Cancer Control and Functional Outcomes After Radical Prostatectomy as Markers of Surgical Quality: Analysis of Heterogeneity Between Surgeons at a Single Cancer Center

Andrew Vickersa,*, Caroline Savagea, Fernando Biancob, John Mulhalla, Jaspreet Sandhu, Bertrand Guillonneau, Angel Cronind, Peter Scardinoc

a Department of Surgery and Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
b Columbia University Division of Urology at Mount Sinai Medical Center, New York, NY, USA
c Department of Surgery and Department of Urology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
d Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

Abstract

Background: Previous studies have shown that complications and biochemical recurrence rates after radical prostatectomy (RP) vary between different surgeons to a greater extent than might be expected by chance. Data on urinary and erectile outcomes, however, are lacking.

Objective: In this study, we examined whether between-surgeon variation, known as heterogeneity, exists for urinary and erectile outcomes after RP.

Design, setting, and participants: Our study consisted of 1910 RP patients who were treated by 1 of 11 surgeons between January 1999 and July 2007.

Intervention: All patients underwent RP at Memorial Sloan-Kettering Cancer Center.

Measurements: Patients were evaluated for functional outcome 1 yr after surgery. Multivariable random effects models were used to evaluate the heterogeneity in erectile or urinary outcome between surgeons, after adjustment for case mix (age, prostate-specific antigen, pathologic stage and grade, comorbidities) and year of surgery.

Results and limitations: We found significant heterogeneity in functional outcomes after RP ($p < 0.001$ for both urinary and erectile function). Four surgeons had adjusted rates of full continence $< 75\%$, whereas three had rates $> 85\%$. For erectile function, two surgeons in our series had adjusted rates $< 20\%$; another two had rates $> 45\%$. We found some evidence suggesting that surgeons' erectile and urinary outcomes were correlated. Contrary to the hypothesis that surgeons “trade off” functional outcomes and cancer control, better rates of functional preservation were associated with lower biochemical recurrence rates.

Conclusions: A patient’s likelihood of recovering erectile and urinary function may differ depending on which of two surgeons performs his RP. Functional preservation does not appear to come at the expense of cancer control; rather, both are related to surgical quality.

© 2010 European Association of Urology. Published by Elsevier B.V. All rights reserved.

* Corresponding author. Health Outcomes Research Group, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA. Tel. +1 646 735 8142; Fax: +1 646 735 0011. E-mail address: vickersa@mskcc.org (A. Vickers).
1. Introduction

Radical prostatectomy (RP) is widely used to treat localized prostate cancer. In a study of unscreened men, RP was shown to improve both overall and cancer-specific survival in comparison with conservative management [1]. However, the operation is also associated with the risk of long-term complications, such as incontinence and erectile dysfunction [2], which can have a profound impact on patients' quality of life.

We have previously shown that cancer control after RP varies between different surgeons to a greater extent than can be accounted for by chance [3]. This type of variation, referred to as heterogeneity, was found even among the most experienced surgeons: Seven had adjusted 5-yr biochemical recurrence rates <10%, and another five had rates >25%. Other investigators have found similarly wide surgeon-to-surgeon variation in complication rates: Begg et al reported that 8% of high-volume surgeons had complications rates following RP above the predicted 99th percentile, whereas 3% had complication rates below the 1st percentile [4].

Given the evidence of heterogeneity in cancer control and complication rates, it is plausible that variations in surgical execution would have a substantial effect on the rates of erectile and urinary dysfunction after RP. On the basis of a cohort accrued at a single academic center, we evaluated whether there is heterogeneity between surgeons for urinary and erectile function following RP, adjusting for case mix and year of surgery. We tested two additional hypotheses. First, we hypothesized that surgeons' urinary and erectile function rates would be positively correlated, indicating they differ in their ability to remove the prostate without damaging the surrounding tissue structures. Second, we hypothesized that surgeons differ in how they trade off cancer control and functional preservation, leading to a negative correlation between functional outcomes and cure rates.

2. Materials and methods

2.1. Patients

A total of 2593 patients were treated by nonsalvage open or laparoscopic RP between January 1999 and July 2007 and were without either urinary or erectile dysfunction at baseline (defined in section 2.2). Patient function was assessed at routine clinical assessments every few months after surgery, leading to interval censoring. Accordingly, we converted the time-to-event data into binary variables for the outcome of 1-yr function. In the main analysis, patients who regained function before 14 mo after surgery (1 yr plus a 2-mo window) were considered to have regained function; patients who were not functional after 10 mo (1 yr minus a 2-mo window) and who did not regain function between 10 and 14 mo were considered nonfunctional. Patients without eligible data (n = 1170) were excluded from the main analyses. As a sensitivity analysis, we included all patients with follow-up data, using the assessment closest to 1 yr. Data were obtained under a waiver from the institutional review board.

2.2. Outcome measures

Functional outcomes were assessed by the treating surgeon and graded on a 5-point scale (Table 1). Our key concern was that reported differences in functional outcomes between surgeons might result from differences in grading by surgeons rather than from differences in patient outcomes. We used definitions of surgeon grading for each outcome that we showed previously to correlate strongly with patient outcome [5,6]: Erectile function was defined as an erectile rigidity score of 1 or 2, with urinary function defined as a score of 1. Moreover, we planned to test our findings in a separate group of patients who had completed functional outcome questionnaires. Biochemical recurrence was defined as postoperative prostate-specific antigen (PSA) value ≥0.10 ng/ml and a confirmatory rise or as salvage treatment.

2.3. Statistical methods

For the end points of erectile and urinary function, we created logistic regression models adjusting for case mix by including the following predictors: age, PSA, extracapsular extension, lymph node involvement, seminal vesicle invasion, pathologic Gleason grade (≤6, 7, or ≥8), comorbidities (0, 1, 2, or ≥3), and year of surgery. We used a fixed effect, parametric random effects regression survival-time model, using a log-logistic survival distribution to model biochemical recurrence rates [7]; covariates were the same as for functional outcomes, except that comorbidities and age were not included.

To determine whether there were differences between surgeons in measured patient characteristics, something that might indicate difference in unmeasured confounders, we tested whether surgeon, entered as a categorical variable in a fixed effects model, was significantly associated with our chosen predictors. Because different surgeons operated in different years and there are important changes in patient characteristics over time, we adjusted these analyses for year of surgery.

To test for heterogeneity between surgeons, a random effect following an inverse Gaussian distribution 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 erectile or urinary function at 1 yr. To produce a forest plot, we used fixed effects models, calculating the predicted probability of each outcome separately for each surgeon using the mean value for covariates. Spearman rank correlation was calculated for the adjusted rates of urinary and erectile function for each surgeon, and for the adjusted rates of functional preservation (return of both urinary and erectile function) and biochemical recurrence. Lastly, to test for an association between a surgeon's annual volume and both functional preservation and biochemical recurrence, volume was entered into the respective fixed effects regression model using restricted cubic splines with knots at the

<table>
<thead>
<tr>
<th>Score</th>
<th>Continence: clinical meaning</th>
<th>Erectile function: clinical meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No pads</td>
<td>Normal full erections</td>
</tr>
<tr>
<td>2</td>
<td>One pad daily</td>
<td>Full but diminished erections satisfactory for sexual activity</td>
</tr>
<tr>
<td>3</td>
<td>Two or three pads daily</td>
<td>Partial erections occasionally satisfactory for sexual activity</td>
</tr>
<tr>
<td>4</td>
<td>Four or more pads daily</td>
<td>Partial erections unsatisfactory for sexual activity</td>
</tr>
<tr>
<td>5</td>
<td>Complete incontinence</td>
<td>No erections</td>
</tr>
</tbody>
</table>
3. Results

Patient characteristics, shown in Table 2, were similar between patients who did and did not provide functional outcome data. The exception is there are more missing data for patients treated laparoscopically, a difference that can be addressed in a subgroup analysis including only open surgeons. Small nonsignificant differences between surgeons were found with respect to follow-up, with surgeons who had better functional outcomes having a slightly higher tendency to follow patients less likely to regain function (Spearman correlation between adjusted rates of function and degree to which patients with less favorable characteristics features were preferentially followed up were, for urinary function: 0.37; \( p = 0.3 \); for erectile function: \(-0.05; p = 0.9\)). As such, differences between surgeon follow-up seem unlikely to explain any observed differences in functional outcomes. Furthermore, there were no significant differences in patient cancer risk or comorbidities by surgeon after adjustment for year of surgery (global \( p = 0.15 \) and \( p = 0.16 \), respectively). Any differences between surgeons regarding functional outcomes are therefore unlikely to be related to selection bias.

For the main analysis, 1423 patients were evaluated for urinary and sexual function at 1 yr; 97 were missing data on clinical variables or recurrence. Of the 1326 patients available for analysis, 43\% (\( n = 576 \)) of patients had recovery of erectile function, 83\% (\( n = 1100 \)) were fully continent, and 40\% (\( n = 529 \)) had returned to both erectile and urinary function; 124 patients experienced a biochemical recurrence after surgery. There was significant heterogeneity between surgeons for both functional outcomes: erectile function (random effects variance 0.048; 95% confidence interval [CI], 0.012–0.168; \( p < 0.001 \)) and urinary function (random effects variance 0.074; 95% CI, 0.024–0.211; \( p < 0.001 \)). Figs. 1 and 2 show forest plots of adjusted

Table 2 – Summary of patient characteristics*

| Variable                        | Followed for functional outcomes \( n = 1910 \) | Missing data on functional outcomes \( n = 531 \) or clinical covariates \( n = 152 \) | \( p \) value for difference*
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yr</td>
<td>58 (53, 63)</td>
<td>59 (54, 64)</td>
<td>0.004</td>
</tr>
<tr>
<td>PSA, ng/ml</td>
<td>5.4 (4.1, 7.4)</td>
<td>5.4 (4.1, 8.1)</td>
<td>0.5</td>
</tr>
<tr>
<td>Pathologic Gleason grade, %</td>
<td></td>
<td></td>
<td>0.2</td>
</tr>
<tr>
<td>≤6</td>
<td>797 (42)</td>
<td>255 (41)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1000 (52)</td>
<td>324 (52)</td>
<td></td>
</tr>
<tr>
<td>≥8</td>
<td>113 (6)</td>
<td>49 (8)</td>
<td></td>
</tr>
<tr>
<td>Nerve sparing, %</td>
<td></td>
<td></td>
<td>0.16</td>
</tr>
<tr>
<td>None</td>
<td>317 (17)</td>
<td>130 (19)</td>
<td></td>
</tr>
<tr>
<td>Unilateral</td>
<td>503 (26)</td>
<td>179 (26)</td>
<td></td>
</tr>
<tr>
<td>Bilateral</td>
<td>825 (43)</td>
<td>266 (39)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>265 (14)</td>
<td>108 (16)</td>
<td></td>
</tr>
<tr>
<td>Positive surgical margins, %</td>
<td>240 (13)</td>
<td>96 (14)</td>
<td>0.3</td>
</tr>
<tr>
<td>Seminal vesicle invasion, %</td>
<td>107 (6)</td>
<td>53 (8)</td>
<td>0.025</td>
</tr>
<tr>
<td>Lymph node invasion, %</td>
<td>75 (4)</td>
<td>34 (5)</td>
<td>0.3</td>
</tr>
<tr>
<td>Extracapsular extension, %</td>
<td>511 (27)</td>
<td>182 (29)</td>
<td>0.4</td>
</tr>
<tr>
<td>Open (vs laparoscopic) surgery</td>
<td>1522 (80)</td>
<td>425 (62)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No. of comorbidities, %</td>
<td></td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>0</td>
<td>964 (50)</td>
<td>380 (56)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>622 (33)</td>
<td>194 (28)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>260 (14)</td>
<td>90 (13)</td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>64 (3)</td>
<td>19 (3)</td>
<td></td>
</tr>
<tr>
<td>Unadjusted outcome, %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continent</td>
<td>1444 (76)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Erectile function</td>
<td>669 (35)</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

PSA = prostate-specific antigen.
* All values are median (interquartile range) or frequency (proportion).
† Obtained using Wilcoxon rank sum test for continuous variables or Fisher's exact test for categorical variables.
1-yr rates of urinary and erectile function, respectively, demonstrating wide variation in both functional outcomes.

We conducted several preplanned sensitivity analyses. Our findings were unaffected if we restricted the analysis to patients who were treated by open surgery (heterogeneity \( p \leq 0.001 \) for both sexual and urinary function) or if we stratified analysis by baseline potency grade (\( p \leq 0.001 \)). Restricting analysis to patients treated by the most experienced surgeons with \( \geq 500 \) prior operations, we also found significant heterogeneity for functional outcomes (\( p = 0.012 \) and \( p < 0.001 \) for sexual and urinary function, respectively). Lastly, when we included in the model nerve-sparing status, itself a marker of surgeon technique, we still observed significant heterogeneity for urinary function (0.106; 95% CI, 0.023–0.378; \( p < 0.001 \)) and sexual function (0.052; 95% CI, 0.004–0.398; \( p = 0.04 \)). This suggests that nerve damage varies by surgeon, even within categories of nerve sparing.

We found evidence to support an association between surgeons’ adjusted urinary and erectile outcomes, although the correlation did not reach statistical significance (Spearman \( \rho \): 0.43; 95% CI, –0.19 to 0.81; \( p = 0.16 \); Fig. 3). We found a statistically significant relationship between functional preservation and recurrence that was in the opposite direction to that we had hypothesized: Surgeons with better functional outcomes also had higher rates of cancer control (Spearman \( \rho \): 0.58; 95% CI, 0.01–0.87; \( p = 0.048 \); Fig. 4). When we expanded the cohort to include all patients with at least one follow-up assessment of functional outcomes (\( n = 1910 \)) but who were not evaluated within the eligible time frame, the results supported our main conclusions (Spearman \( \rho \): 0.53; 95% CI, –0.03 to 0.84; \( p = 0.06 \) for correlation between urinary and erectile outcomes and Spearman \( \rho \): 0.71; 95% CI, 0.26–0.91; \( p = 0.007 \) for the correlation between functional preservation and biochemical recurrence).

Lastly, we found an association between surgeons’ annual volumes and patient outcomes. Surgeons with higher volumes had significantly better functional preservation than those with lower volumes (\( p = 0.005 \)). For a patient with the mean level of all covariates, the predicted probability of experiencing recovery of both erectile and urinary function at 1 yr was 21% if treated by a surgeon with an annual volume of 25 cases; this probability increased to 47% if the surgeon had an annual volume of 100.

As a final check of our analysis, we used recently acquired patient-reported data from 1053 men who completed functional outcome questionnaires as part of a new initiative to obtain patient reports as a routine part of clinical care. We did not use these data in our main analysis because only a limited number of surgeons were represented and the patients were treated recently and thus did not have sufficient follow-up for the analysis of biochemical recurrence. Urinary function and sexual function were measured on a 0- to 21-point scale and a 0- to 30-point
scale, respectively, with higher scores indicating better function. We found that surgeons with rates of functional preservation above the median for the main analysis also had superior patient-reported outcomes for both sexual and urinary function (adjusted urinary and sexual function scores of 0.9 and 1.4 points higher; \( p = 0.003 \) and \( p = 0.03 \), respectively).

4. Discussion

We have previously shown heterogeneity between surgeons for rates of cancer control after RP. Here we show statistically significant between-surgeon variation for functional outcomes. The adjusted rates of good erectile function at 1 yr ranged from 8% to 49%, and rates of continence ranged from 64% to 97%. For 3 of 11 surgeons, the 95% CIs for continence rates excluded the mean for the group; CIs for erectile outcomes excluded the mean for the group for 2 of 11 surgeons. To our knowledge, this is the first study to show that erectile and urinary function after RP vary between surgeons and therefore that a patient’s risk of functional impairment can depend on which of two surgeons does his RP.

We also found some evidence that surgeons with superior continence rates also had better outcomes for erectile function. Critically, we found that improved function outcome did not come at the expense of cancer control; indeed, surgeons with higher rates of functional preservation also had lower rates of recurrence. Better surgeons are able to remove cancer and preserve function; less able surgeons compromise function without complete excision of all tumor cells.

Our study has several limitations. First, we were not able to adjust for each surgeon’s lifetime experience. To use experience as a covariate requires a good estimation of the association between experience and outcome. Such an estimate, the learning curve, depends on having near complete case histories for all surgeons; otherwise what is compared are differences between surgeons, not surgeon experience. Functional outcomes were not well recorded until recently, leading to missing data during the early part of most surgeons’ careers. We have previously shown that surgical experience is strongly associated with improved cancer control [7]. Other studies have found a strong association between surgeon volume, which is highly correlated with experience, and complication rates following RP [8,9]. Thus we suspect that experience likely influences both erectile and urinary outcomes after RP. That said, significant heterogeneity was observed for functional outcomes when the analysis was restricted to the most experienced surgeons. From a patient’s point of view it may not matter why two surgeons differ—i.e., is it because one is more experienced or the other has less inherent ability?—it matters only that they do. Second, any observational study is subject to unmeasured confounders, and it is possible that between-surgeon differences in unmeasured confounders might lead to differences in outcome. In particular, there may have been subtle differences in baseline function between patients given the same grade. We did not use baseline questionnaires of erectile function as we have concerns that asking patients about sexual activity shortly after a diagnosis of cancer is likely to be uninformative because anxiety interferes with sexual desire and performance. However, we found no differences between surgeons for any measured patient characteristics, and we carefully adjusted for major confounders in our statistical model. Furthermore, it seems unlikely that unmeasured confounders could explain the large differences we observed in this study, such as a continence rate >90% compared with one of only 65%.

The use of surgeon-reported outcomes is a clear limitation of our study and raises the possibility that we may be observing variation in outcome assessment rather than variation in outcome. Again, we find this an inadequate explanation of our results. It is implausible that differences in ascertainment could explain the very large differences we observed (e.g., a 30% absolute difference in urinary function). Moreover, when we assessed function directly reported by patients themselves, surgeons with better rates in our main analysis also had better patient-reported outcomes. The strong association we report between volume and functional outcome, comparable with what other workers have demonstrated for complication rates [4], also argues against an explanation of our findings in terms of biases in surgeon assessment, as does the positive correlation found between functional end point and recurrence, which is an objective criterion of surgical outcome.

As is common in longitudinal studies of functional outcomes, we had incomplete follow-up. Yet we found that surgeons with better functional outcomes preferentially followed up patients with less favorable characteristics, suggesting that bias is a highly unlikely explanation for the differences in outcome we observed between surgeons.

A final possible limitation is that our sample included only academic surgeons practicing at a single high-volume cancer center. We would counter that, conversely, this fact is a strength of our findings. If we were to see consistent results anywhere in RP, one would hope it would be within a single specialist institution. That we see large variation at our center suggests such variation is ubiquitous in urologic oncology.

These findings have several implications. First, the use of averages in patient counseling during initial decision making is problematic. For example, a patient might be told that if he undergoes RP, he has an 85% probability of being free of disease at 5 yr but runs a 20% risk of urinary problems and has a slightly less than even chance of recovering erectile function within a year of surgery. But as we have shown, the man’s true probability of any of these events varies widely depending on his surgeon. Second, our findings raise questions about studies comparing the outcomes between different surgical approaches. In a typical study, Ficarra et al reported a 9% absolute higher probability of continence at 1 yr for robotic as compared with open RP. The authors concluded that the robotic approach leads to improved urinary function [10]. But this difference between techniques is dwarfed by the 30% and
40% between-surgeon differences we report here. It is thus plausible that the results of Ficarra et al reflect not so much an inherent advantage of the robotic technique but differences in surgical skill between the two robotic versus four open surgeons who happened to take part in the study. Our results also have implications for research. There is a clear need for research both to identify specific aspects of technique associated with superior outcomes and to explore educational interventions and thus reduce variability in surgical outcomes.

5. Conclusions

Rates of urinary incontinence and erectile dysfunction vary widely between surgeons, suggesting that some patients are experiencing less than optimal outcomes. Strong correlations between functional and oncologic outcomes indicate that this variation results from differences in technique, rather than from differences in how surgeons trade off functional preservation and cancer control. Urologic surgeons need to address the research, educational, and clinical issues raised by heterogeneity to uniformly deliver high-quality care to patients with localized prostate cancer.

Author contributions: Andrew Vickers had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Vickers, Scardino.
Acquisition of data: Scardino, Bianco, Mulhall, Sandhu, Guillonneau.
Analysis and interpretation of data: Vickers, Savage, Cronin.
Drafting of the manuscript: Vickers, Savage.
Critical revision of the manuscript for important intellectual content: Bianco, Mulhall, Sandhu, Guillonneau, Scardino, Vickers.
Statistical analysis: Vickers, Savage, Cronin.
Obtaining funding: Vickers, Scardino.
Administrative, technical, or material support: Scardino.
Supervision: Vickers.
Other (specify): None.

Financial disclosures: I certify that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: Supported in part by funds from David H. Koch provided through the Prostate Cancer Foundation, the Sidney Kimmel Center for Prostate and Urologic Cancers, and a PS0-CA92629 SPORE grant from the National Cancer Institute to Peter Scardino.

References