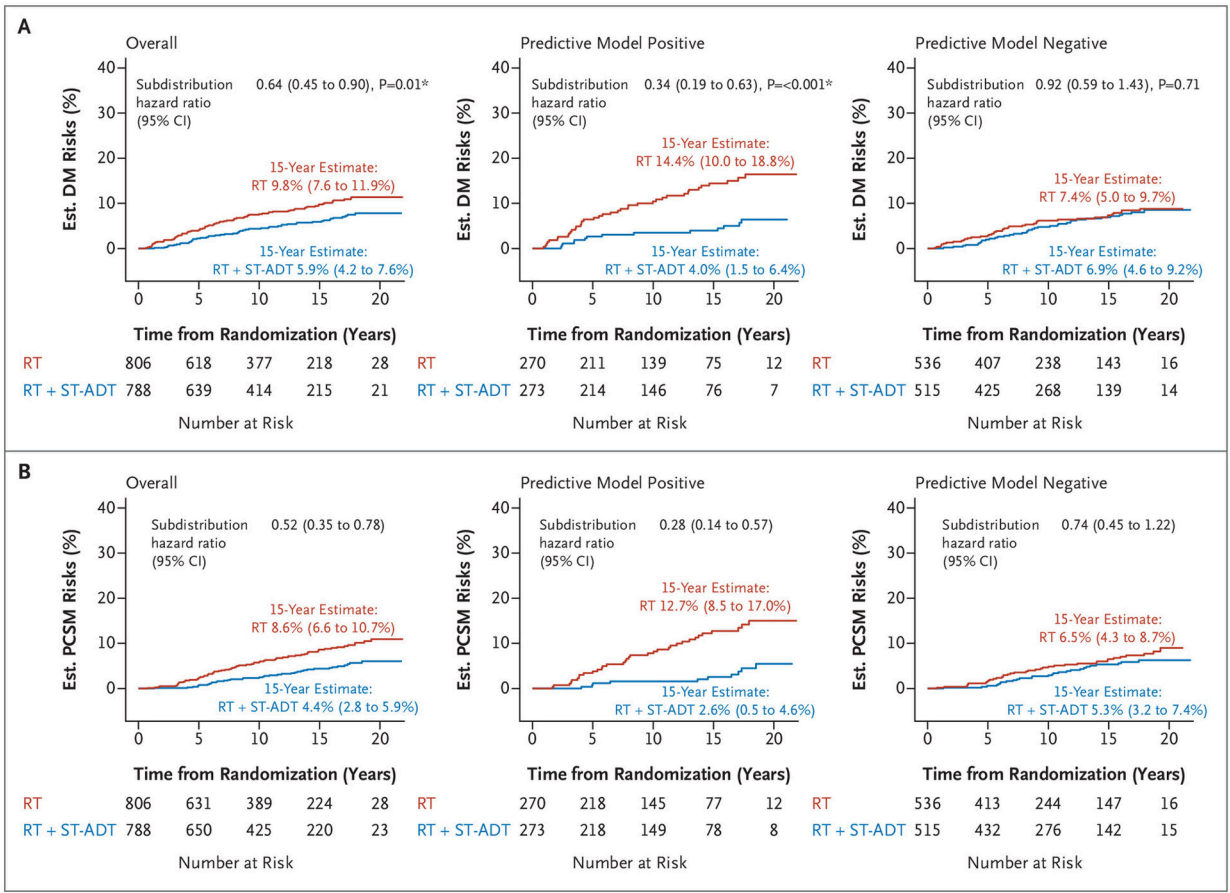
Figure 2.
Cumulative Incidence in the Validation Cohort (NRG/RTOG 9408) of Histopathology-Imaged Patients by Artificial Intelligence–Predictive Model Subgroups for (Panel A) Distant Metastasis and (Panel B) Prostate Cancer–Specific Mortality.

CI denotes confidence interval; DM, distant metastasis; Est., estimated; PCSM, prostate cancer–specific mortality; RT, radiotherapy; and ST-ADT, short-term androgen deprivation therapy. * Denotes P value <0.05.

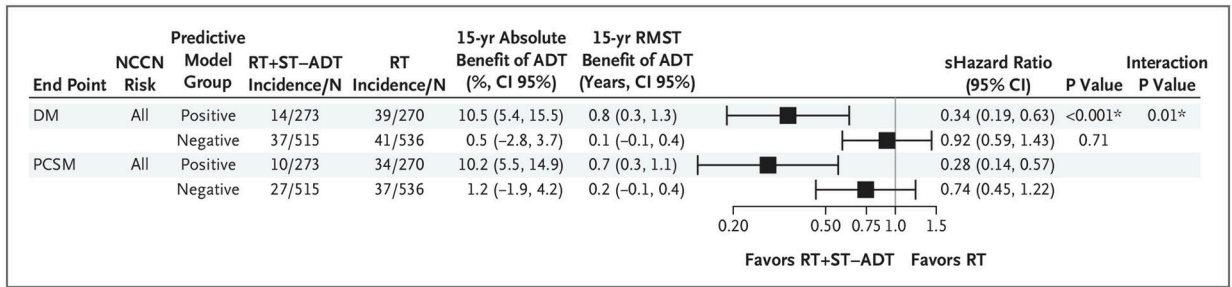| End Point	NCCN Risk	Predictive Model Group	RT+ST–ADT Incidence/N	RT Incidence/N	15-yr Absolute Benefit of ADT (%, CI 95%)	15-yr RMST Benefit of ADT (Years, CI 95%)		sHazard Ratio (95% CI)	P Value	Interaction P Value
DM	All	Positive	14/273	39/270	10.5 (5.4, 15.5)	0.8 (0.3, 1.3)		0.34 (0.19, 0.63)	<0.001*	0.01*
		Negative	37/515	41/536	0.5 (–2.8, 3.7)	0.1 (–0.1, 0.4)		0.92 (0.59, 1.43)	0.71	
PCSM	All	Positive	10/273	34/270	10.2 (5.5, 14.9)	0.7 (0.3, 1.1)		0.28 (0.14, 0.57)		
		Negative	27/515	37/536	1.2 (–1.9, 4.2)	0.2 (–0.1, 0.4)		0.74 (0.45, 1.22)		

0.20   0.50 0.75 1.0   1.5

Favors RT+ST–ADT   Favors RT

Figure 3.
Forest Plots for All End Points in Positive and Negative Predictive Model Groups of NRG/
RTOG 9408 (Validation Set) for All Patients.

ADT denotes androgen deprivation therapy; CI, confidence interval; DM, distant metastasis;
N, number of patients; NCCN, National Comprehensive Cancer Network; PCSM, prostate
cancer–specific mortality; RMST, restricted mean survival time; RT, radiation therapy; ST-
ADT, short-term androgen deprivation therapy; and yr, year. * Denotes P value <0.05.

Table 1.

Patient Baseline Characteristics for NRG/RTOG 9408.[*]

Characteristic	NRG/RTOG 9408 Full Cohort (N=1974)			NRG/RTOG 9408 Imaged Cohort (n=1594)	
	Overall (N=1974)	Imaged (n=1594)	Not Available (n=380)	RT (n=806)	RT + ST-ADT (n=788)
Group					
RT	990 (50.2)	806 (50.6)	184 (48.4)	—	—
RT + ST-ADT	984 (49.8)	788 (49.4)	196 (51.6)	—	—
Age, years					
Median (IQR)	71 (66–74)	71 (66–74)	70 (66–74)	71 (66–74)	70 (66–74)
Missing	1	0	1	—	—
Race					
Black	394 (20.0)	306 (19.2)	88 (23.2)	150 (18.6)	156 (19.8)
White	1,497 (75.8)	1,220 (76.5)	277 (72.9)	624 (77.4)	596 (75.6)
Other	80 (4.1)	65 (4.1)	15 (3.9)	31 (3.8)	34 (4.3)
Unknown	3 (0.2)	3 (0.2)	0 (0.0)	1 (0.1)	2 (0.3)
KPS					
70–80	154 (7.8)	126 (7.9)	28 (7.4)	60 (7.4)	66 (8.4)
90–100	1819 (92.2)	1468 (92.1)	351 (92.6)	746 (92.6)	722 (91.6)
Missing	1	0	1	—	—
Baseline PSA, ng/ml					
Median (IQR)	8 (6–12)	8 (6–12)	7 (5–10)	8 (6–12)	8 (6–12)
<4	209 (10.6)	145 (9.1)	64 (16.9)	66 (8.2)	79 (10.0)
4–10	1089 (55.2)	874 (54.8)	215 (56.7)	448 (55.6)	426 (54.1)
10–20	669 (33.9)	570 (35.8)	99 (26.1)	288 (35.7)	282 (35.8)
>20	6 (0.3)	5 (0.3)	1 (0.3)	4 (0.5)	1 (0.1)
Missing	1	0	1	—	—
Tumor stage					
T1	962 (48.8)	775 (48.6)	187 (49.3)	379 (47.0)	396 (50.3)
T2	1011 (51.2)	819 (51.4)	192 (50.7)	427 (53.0)	392 (49.7)
Missing	1	0	1	—	—
Nodal stage					
N0	80 (4.1)	67 (4.2)	13 (3.4)	33 (4.1)	34 (4.3)
Nx	1893 (95.9)	1527 (95.8)	366 (96.6)	773 (95.9)	754 (95.7)
Missing	1	0	1	—	—
Gleason score					
<7	1212 (62.9)	969 (62.2)	243 (65.7%)	475 (60.6%)	494 (63.9%)
7	535 (27.8)	437 (28.1)	98 (26.5%)	233 (29.7%)	204 (26.4%)
8–10	180 (9.3)	151 (9.7)	29 (7.8)	76 (9.7)	75 (9.7)
Missing	47	37	10	22	15
Risk group					
High	180 (9.3)	151 (9.7)	29 (7.8)	76 (9.7)	75 (9.7)
Intermediate	1071 (55.6)	878 (56.4)	193 (52.2)	453 (57.8)	425 (55.0)

Characteristic	NRG/RTOG 9408 Full Cohort (N=1974)			NRG/RTOG 9408 Imaged Cohort (n=1594)	
	Overall (N=1974)	Imaged (n=1594)	Not Available (n=380)	RT (n=806)	RT + ST-ADT (n=788)
Low	676 (35.)	528 (33.9)	148 (40.0)	255 (32.5)	273 (35.3)
Missing	47	37	10	22	15

*
Values are presented as No. (%) unless indicated otherwise. Note that some percentages may not add up to 100% because of rounding. Karnofsky performance status (KPS) scores range from 0 to 100. A higher score indicates the patient having better ability to carry out daily activities. IQR denotes interquartile range; n, number of patients; NRG/RTOG, NRG Oncology/Radiation Therapy Oncology Group; PSA, prostate-specific antigen; RT, radiation therapy; and ST-ADT, short-term androgen deprivation therapy.