Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2016 June 1; 95(2): 703–706. doi:10.1016/j.ijrobp.2016.01.034.

Validating a Local Failure Risk Stratification for Use in Prospective Studies of Adjuvant Radiation for Bladder Cancer

Brian C. Baumann, MD¹, Jiwei He, PhD², Wei-Ting Hwang, PhD², Kai N. Tucker, BA¹, Justin E. Bekelman, MD¹, Harry W. Herr, MD³, Seth P. Lerner, MD⁴, Thomas J. Guzzo, MD, MPH⁵, S. Bruce Malkowicz, MD⁵, and John P. Christodouleas, MD, MPH¹

¹University of Pennsylvania, Department of Radiation Oncology, Philadelphia, PA

²University of Pennsylvania, Department of Epidemiology and Biostatistics, Philadelphia, PA

³Memorial Sloan Kettering Cancer Center, New York, NY

⁴Baylor College of Medicine, Department of Urology, Houston, TX

⁵University of Pennsylvania, Department of Urology, Philadelphia, PA

Abstract

Purpose/Objectives—To inform prospective trials of adjuvant radiation therapy (adj-RT) for bladder cancer after radical cystectomy, a local-regional failure (LF) risk stratification was proposed. This stratification was developed and validated using surgical databases that may not reflect the outcomes expected in prospective trials. Our purpose was to assess sources of bias that may impact the stratification model's validity or alter the LF risk estimates for each subgroup: time bias due to evolving surgical techniques; trial accrual bias due to inclusion of patients who would be ineligible for adj-RT trials due to early disease progression, death, or loss to follow-up shortly after cystectomy; bias due to different statistical methodologies to estimate LF; and subgrouping bias due to different definitions of the LF subgroups.

Methods and Materials—The LF risk stratification was developed using a single-institution cohort (n=442, 1990–2008) and the multi-institutional SWOG 8710 cohort (n=264, 1987–1998) treated with radical cystectomy +/- chemotherapy. We evaluated the sensitivity of the stratification to sources of bias using Fine-Gray regression and Kaplan-Meier analyses.

Results—Year of radical cystectomy was not associated with LF risk on univariate or multivariate analysis after controlling for risk group. Using more stringent inclusion criteria, 26 SWOG patients (10%) and 60 (14%) from the single-institution cohort were excluded. Analysis of the remaining patients confirmed 3 subgroups with significantly different LF risk with 3-year rates of 7%, 17%, and 36%, respectively ($p<0.01$), nearly identical to the rates without correcting for trial accrual bias. Kaplan-Meier techniques estimated higher subgroup LF rates than competing risk analysis. The subgroup definitions used in the NRG-GU001 adj-RT trial were validated.

Corresponding author: John P. Christodouleas, MD, MPH, 3400 Civic Center Blvd, PCAM 2, University of Pennsylvania, Department of Radiation Oncology, Philadelphia, PA 19103, christojo@uphs.upenn.edu, Phone: 215-662-2428.

Conflict of interest: J.C. discloses employment at Elekta AB. The authors have nothing else to disclose.

Conclusions—These sources of bias did not invalidate the LF risk stratification or substantially change the model's LF estimates.

SUMMARY

To inform trials of adjuvant radiotherapy for bladder cancer, a local-regional failure (LF) risk stratification was developed. We assessed the impact of multiple potential biases on the model's validity for predicting outcomes. The predictions were not invalidated by time bias from evolving surgical techniques, use of different statistical methods to estimate LF risk, subgrouping bias due to differing definitions for the LF subgroups, or trial accrual bias from including patients ineligible for adjuvant RT trials.

INTRODUCTION

Patients with advanced bladder cancer (stage pT3) have a five-year overall survival of ~50% after cystectomy and pelvic lymphadenectomy +/- chemotherapy (1) with approximately half of recurrences in the pelvis, either as isolated failures or synchronous with distant metastases (2, 3). Multiple organizations, including the NRG, are developing prospective trials of adjuvant radiation (adj-RT) to reduce the risk of local-regional failure (LF).

To inform the design of such trials, a LF risk model was developed based on two cohorts totaling 706 cystectomy patients stratified into three LF risk groups based on pathologic T-stage, margin status, and number of nodes identified at surgery (2). The low-risk subgroup has stage pT2 disease; the intermediate-risk subgroup has pT3 disease, negative margins, and ≤10 nodes identified; and the high-risk subgroup has pT3 disease with positive margins OR >10 nodes identified. This stratification was developed and externally validated (4, 5) using historical surgical databases that may not reflect the outcomes expected in the observation arm of a contemporary prospective trial.

We examined four potential sources of bias within our risk stratification that could affect its validity or substantially change the absolute LF risk estimate for each subgroup.

METHODS

The LF risk stratification was developed using a single-institution cohort (n=442, 1990–2008) and the multi-institutional SWOG 8710 cohort (n=264, 1987–1998) treated with cystectomy +/- chemotherapy. Details of the cohorts were described previously (1–3, 6).

We assessed time bias by determining if improving surgical techniques over time would alter the risk stratification results. We assessed trial accrual bias by excluding patients from the model's database who would be ineligible for adj-RT trials because they died, developed distant metastasis and/or LF, or were lost to follow-up within 90 days of surgery or 45 days of post-operative chemotherapy. Failure to exclude such patients might overestimate the LF rates expected in the control arm of an adj-RT trial. We also evaluated the effect of different statistical methods for estimating LF rates as a potential source of bias by comparing the LF rates for each subgroup in the original model, which were derived using the competing risk method of cumulative incidence, to LF rates calculated using the Kaplan-Meier method

commonly used in trial power analyses. We assessed subgrouping bias by evaluating the impact on predictions of LF of the alternative subgroup definitions used in the new NRG-GU001 trial of adj-RT that excludes pT3a patients from the intermediate-risk subgroup. Finally, we calculated the absolute LF rates expected in the control arm of clinical trials, including NRG-GU001, accounting for these different sources of bias.

RESULTS

Year of cystectomy as a proxy for evolving surgical techniques was not associated with LF risk on univariate Fine-Gray competing risk analysis (subhazard ratio 1.01, 95% CI 0.98–1.03, $p=0.65$). Adjusting for risk group, the year of cystectomy was not a significant predictor of LF on competing risk multivariate analysis (Table 1).

Eighty-six patients from the risk model's original database who developed LF, distant metastases, died, or were lost to follow up within 90 days of surgery or 45 days of post-operative chemotherapy were excluded as these patients are unlikely to be enrolled in a prospective radiation trial. Using this modified database, 3-year cumulative incidence of LF for the three risk subgroups were estimated and compared using Gray's test. The three subgroups continued to have significantly different LF rates of 7%, 17%, and 36%, respectively ($p<0.01$), that are nearly identical to the rates seen in the original, full patient cohort (Figure 1/Table 2).

We repeated the analysis using Kaplan-Meier methods rather than the original competing risk assessment. Kaplan-Meier analysis confirmed that the three subgroups still had significantly different LF risks with three-year LF rates of 8%, 25%, and 57%, respectively ($p<0.001$), that were larger for each subgroup than the estimated LF risks originally reported using competing risk techniques (Figure 2/Table 2).

NRG-GU001 adopted our risk stratification with the exception that patients with pT3a disease, negative margins, and 10 nodes identified were re-classified as low-risk. We analyzed the validity of this modification using data from the single-institution 442 patient subset that stratified patients as pT3a versus pT3b. Using Kaplan-Meier techniques and the NRG modification, the estimated two and three-year LF rates remain significantly different for intermediate-risk [19% (95% CI 14–25%) and 26% (95% CI 19–33%), respectively] and high-risk disease [34% (95% CI 27–41%) and 46% (95% CI 37–54%), respectively] ($p<0.01$) (Figure 3A). Both two and three-year LF rates for the 35 patients in the original model with pT3a disease who were excluded from the NRG-modified intermediate-risk subgroup are 12% which is significantly smaller than the LF rates in the NRG-modified intermediate-risk group ($p<0.01$), thereby justifying the exclusion of pT3a patients from the model while continuing to show that our model stratifies patients into distinct risk groups even when subgroup criteria are modified (Figure 3B).

To assess the expected LF rate in the control arm of an adj-RT trial, we accounted for trial accrual bias by applying the more stringent inclusion criteria to our model's original database and then analyzed the remaining patients using Kaplan-Meier techniques. This analysis confirmed that the three subgroups still had significantly different three-year LF

rates of 8%, 22%, and 50%, respectively ($p<0.001$), which could serve as the LF estimates to be expected in the absence of adjuvant radiation (Figure 4/Table 2).

In summary, the sources of bias we examined did not invalidate the LF risk stratification or substantially reduce the model's LF estimates.

Acknowledgments

Funding sources: This investigation was supported in part by National Institutes of Health (NIH)/National Cancer Institute (NCI)/National Clinical Trials Network (NCTN) grants CA180888 and CA180819

References

1. Grossman HB, Natale RB, Tangen CM, et al. Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. *N Engl J Med*. 2003; 349:859–866. [PubMed: 12944571]
2. Christodouleas JP, Baumann BC, He J, et al. Optimizing bladder cancer locoregional failure risk stratification after radical cystectomy using SWOG 8710. *Cancer*. 2014; 120:1272–1280. [PubMed: 24390799]
3. Baumann BC, Guzzo TJ, He J, et al. A novel risk stratification to predict local-regional failures in urothelial carcinoma of the bladder after radical cystectomy. *Int J Radiat Oncol Biol Phys*. 2013; 85:81–88. [PubMed: 22543204]
4. Ku JH, Kim M, Jeong CW, et al. Risk prediction models of locoregional failure after radical cystectomy for urothelial carcinoma: external validation in a cohort of korean patients. *Int J Radiat Oncol Biol Phys*. 2014; 89:1032–1037. [PubMed: 25035206]
5. Froehner M, Novotny V, Wirth MP, et al. External validation of a model to predict locoregional failure after radical cystectomy. *Cancer*. 2014
6. Herr HW, Faulkner JR, Grossman HB, et al. Surgical factors influence bladder cancer outcomes: a cooperative group report. *J Clin Oncol*. 2004; 22:2781–2789. [PubMed: 15199091]

Patients at Risk					
High risk	100	6	3	1	0
Intermediate risk	125	28	18	5	2
Low risk	395	181	89	50	17

Figure 1.
Cumulative incidence of LF by subgroup, accounting for trial accrual bias

Patients at Risk					
High risk	138	6	3	1	0
Intermediate risk	153	28	18	5	2
Low risk	415	181	89	50	17

Figure 2.
Local recurrence-free survival (LRFS) by subgroup (Kaplan-Meier)

A

Patients at Risk				
High risk	92	4	1	0
Intermediate risk	79	11	5	1

B

Patients at Risk				
High risk	92	4	1	0
Intermediate risk	79	11	5	1
Excluded from NRG	35	7	5	0

Figure 3.

(A) LRFS using the modified NRG risk stratification (XXX cohort). (B) LRFS for the intermediate-risk pT3a patients excluded from NRG-GU001.

Patients at Risk					
High risk	100	6	3	1	0
Intermediate risk	125	28	18	5	2
Low risk	395	181	89	50	15

Figure 4.
LRFS by subgroup (Kaplan-Meier) after adjusting for trial accrual bias

Table 1

Competing risk multivariate analysis of LF as a function of year of cystectomy, controlling for LF risk group

Year of Cystectomy	Subhazard Ratio	95% CI	p-value
Penn	1.00	(0.96 – 1.04)	0.85
SWOG	0.99	(0.88 – 1.10)	0.81
Combined cohort	1.00	(0.97 – 1.03)	0.98

Table 2

Rates of LF by risk group in the combined cohort, both adjusted and unadjusted for trial accrual bias, using cumulative incidence and Kaplan-Meier techniques

Subgroup	3 year Cumulative incidence of Local Failure			3 year Kaplan-Meier Local Failure Rate		
	Local Failure Unadjusted	95% CI	Local Failure - Adjusted	95% CI	Local Failure - Adjusted	95% CI
Low	7%	5 – 10%	7%	7 – 10%	8%	7 – 8%
Intermediate	18%	12 – 25%	17%	11 – 21%	22%	17 – 26%
High	40%	31 – 48%	36%	26 – 46%	50%	43 – 57%