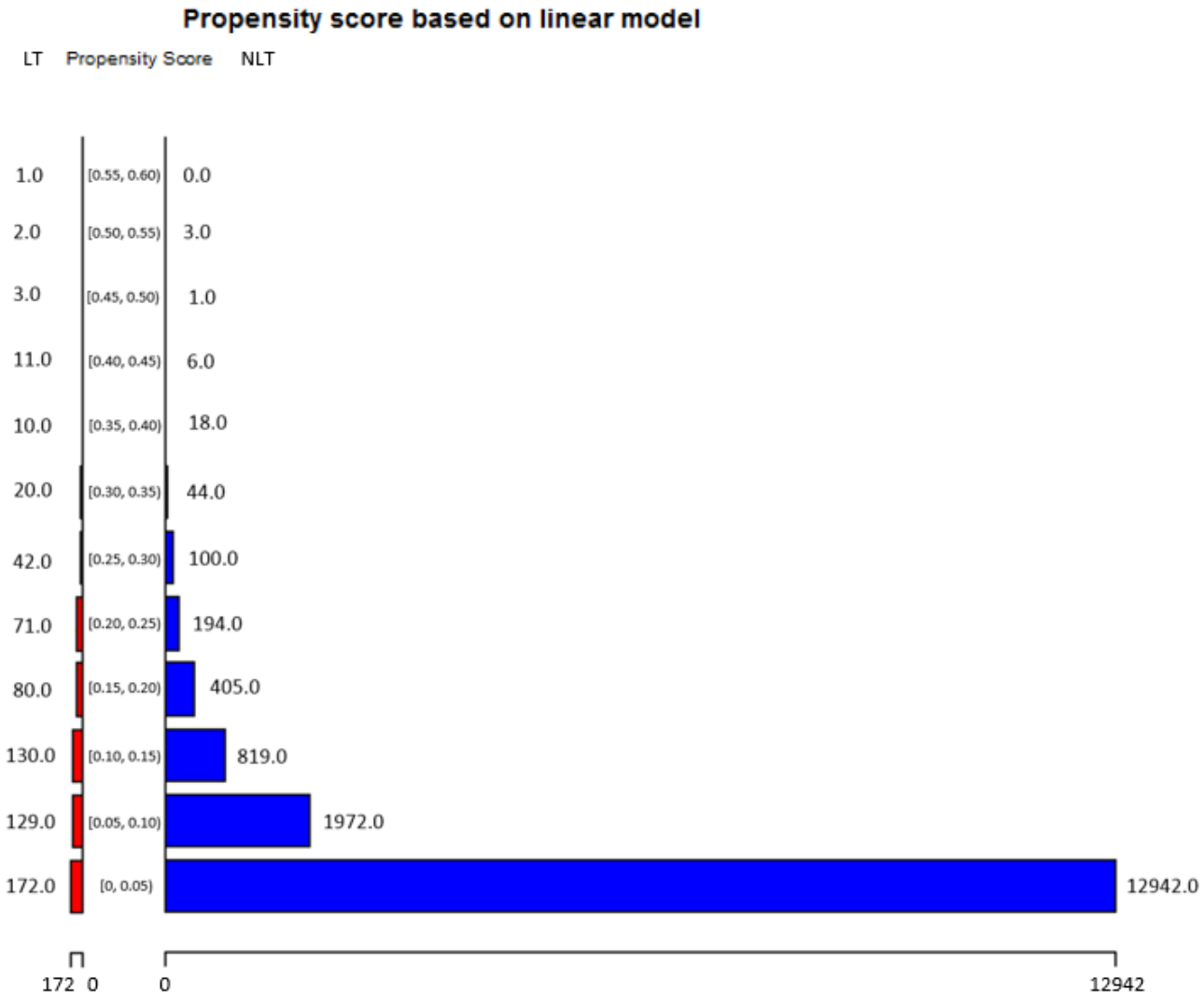
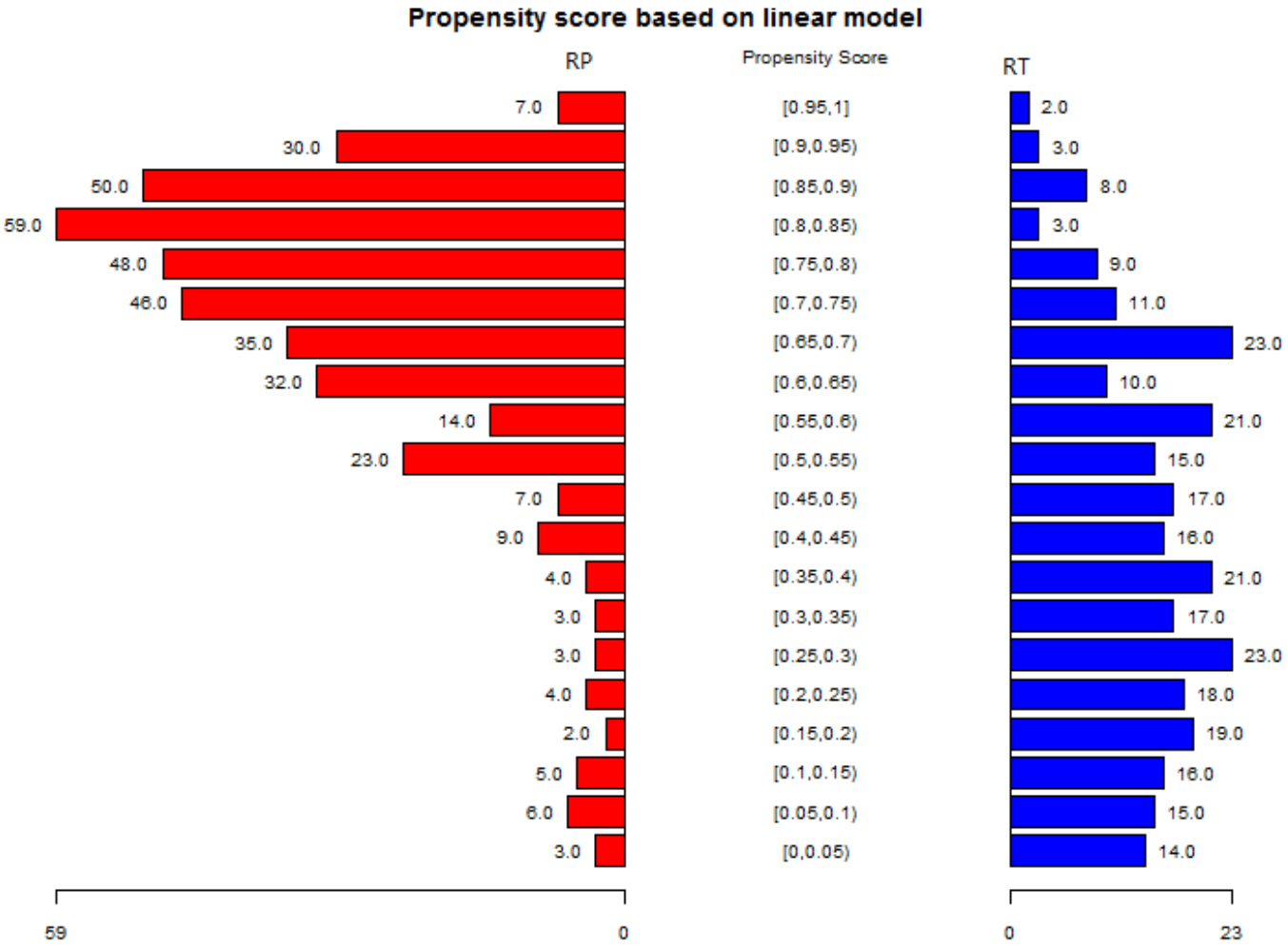
Supplement Table 1: Propensity score parameter list

the variables used in calculating the propensity matching	Age, marital status, race, PSA, GS, TNM stages			
Propensity scoring algorithm	Logistic regression model			
C-statistical	0.8431		0.9105	
Matching method	Greedy matching within specified caliper distances			
Distance metric	0.05			
Matching ratio	(No local treatment) 1:1 (Local treatment)		(Radiation therapy) 1:1 (Radical prostatectomy)	
Use of replacement	With replacement			
Matching sample size	NLT: 667 cases	Total: 1334 cases	RT: 168 cases	Total: 336 cases
	LT: 667 cases		RP: 168 cases	

Supplement Figure 1: Propensity score matching schematic diagram





Supplement Table 2: sensitivity analysis from propensity score matching (PSM)

Several models were used to examine robustness of the treatment effects in the comparison of NLT vs LT and RT vs RP:

(1) Inverse probability of treatment weighting (IPTW) logistic regression model
(2) Standard mortality ratio weighting (SMRW) logistic regression model
(3) Covariate adjustment propensity score (CAPS) model
(4) Double propensity score (DPS) adjusted model
(5) Propensity score (PPS) stratified model

Inverse probability of treatment weighting (IPTW)

IPTW attempts to estimate the effect of treatment among the same individuals [1]. It relies on the assumption that the distribution of risk factors in patients received treatment is equal to that found in all individuals, thus making treatment choices balanced [2, 3].

Standard mortality ratio weighting (SMRW)

SMRW attempts to estimate the standard effect measure that considers the treatment group as the standard population, making the distribution of risk factors in all patients is equal to that found in the treatment group [4].

Covariate adjustment propensity score (CAPS)

After performing propensity score matching calculated by covariates, the confounders in the two treatment groups can be expressed as propensity score. Distribution of several covariates differs after matching (Table 2), indicating that adjustment of PS is necessary for validation.

Double propensity score (DPS) adjusted

Considering unbalanced baseline characteristics after matching, additional double propensity score adjusted model was performed in the matched cohort to further overcome the deviation caused by confounders.

Propensity score stratified

After straitified the whole population into five groups according to propensity score, individual with similar conditions were combined (eg. Those with poor conditions or better conditions).

	OS	CSM
NLT vs LT		
IPTW model		
Non-adjusted	**1.05 (1.03, 1.08)**	**1.17 (1.14, 1.21)**
Adjusted	0.96 (0.94, 0.99)	**1.03 (1.00, 1.06)**
SMRW model		
Non-adjusted	0.56 (0.48, 0.65)	0.54 (0.45, 0.64)
Adjusted	0.58 (0.49, 0.69)	0.58 (0.48, 0.71)
CAPS model		
Non-adjusted	0.39 (0.35, 0.44)	0.39 (0.34, 0.45)
Covariate PS as continuous	0.57 (0.50, 0.65)	0.57 (0.49, 0.66)
Covariate PS as categorical (divided into five groups)	0.51 (0.45, 0.58)	0.52 (0.45, 0.60)
DPS model	0.57 (0.49, 0.68)	0.54 (0.45, 0.66)
PPS stratified model		
Q1	**1.15 (1.10, 1.21)**	**1.36 (1.29, 1.43)**
Q2	**1.24 (1.18, 1.30)**	**1.34 (1.27, 1.42)**
Q3	0.99 (0.93, 1.05)	**1.11 (1.04, 1.19)**
Q4	0.68 (0.63, 0.73)	0.71 (0.66, 0.77)
Q5	0.38 (0.35, 0.41)	0.36 (0.33, 0.40)
RT vs RP		
IPTW model		
Non-adjusted	0.63 (0.54, 0.73)	0.65 (0.55, 0.77)
Adjusted	0.73 (0.60, 0.88)	0.72 (0.58, 0.89)
SMRW model		
Non-adjusted	0.43 (0.34, 0.53)	0.38 (0.29, 0.49)
Adjusted	0.48 (0.36, 0.65)	0.47 (0.33, 0.66)
CAPS model		
Non-adjusted	0.47 (0.37, 0.59)	0.43 (0.32, 0.56)
Covariate PS as continuous	0.59 (0.44, 0.81)	0.61 (0.43, 0.88)
Covariate PS as categorical (divided into five groups)	0.51 (0.36, 0.71)	0.52 (0.35, 0.77)
DPS model	0.64 (0.41, 0.98)	0.74 (0.45, 1.21)
PPS stratified model		
Q1	0.62 (0.59, 0.65)	0.58 (0.55, 0.61)
Q2	0.17 (0.14, 0.20)	0.18 (0.15, 0.23)
Q3	0.62 (0.52, 0.74)	0.72 (0.60, 0.88)
Q4	0.18 (0.14, 0.22)	0.09 (0.07, 0.13)
Q5	0.26 (0.19, 0.36)	0.37 (0.25, 0.55)

OS = overall survival, CSM = cancer specific mortality, PS = propensity score
IPTW model and SMRW model: adjust for age, marital status, race, prostate specific antigen, Gleason score, TNM stages
CAPS model: adjust for propensity score
DPS model: NLT vs LT, adjusted for marital status, T stage; RT vs RP, adjusted for T stage, N stage and Gleason score

The effects of treatment in the IPTW model changed, probably showing that a fraction of patients received LT had poor conditions and led to death easier. And this model enlarged this consequence, causing the outcome unstable slightly. PPS straitified model also obtain that for those with lower PPS, patients might not benefit from LT. However, other sensitivity analyses all showed LT could reduce risks of death, reflecting general situation among the whole population.

Supplement Table 3: statistical methods for the instrumental variable analysis
The instrument variable (IV) for this study is the use of each therapy in different region per year (therapy cases/region/yr). For the IV to be considered valid, we performed a three-step analysis.

Step 1: to confirm that IV must be associated with the treatment.
To confirm the association between treatment and IV, in our study IV was calculated by the frequency of treatment use. Besides, it is widely considered that IV is strong with F-statistic > 10 [5, 6]. Our analysis showed that IV is highly correlated with treatment (F statistic = 652.798, p<0.001 and F statistic = 41.019, p<0.001, respectively).

Step 2: to confirm that IV must not be associated with the outcome.
From multivariate analysis, IV was not correlated with OS (p=0.989) and CSM (p=0.590) in the comparison of NLT vs LT.
As for RT vs RP, IV was also not correlated with OS (p=0.593) and CSM (p=0.463)

	OS	CSM
Frequency of therapy use per 10% increase	1.00 (0.93, 1.08) p=0.989	1.02 (0.94, 1.12) p=0.590
Marital status		
Married	1.0	1.0
Single	1.22 (1.15, 1.29)	1.12 (1.05, 1.20)
Divorced/Widowed	1.17 (1.11, 1.22)	1.12 (1.06, 1.19)
Unknown	0.92 (0.85, 0.99)	0.90 (0.82, 0.98)
Age	1.02 (1.02, 1.03)	1.02 (1.01, 1.02)
Race		
White	1.0	1.0
Black	0.99 (0.94, 1.04)	0.97 (0.91, 1.02)
Other	0.81 (0.74, 0.88)	0.77 (0.70, 0.86)
Unknown	0.32 (0.21, 0.49)	0.16 (0.08, 0.32)
T stage		
T1	1.0	1.0
T2	1.06 (1.00, 1.11)	1.07 (1.00, 1.13)
T3	0.99 (0.92, 1.07)	1.03 (0.94, 1.12)
T4	1.36 (1.27, 1.45)	1.45 (1.35, 1.56)
Unknown	1.37 (1.29, 1.46)	1.40 (1.30, 1.50)
N stage		
N0	1.0	1.0
N1	1.06 (1.01, 1.11)	1.11 (1.04, 1.17)
Unknown	1.08 (1.03, 1.14)	1.10 (1.04, 1.16)
M stage		
M1a	1.0	1.0

M1b	1.39 (1.27, 1.52)	1.57 (1.41, 1.75)
M1c	1.71 (1.55, 1.88)	1.94 (1.73, 2.18)
Gleason score		
GS ≤ 6	1.0	1.0
GS = 7	0.86 (0.78, 0.95)	0.90 (0.79, 1.01)
GS ≥ 8	1.15 (1.05, 1.25)	1.31 (1.18, 1.46)
Unknown	1.10 (0.99, 1.22)	1.22 (1.08, 1.38)
PSA	1.00 (1.00, 1.00)	1.01 (1.00, 1.01)
Treatment		
NLT	1.0	1.0
LT	0.56 (0.42, 0.76)	0.63 (0.45, 0.87)
Year of Diagnosis		
2004	1.0	1.0
2005	1.10 (1.01, 1.20)	1.12 (1.01, 1.23)
2006	0.99 (0.91, 1.08)	0.99 (0.90, 1.10)
2007	1.05 (0.97, 1.15)	1.00 (0.90, 1.11)
2008	0.98 (0.90, 1.06)	0.97 (0.87, 1.07)
2009	0.96 (0.88, 1.05)	0.95 (0.86, 1.05)
2010	0.97 (0.86, 1.08)	0.96 (0.85, 1.09)
2011	0.97 (0.86, 1.08)	0.94 (0.82, 1.07)
2012	0.89 (0.80, 1.00)	0.88 (0.77, 1.00)
2013	0.95 (0.84, 1.06)	0.94 (0.83, 1.08)
2014	0.95 (0.83, 1.07)	0.92 (0.79, 1.06)
2015	0.76 (0.64, 0.91)	0.84 (0.68, 1.03)
Region		
Pacific Coast	1.0	1.0
East	1.14 (1.02, 1.28)	1.14 (1.00, 1.31)
Northern Plains	1.13 (0.84, 1.51)	1.21 (0.86, 1.69)
Southwest	1.10 (0.79, 1.54)	1.21 (0.82, 1.79)

	OS	CSM
Frequency of therapy use per 10% increase	1.07 (0.84, 1.36) p = 0.593	1.11 (0.84, 1.47) p = 0.463
Marital status		
Married	1.0	1.0
Single	1.25 (0.81, 1.91)	1.30 (0.79, 2.14)
Divorced/Widowed	1.36 (0.94, 1.97)	1.07 (0.68, 1.69)
Unknown	0.94 (0.53, 1.66)	0.88 (0.46, 1.67)
Age	1.02 (1.01, 1.04)	1.01 (1.00, 1.03)
Race		
White	1.0	1.0
Black	0.96 (0.66, 1.40)	0.81 (0.51, 1.27)
Other	1.05 (0.62, 1.79)	1.02 (0.56, 1.87)
Unknown	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
T stage		
T1	1.0	1.0
T2	1.03 (0.69, 1.54)	0.75 (0.46, 1.22)
T3	1.16 (0.73, 1.84)	1.05 (0.61, 1.78)

T4	2.09 (1.26, 3.45)	1.90 (1.08, 3.35)
Unknown	1.88 (1.06, 3.31)	1.68 (0.90, 3.14)
N stage		
N0	1.0	1.0
N1	1.13 (0.78, 1.63)	1.23 (0.81, 1.85)
Unknown	0.94 (0.63, 1.42)	0.94 (0.58, 1.51)
M stage		
M1a	1.0	1.0
M1b	1.70 (1.01, 2.84)	3.02 (1.54, 5.94)
M1c	2.59 (1.50, 4.50)	4.35 (2.13, 8.86)
Gleason score		
GS ≤ 6	1.0	1.0
GS = 7	1.48 (0.78, 2.79)	1.45 (0.61, 3.43)
GS ≥ 8	2.87 (1.60, 5.16)	4.06 (1.87, 8.80)
Unknown	2.31 (1.09, 4.88)	2.15 (0.81, 5.76)
PSA	1.01 (1.01, 1.01)	1.01 (1.01, 1.02)
Treatment		
RT	1.0	1.0
RP	0.56 (0.40, 0.80)	0.58 (0.38, 0.88)
Year of Diagnosis		
2004	1.0	1.0
2005	0.74 (0.44, 1.26)	0.80 (0.43, 1.48)
2006	0.73 (0.42, 1.24)	0.72 (0.38, 1.36)
2007	1.04 (0.60, 1.82)	1.00 (0.52, 1.92)
2008	0.75 (0.43, 1.29)	0.61 (0.31, 1.19)
2009	0.89 (0.51, 1.55)	0.84 (0.44, 1.60)
2010	0.80 (0.38, 1.68)	1.10 (0.46, 2.66)
2011	1.01 (0.48, 2.13)	1.11 (0.45, 2.75)
2012	0.79 (0.33, 1.90)	1.00 (0.36, 2.82)
2013	0.64 (0.27, 1.51)	0.81 (0.29, 2.27)
2014	0.47 (0.19, 1.16)	0.66 (0.24, 1.87)
2015	0.40 (0.05, 3.20)	0.71 (0.08, 6.01)
Region		
Pacific Coast	1.0	1.0
East	1.39 (1.02, 1.89)	1.22 (0.84, 1.76)
Northern Plains	1.31 (0.71, 2.39)	1.68 (0.85, 3.34)
Southwest	1.58 (0.79, 3.17)	1.68 (0.75, 3.78)

Step 3: to determine whether the IV model could estimate the relationship between treatment and outcome

The Durbin-Wu-Hausman test was performed. The result indicated that the residual can affect the relationship for NLT vs LT (F statistic = 58.410, $p<0.001$ for CSM and F statistic = 85.459, $p<0.001$ for OS, respectively), as well as RT vs RP (F statistic = 6.986, $p=0.008$ for CSM and F statistic = 8.408, $p=0.004$ for OS, respectively), thus IV model could estimate the effect of RP and RT more accurately than the standard model (Cox multivariate regression) [5].

References

1 Curtis LH, Hammill BG, Eisenstein EL, Kramer JM, Anstrom KJ. Using inverse probability-weighted estimators in comparative effectiveness analyses with observational databases. Med Care. 2007;45(10 Supl 2):S103-S107.

2 Robins JM. Marginal structural models. In: 1997 Proceedings of the Section on Bayesian Statistical Science. Alexandria, VA: American Statistical Association, 1998:1–10.

3 Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550–60.

4 Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. Epidemiology 2003;14:680–6.

5 Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med. 2014;33(February):2297-2340.

6 Rassen J a, Brookhart MA, Glynn RJ, Mittleman M a, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. J Clin Epidemiol. 2009;62(12):1233-1241.