Published in final edited form as:

J Urol. 2010 March ; 183(3): 977–982. doi:10.1016/j.juro.2009.11.015.

Variations among experienced surgeons in cancer control after open radical prostatectomy

Fernando J. Bianco Jr., MD, Andrew J. Vickers, DPhil, Angel M. Cronin, MS, Eric A. Klein, MD, James A. Eastham, MD, J. Edson Pontes, MD, and Peter T. Scardino, MD

Departments of Surgery, Division of Urology (FJB, JAE, PTS) and Epidemiology and Biostatistics (AJV, AMC), Memorial Sloan-Kettering Cancer Center, New York, NY; Glickman Urological and Kidney Institute (EAK), Cleveland Clinic, Cleveland, OH; Department of Urology (JEP), Wayne State University, Detroit, MI

Abstract

Purpose—It has been demonstrated that complications and functional outcomes after prostate surgery vary between different surgeons to a greater extent than might be accounted for by chance. This type of excessive variation is known as “heterogeneity.” In this study, we explored whether there is also heterogeneity among high-volume surgeons with respect to cancer control after surgery.

Patients and Methods—The study cohort consisted of 7,725 patients with clinically localized prostate cancer treated by open radical prostatectomy at four major US academic medical centers 1987 – 2003 by one of 54 surgeons. We defined biochemical recurrence by a serum PSA level $\geq$ 0.4 ng/mL followed by a subsequent higher PSA level. Multivariable random effects models were used to evaluate the heterogeneity in prostate cancer recurrence between surgeons, after adjustment for case-mix (PSA, pathological stage and grade), year of surgery and surgeon experience.

Results—We found statistically significant heterogeneity in prostate cancer recurrence rates ($p=0.002$) independent of surgeon experience. Seven experienced surgeons in our series had adjusted five-year prostate cancer recurrence rates less than 10%, while another five experienced surgeons had rates that exceed 25%. Significant heterogeneity remained in sensitivity analyses adjusting for possible differences in follow-up, patient selection and stage migration.

Conclusions—A patient’s risk of recurrence may differ depending on which of two surgeons he sees, even if they have similar levels of experience. Surgical randomized trials are imperative to determine and characterize the roots of these variations

Keywords

prostatic neoplasms; surgery; prostatectomy outcomes; recurrence. Surgeon; Experience; Volume. Variability; Heterogeneity; Differences

Corresponding author: Fernando J. Bianco Jr., M.D., Consultant, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY.

Current Address:

Columbia University Division of Urology at Mount Sinai Medical Center, 4302 Alton Road, Suite 540, Miami Beach, FL 33140, Tel: (305) 606-1761, Fax: (305) 674-2899, fb2271@columbia.edu

Drs. Bianco and Vickers and would like to be considered as joint first authors on this paper.

Parts of this study were presented during the 2007 Meeting of the American Urological Association (AUA) and was awarded as the Best Clinical Research Paper by the AUA and the Society of Urologic Oncology

Introduction

Radical prostatectomy is a mainstay of treatment for localized prostate cancer. The procedure has been demonstrated in a randomized clinical trial to improve both overall and cancer-specific survival in men with intermediate- to high-risk prostate cancer.¹ Nonetheless, given the high degree of skill required for RP, it is plausible that cancer control outcomes may vary between different surgeons.

We distinguish two types of variation. The first is associated with readily identifiable characteristics of a surgeon, such as yearly caseload (“surgeon volume”), total lifetime surgical experience, and the number of cases treated at the hospital where the surgeon practices (“hospital volume”). Evidence of this type of variation is provided by studies that show an association between surgeon characteristics and patient outcomes.^{2–3}

The second type of variation is related to unmeasured differences in approach (the expected surgical plan) and technique (how a particular step in the operation is executed). These aspects of approach and technique associated with surgical success are often unknown and are rarely documented adequately. Accordingly, evidence that differences in approach and technique affect outcome must be indirect: the degree of variation in outcome between different surgeons is compared to the degree of variation expected by chance; if variation is higher than expected, “heterogeneity” is reported, and it may be concluded that unmeasured differences between surgeons are responsible for the differences in observed results.

Several studies have investigated heterogeneity in outcome between surgeons. In a study of RP, 8% of high-volume surgeons had postoperative complication rates above the predicted 99th percentile, whereas 3% had rates, below the first percentile. This heterogeneity in morbidity outcomes was not explained by chance and suggests differences in surgical execution.^{4,5}

The possibility and degree of heterogeneity in cancer control outcomes has not previously been investigated. This sort of heterogeneity has particular relevance to patient care. For example, we have reported that sufficiently experienced surgeons have near-zero recurrence rates for organ-confined prostate cancer.⁶ Evidence for or against heterogeneity would be informative as to whether this effect could be generalized to all surgeons with high levels of experience, or whether it was specific to the highly experienced surgeons in our data set. In this study, we evaluate whether there is heterogeneity in recurrence rates after radical prostatectomy. This is a good model to study heterogeneity because the outcome measure - BCR - is objective and well standardized. In our analyses we controlled for case mix, year of surgery, and surgeon experience; the question under study was whether patients can expect similar outcomes from surgeons with similar levels of experience.

Patients and Methods

Sources of data and study design

Our study cohort has been described previously.³ In brief, a total of 7,765 treatment naïve prostate cancer patients underwent open radical retropubic prostatectomy between January 1987 and December 2003 by one of 72 surgeons at one of four participating institutions: Memorial Sloan-Kettering Cancer Center (New York, NY), Baylor College of Medicine (Houston, TX), Wayne State University Harper University Hospital (Detroit, MI), and the Cleveland Clinic (Cleveland, OH). A surgeon's first radical prostatectomy was defined as the initial case performed after completing training in a urologic residency program accredited by the Accreditation Council for Graduate Medical Education. For surgeons that pursued further training, the first case was defined in terms of completion of fellowship

training. Surgeons whose RP experience began at a non-study institution were asked to provide their prior caseload. Surgeons with less than five cases in the cohort were excluded. All information was obtained with appropriate institutional review board waivers, and data were de-identified before analysis.

Outcome Measure

Follow-up visits were in compliance with the standard-of-care clinical practice at each institution's clinics. This consisted of serum PSA measurements every 3–4 months during the first postoperative year, semiannually the second year, and annually thereafter. Digital rectal examinations were performed annually, or if there was evidence of a rise in PSA. For our outcome measure we defined prostate cancer recurrence as a serum PSA of more than 0.4 ng/mL followed by a subsequent higher PSA level (ie, BCR).⁷ In rare cases (eg, <1% in the Memorial Sloan-Kettering Cancer Center data set), secondary treatment was initiated for patients who did not meet the strict criteria for recurrence yet had a rising PSA; such treatment was counted as an event

Statistical methods

Our statistical methods followed those of our prior paper on the learning curve.³ In brief, we created a multivariable, parametric random-effects regression survival-time model, using a log-logistic survival distribution to model hazard over time. We adjusted for case mix by including PSA, stage, grade, year of surgery and surgeon experience as covariates in the model. For each patient, “surgeon experience” was coded as the number of RPs conducted by the surgeon prior to the patient's operation and was calculated directly from the data set. Since the relationship between surgeon experience and BCR is nonlinear we used restricted cubic splines with knots at the quartiles. We did not cluster by institution because there is no plausible mechanism for how an institution could modify BCR rates independent of a surgeon.

To test for heterogeneity between surgeons, a random effect was included in the model for each surgeon to allow for patients treated by that surgeon to be at a higher or lower likelihood of subsequent recurrence. The random effects were assumed to follow an inverse Gaussian distribution, and the variance of the random effects was estimated to evaluate its departure from zero: if the random effects variance is significantly different from zero, this would suggest that there is heterogeneity between surgeons that cannot be explained by the covariates in the model.

The overall objective of this study was to determine whether oncologic outcome varies between surgeons due to differences in surgical technique. As such, we did not include margin status in our initial model because margins depend on technique^{8,9} it is reasonable to suppose that a surgeon with poor technique will have a higher positive margin rate and more recurrences; controlling away the difference in margin rates might lead us to a false conclusion of no differences in recurrence. To determine whether differences in technique could affect outcome beyond achieving clear margins we performed an additional analysis including surgical margin status as a covariate in the multivariable model. Statistical analyses were performed using Stata 9.2 (Stata Corp., College Station, TX).

Results

A total of 7,725 patients treated by one of 54 surgeons met eligibility criteria. Clinical and pathological characteristics are shown in Table 1, separately by institution. We observed 1,247 BCR events. Median follow-up for patients without BCR was four years. From the multivariable random effects model, there was significant heterogeneity in prostate cancer

recurrence between surgeons (random effects variance 0.050; 95% C.I. 0.014, 0.173; $P = 0.002$). Figure 1 shows the adjusted five-year probability of freedom from BCR by surgeon, after adjustment for lifetime experience, case mix (EPE, SVI, LNI), and year of surgery. For ease of visual interpretation, the figure shows only those surgeons who performed 40 or more total cases. The degree of variation is of clear clinical relevance: seven surgeons in our series had adjusted five-year BCR rates less than 10%, while five had rates in excess of 25%. Furthermore, the 95% confidence interval for 15 surgeons was completely outside the mean.

Significant heterogeneity in BCR rates remained after adjusting for margin status (figure 2; $P = 0.001$; random effects variance = 0.063). This suggests that differences in outcome between surgeons go over and above removing all tissue positive for cancer by gross pathology.

We conducted various sensitivity analyses to check the robustness of our findings. In our main analyses, we included all surgeons regardless of the total number of cases they had performed. It is plausible that some surgeons have poor results and are unable to develop a practice, and that these surgeons are responsible for the heterogeneity in our data set. We therefore performed three separate analyses restricting the data set to surgeons with at least 40, 100, or 250 total cases, resulting in the inclusion of 36, 16, and 10 surgeons, respectively. One possible cause of heterogeneity is difference in intensity of follow-up: if some surgeons followed their patients more regularly than others, even if underlying event rates were similar heterogeneity might be observed. To adjust for this effect, we converted our time-to-event data into the binary variable of recurrence at three years, excluding patients censored before this time. The results of our sensitivity analyses are shown in Table 2.

We conducted an additional sensitivity analysis related to stage migration shown in Table 3. To determine whether stage shift might be responsible for our results. We simulated the effects of stage shift more aggressively by including nonlinear terms for year and interaction terms between year and Gleason score, and between year and pathologic stage. The heterogeneity found was statistically significant. We conducted a sensitivity analysis restricted to patients treated after 1995 based on our prior observation³ that by that year the stage shift was largely accounted for. We found no statistically significant heterogeneity ($P = 0.3$). To determine whether this result might be associated with lower power due to the reduction in sample size, we conducted a simulation study in which we randomly deleted patients from the data set to create a cohort of a similar size to the post-1995 group. Heterogeneity was statistically significant in approximately 45% of these simulations. In addition, we performed a sub-analysis controlling for men with tumors pathologically confined to the prostate ($n = 5,316$ with 402 events) contemplating that selection bias would be negligible as this represents a more homogenous group of patients with the best prognosis. In this sensitivity analysis of pT2 patients we found significant variability among patients risk of recurrence attributable to surgeon variations – p value for heterogeneity = 0.004. We concluded that although stage migration may explain some of the heterogeneity observed in our main analysis, the preponderance of evidence supports clinically and statistically significant heterogeneity in prostate cancer control outcomes between surgeons.

Discussion

We have previously reported an important association between surgical experience and cancer control after radical prostatectomy, which we described in terms of a surgical learning curve.³ We have also analyzed the learning curve separately by pathologic stage, and reported that recurrences are extremely rare in patients with organ-confined disease

treated by the most experienced surgeons.⁶ These findings might suggest that experience is deterministic of outcome, such that, the results of an inexperienced surgeon are never superior to those of a surgeon with greater experience. Here we report statistically significant heterogeneity in oncologic outcome after adjusting for both tumor characteristics and surgeon experience. Thus, a patient's risk of BCR may differ depending on which of two surgeons he sees, even if they have similar levels of experience. We conclude that the oncologic results of radical prostatectomy can vary by both measured and unmeasured characteristics of the treating surgeon.

Although there have been previous reports showing heterogeneity for surgical margins⁹, this is the first large-scale demonstration that recurrence after radical prostatectomy varies between surgeons, even when taking into consideration surgeon experience. For 15 (42%) of the 36 surgeons with total lifetimes experience of 40 cases or more, the 95% confidence intervals for BCR rate excluded the mean for the group as a whole. Moreover, the extent of this variability is of clear relevance: seven surgeons in our series had adjusted BCR rates of less than 10%, while five had BCR rates in excess of 25%. A 15% absolute difference in BCR rates is higher than that observed for adjuvant chemotherapy for colon¹⁰ or breast cancer¹¹ and comparable to that associated with adjuvant radiation therapy after radical prostatectomy in high-risk patients.¹²

Our results have implications for both research and clinical practice. Heterogeneity between surgeons suggests that any randomized trial involving radical prostatectomy—whether the procedure itself or with adjuvant therapy—should stratify by surgeon, on the grounds that a chance imbalance between groups of surgeons may accentuate or attenuate differences between groups. Heterogeneity might also become the focus of research: what aspects of surgical technique explain the different results between surgeons? Systematic research is required to identify the critical aspects of radical prostatectomy that are associated with cancer control. It is known that patients may recur even if they have negative surgical margins¹³; our results suggest that whatever the mechanism, the surgeon may play a role. Evaluation of the expected surgical plan and how it was executed may be an initial step. With respect to clinical practice, it is clear that regionalization of cancer care, such that a higher proportion of patients are treated by surgeons with greater experience, is insufficient to guarantee optimal outcomes after radical prostatectomy. Undesirable variability in surgical outcomes could be addressed by continuing practical education—eg, by using surgical simulators or having surgeons observe each other's operations in person. The use of video has proven pivotal to improving outcomes¹⁴; thus increased exposure via video to prostatectomy procedures—including laparoscopic and robotically assisted—with enhanced magnification and controlled homeostasis should serve as fertile ground. There may also be a role for individualized feedback between peers, so that surgeons can monitor their execution and optimize their results.

There are several limitations associated with the retrospective, observational nature of our data. First, differences between surgeons may result from confounding by unmeasured differences in case mix. Yet it seems unlikely that such differences would completely explain our results, given the large effect sizes we noted, such as a difference between <10% and >25% adjusted BCR rates. While one may speculate about patient factors that vary between surgeons—such as screening history, socioeconomic status, race, body mass index, or age—it is difficult to believe that such factors would change recurrence rates two- or threefold. Second, we used a time-to-event endpoint to model to our data, and this can be influenced by intensity of follow-up. However, a sensitivity analysis using recurrence as a binary variable failed to affect our findings. Institutional differences are also an unlikely explanation for heterogeneity: despite the loss in power associated with subgroup analysis, we did find heterogeneity between surgeons in one of the three institutions studied. Third,

our data set includes patients treated during the period of stage migration in prostate cancer. Evidence of heterogeneity remained after careful statistical adjustment for stage migration: notwithstanding, we cannot definitively exclude the possibility that stage migration might explain our results, yet heterogeneity was particularly conspicuous in patients with organ confined prostate cancer.

In sum, heterogeneity in medical outcomes is undesirable and suggests that some patients are experiencing less than optimal outcomes. It is incumbent upon urologic surgeons to address the research, educational, and clinical issues raised by heterogeneity in order to uniformly deliver high-quality care to patients with localized prostate cancer.

Acknowledgments

This research was funded in part by a P50-CA92629 SPORE grant from the National Cancer Institute and by the Allbritton Fund and the Koch Foundation. F.J.B is supported in part by the American Urological Association Foundation and a training grant (T32-82088) from the National Institutes of Health.

References

1. Bill-Axelson A, Holmberg L, Ruutu M, Haggman M, Andersson SO, Bratell S, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med*. 2005; 352:1977. [PubMed: 15888698]
2. Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. *N Engl J Med*. 2003; 349:2117. [PubMed: 14645640]
3. Vickers AJ, Bianco FJ, Serio AM, Eastham JA, Schrag D, Klein EA, et al. The surgical learning curve for prostate cancer control after radical prostatectomy. *J Natl Cancer Inst*. 2007; 99:1171. [PubMed: 17652279]
4. Begg CB, Riedel ER, Bach PB, Kattan MW, Schrag D, Warren JL, et al. Variations in morbidity after radical prostatectomy. *N Engl J Med*. 2002; 346:1138. [PubMed: 11948274]
5. Bianco FJ Jr, Riedel ER, Begg CB, Kattan MW, Scardino PT. Variations among high volume surgeons in the rate of complications after radical prostatectomy: further evidence that technique matters. *J Urol*. 2005; 173:2099. [PubMed: 15879851]
6. Vickers AJ, Bianco FJ, Gonen M, Cronin AM, Eastham JA, Schrag D, et al. Effects of pathologic stage on the learning curve for radical prostatectomy: evidence that recurrence in organ-confined cancer is largely related to inadequate surgical technique. *Eur Urol*. 2008; 53:960. [PubMed: 18207316]
7. Stephenson AJ, Kattan MW, Eastham JA, Dotan ZA, Bianco FJ Jr, Lilja H, et al. Defining biochemical recurrence of prostate cancer after radical prostatectomy: a proposal for a standardized definition. *J Clin Oncol*. 2006; 24:3973. [PubMed: 16921049]
8. Chun FK, Briganti A, Antebi E, Graefen M, Currlin E, Steuber T, et al. Surgical volume is related to the rate of positive surgical margins at radical prostatectomy in European patients. *BJU Int*. 2006; 98:1204. [PubMed: 17034506]
9. Eastham JA, Kattan MW, Riedel E, Begg CB, Wheeler TM, Gerigk C, et al. Variations among individual surgeons in the rate of positive surgical margins in radical prostatectomy specimens. *J Urol*. 2003; 170:2292. [PubMed: 14634399]
10. Gill S, Loprinzi CL, Sargent DJ, Thome SD, Alberts SR, Haller DG, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *J Clin Oncol*. 2004; 22:1797. [PubMed: 15067028]
11. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005; 365:1687. [PubMed: 15894097]
12. Bolla M, van Poppel H, Collette L, van Cangh P, Vekemans K, Da Pozzo L, et al. Postoperative radiotherapy after radical prostatectomy: a randomised controlled trial (EORTC trial 22911). *Lancet*. 2005; 366:572. [PubMed: 16099293]

13. Swindle P, Eastham JA, Ohori M, Kattan MW, Wheeler T, Maru N, et al. Do margins matter? The prognostic significance of positive surgical margins in radical prostatectomy specimens. *J Urol*. 2005; 174:903. [PubMed: 16093984]
14. Walsh PC, Marschke P, Ricker D, Burnett AL. Use of intraoperative video documentation to improve sexual function after radical retropubic prostatectomy. *Urology*. 2000; 55:62. [PubMed: 10654896]

Figure 1.

Forest plot of 5-year predicted probability of freedom from recurrence by surgeon. The probabilities are for a patient with the mean level of all covariates (PSA, Gleason score, EPE, SVI, LNI, year of surgery) treated when the surgeon a minimum of 40 prior cases. The vertical line represents the mean adjusted 5-year probability of freedom from biochemical recurrence among all surgeons.

Figure 2.

Forest plot of 5-year predicted probability of freedom from recurrence by surgeon. The probabilities are for a patient with the mean level of all covariates plus surgical margin status treated when the surgeon a minimum of 40 prior cases. The vertical line represents the mean adjusted 5-year probability of freedom from biochemical recurrence among all surgeons.

Table 1

Clinical and Pathological Patient Characteristics by Institution*

	Number of Patients	Institution		
		CCF	MSKCC ⁺	WSU
		1853	4168	1704
Clinical patient characteristics	Preoperative PSA (ng/ml)	6.20 (4.70, 9.00)	6.89 (4.9, 10.6)	6.80 (5.09, 10.1)
	Age at RP (years)	61 (56, 66)	61 (56, 65)	63 (56, 67)
	Pathological Gleason grade			
	≤ 5	64 (3%)	232 (6%)	132 (8%)
	6	637 (34%)	1819 (44%)	564 (33%)
	7	1054 (57%)	1831 (44%)	892 (52%)
	8	69 (4%)	177 (4%)	101 (6%)
	≥ 9	29 (2%)	109 (3%)	15 (1%)
	Extracapsular extension (ECE)	627 (34%)	1213 (29%)	408 (24%)
	Seminal vesicle invasion (SVI)	148 (8%)	353 (8%)	190 (11%)
	Lymph node involvement (LNI)	43 (2%)	176 (4%)	70 (4%)
	Non-organ-confined cancer (any of ECE, SVI, or LNI)	646 (35%)	1309 (31%)	454 (27%)
Surgeon characteristics	Number of surgeons	18	12	24
	Total number of cases performed by surgeon			
	<40	9 (50%)	0 (0%)	8 (33%)
	40–99	6 (33%)	3 (25%)	12 (50%)
	100–249	1 (6%)	3 (25%)	2 (8%)
	≥ 250	2 (11%)	6 (50%)	2 (8%)
Time frame	Number of operations performed by year			
	1987–1990	64 (3%)	331 (8%)	37 (2%)
	1991–1995	392 (21%)	1076 (26%)	746 (44%)
	1996–2000	743 (40%)	1447 (35%)	804 (47%)
	2001–2003	654 (35%)	1314 (32%)	117 (7%)
Unadjusted	Positive surgical margins	514 (28%)	871 (21%)	653 (38%)

	Institution		
	CCF	MSKCC ⁺	WSU
Number of Patients			
	1853	4168	1704
Outcomes	20%	18%	22%
	Five-year recurrence-free probability		

* Data are median (interquartile range) or frequency (percent).

⁺ Includes 1 surgeon practiced at Baylor College of Medicine and MSKCC.

Abbreviations: CCF: Cleveland Clinic Foundation; MSKCC: Memorial Sloan-Kettering Cancer Center; WSU: Wayne State University.

Table 2

Sensitivity Analyses

	<i>P</i> Value for Heterogeneity	Random Effects Variance (95% CI)
Main analysis	0.002	0.050 (0.014, 0.173)
Surgeons who built up a practice		
Include only surgeons with ≥ 40 total surgeries	0.002	0.055 (0.015, 0.202)
Include only surgeons with ≥ 100 total surgeries	0.028	0.026 (0.004, 0.174)
Include only surgeons with ≥ 250 total surgeries	0.045	0.018 (0.002, 0.129)
Include surgical margin status in multivariable model	0.001	0.063 (0.019, 0.209)
Recurrence as binary variable to adjust for differences in intensity of follow-up	0.002	0.043 (0.014, 0.128)

Table 3**Sensitivity Analyses Related to Stage Migration**

	<i>P</i> Value for Heterogeneity	Random Effects Variance (95% CI)
Main analysis	0.002	0.050 (0.014, 0.173)
Patients treated after 1995	0.3	0.016 (0.000, 0.598)
Additional modeling for year of surgery		
Splines for year of surgery	0.018	0.030 (0.006, 0.146)
+ Interaction between year and stage	0.039	0.026 (0.005, 0.143)
+ Interaction between year and grade	0.026	0.027 (0.005, 0.140)
+ Both interactions	0.051	0.023 (0.004, 0.141)
Only the interactions (no splines)	0.015	0.036 (0.009, 0.147)
Restricted to patients with pathologically organ confined (pT2) prostate cancers (n=5316)	0.004	0.129 (0.031, 0.541)