Individualized Estimation of the Benefit of Radical Prostatectomy from the Scandinavian Prostate Cancer Group Randomized Trial

Andrew Vickersa,*, Caroline Bennetteb, Gunnar Steineckc, Hans-Olov Adami, Jan-Erik Johansson, Anna Bill-Axelson, Juni Palmgren, Hans Garmog, Lars Holmberg

*Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; bUniversity of Washington, Pharmaceutical Outcomes Research and Policy Program, Seattle, WA, USA; cKarolinska Institute, Stockholm, Sweden; dHarvard School of Public Health, Boston, MA, USA; eÖrebro University Hospital, Örebro, Sweden; fUniversity Hospital Uppsala, Uppsala, Sweden; gKing’s College London School of Medicine, London, UK, and Regional Oncologic Centre Uppsala/Orebro, Uppsala, Sweden; hKing’s College London School of Medicine, London, UK, and Regional Oncologic Centre Uppsala/Orebro, Uppsala, Sweden

Abstract

Background: Although there is randomized evidence that radical prostatectomy improves survival, there are few data on how benefit varies by baseline risk.

Objective: We aimed to create a statistical model to calculate the decrease in risk of death associated with surgery for an individual patient, using stage, grade, prostate-specific antigen, and age as predictors.

Design, setting, and participants: A total of 695 men with T1 or T2 prostate cancer participated in the Scandinavian Prostate Cancer Group 4 trial (SPCG-4).

Intervention: Patients in SPCG-4 were randomized to radical prostatectomy or conservative management.

Outcome measurements and statistical analysis: Competing risk models were created separately for the radical prostatectomy and the watchful waiting group, with the difference between model predictions constituting the estimated benefit for an individual patient.

Results and limitations: Individualized predictions of surgery benefit varied widely depending on age and tumor characteristics. At 65 yr of age, the absolute 10-yr risk reduction in prostate cancer mortality attributable to radical prostatectomy ranged from 4.5% to 17.2% for low- versus high-risk patients. Little expected benefit was associated with surgery much beyond age 70. Only about a quarter of men had an individualized benefit within even 50% of the mean. A limitation is that estimates from SPCG-4 have to be applied cautiously to contemporary patients.

Conclusions: Our model suggests that it is hard to justify surgery in patients with Gleason 6, T1 disease or in those patients much above 70 yr of age. Conversely, surgery seems unequivocally of benefit for patients who have Gleason 8, or Gleason 7, stage T2. For patients with Gleason 6 T2 and Gleason 7 T1, treatment is more of a judgment call, depending on patient preference and other clinical findings, such as the number of positive biopsy cores and comorbidities.

© 2012 European Association of Urology. Published by Elsevier B.V. All rights reserved.

* Corresponding author. Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA.
E-mail address: vickersa@mskcc.org (A. Vickers).
1. **Introduction**

Prostate cancer is the most commonly diagnosed cancer in men. This year, >200 000 new cases of prostate cancer will be diagnosed in US men, more than for lung and colorectal cancer combined [1]. Each of those men will have to make a difficult life-changing decision about their initial treatment.

There is lack of randomized trials comparing the initial treatment options for prostate cancer. The dearth of randomized data has led to a rather nihilistic approach to treatment recommendations. For example, a review on “Comparative Effectiveness and Harms of Treatments for Clinically Localized Prostate Cancer,” prepared for the Agency for Healthcare Research and Quality (AHRQ), concludes that “assessment of the comparative effectiveness and harms of localized prostate cancer treatments is difficult because of limitations in the evidence” [2]. An allied problem is the lack of individualized advice, critical in prostate cancer due to the heterogeneity of disease progression. As the AHRQ report puts it, “few high-quality data [are] available on the comparative effectiveness of treatments based on … PSA levels, histologic score, and [stage] to identify low-, intermediate-, and high-risk tumors” [2].

One randomized trial has reported a comparison between different initial treatments for localized prostate cancer [3]. The Scandinavian Prostate Cancer Group 4 (SPCG-4) trial included 695 men randomly assigned to radical prostatectomy or watchful waiting. After a median follow-up for survivors of 12 yr, radical prostatectomy reduced the risk of prostate cancer metastasis and death (relative risk: 0.65; p < 0.05 for both), with an absolute risk reduction of 6.7% and 5.4%, respectively. In this paper, we analyze data from SPCG-4 for the comparative effectiveness of treatments based on Gleason grade, stage, and PSA, based on an analysis of the entire cohort and then for men assigned to radical prostatectomy compared with watchful waiting by age, stratified by Gleason score and stage. PSA was not a strong predictor in the model, and the benefit of surgery was not importantly affected by PSA level. We chose 10 ng/ml to illustrate our results because it was close to the median.

Overall, younger men with more aggressive disease experienced a larger reduction in risk of prostate cancer death with radical prostatectomy than older men with lower risk cancer. For example, the 10-yr risk of death from prostate cancer decreased from 24% to 9% for a 60-yr-old man with Gleason 7, stage T2 disease (Table 2) but from 4% to 3% for a 70-yr-old with Gleason 6 T1. Little expected benefit was associated with surgery beyond age 70, even among those with aggressive disease (Fig. 1; Table 2). At younger ages, the hypothesis that the benefit of radical prostatectomy, in terms of absolute risk reduction, would be greater among men with a higher baseline risk of metastasis or death from prostate cancer, a traditional assumption in clinical epidemiology [9]. We assumed that, due to the effects of surgery on urinary and erectile function, men will not opt for surgery unless it offers an important risk reduction.

Our modeling approach was to create a risk score using Gleason grade, stage, and PSA, based on an analysis of the entire cohort and then include the risk score along with age (restricted cubic splines with knots at the tertiles) in separate models for each treatment group.

We examined both death from prostate cancer and prostate cancer metastasis following up to the end of 2006. Due to the high rate of death from other causes, especially at older ages, we chose to use competing risks regression methods. All analyses were conducted using Stata v.11.0 (Stata Corp., College Station, TX, USA).

2. **Methods**

The SPCG-4 trial was described previously [3]. In brief, men with T1 or T2 prostate cancer, prostate-specific antigen (PSA) <50 ng/ml, and negative bone scan were randomized to radical prostatectomy or watchful waiting, with blinded evaluation of cause of death. Surgery was aborted in the case of positive lymph nodes (7%) and, in contradistinction to contemporary “active surveillance” approaches [4,5], patients in the watchful waiting arm were not followed with a careful protocol of repeat biopsy, with a crossover to active treatment on evidence of disease progression. Approximately 15% of patients in the watchful waiting arm were not followed with a careful protocol of repeat biopsy, with a crossover to active treatment on evidence of disease progression. In comparison, rates of crossover to surgery or radiotherapy are in the order of 30–45% in active surveillance cohorts [6–8].

Our aim was to calculate individualized estimates of benefit. We hypothesized that the benefit of radical prostatectomy, in terms of absolute risk reduction, would be greater among men with a higher baseline risk of metastasis or death from prostate cancer, a traditional assumption in clinical epidemiology [9]. We assumed that, due to the effects of surgery on urinary and erectile function, men will not opt for surgery unless it offers an important risk reduction.

Our modeling approach was to create a risk score using Gleason grade, stage, and PSA, based on an analysis of the entire cohort and then include the risk score along with age (restricted cubic splines with knots at the tertiles) in separate models for each treatment group.

Table 1 – Patient characteristics

<table>
<thead>
<tr>
<th>Age, yr</th>
<th>SPCG-4 RP</th>
<th>SPCG-4 WW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 347</td>
<td>n = 348</td>
</tr>
<tr>
<td>PSA, ng/ml</td>
<td>10.0 (6.05, 17.6)</td>
<td>9.3 (5.5, 16.0)</td>
</tr>
<tr>
<td>n = 685</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1b</td>
<td>33 (10)</td>
<td>50 (14)</td>
</tr>
<tr>
<td>T1c</td>
<td>43 (12)</td>
<td>38 (11)</td>
</tr>
<tr>
<td>T2</td>
<td>270 (78)</td>
<td>259 (74)</td>
</tr>
<tr>
<td>Gleason score (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6</td>
<td>210 (70)</td>
<td>212 (67)</td>
</tr>
<tr>
<td>7</td>
<td>77 (26)</td>
<td>82 (26)</td>
</tr>
<tr>
<td>≥8</td>
<td>14 (5)</td>
<td>21 (7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>46 (13)</td>
<td>33 (9)</td>
</tr>
</tbody>
</table>

SPCG-4 = Scandinavian Prostate Cancer Group 4; RP = radical prostatectomy; WW = watchful waiting; PSA = prostate-specific antigen.

* All values are median (interquartile range) or frequency (proportion).
absolute risk reduction associated with surgery was highly dependent on tumor characteristics. For 65-yr-olds, the absolute 10-yr risk reduction associated with surgery ranged from 4% for a man with Gleason 6, stage T1 disease to 17% for a man with Gleason 8, stage 2 disease (Table 2). In some cases, it appears that men >70 yr have slightly poorer prostate cancer mortality than controls, likely due to random variation between similar estimates.

Given the wide variation in risk by baseline features, we assessed what proportion of men in this study were at “average” risk, and thus for whom the estimate from the randomized trial (4.6% risk difference; number needed to treat: 22) would have been an appropriate estimate. Only 27% of men in this study had a predicted number needed to treat between 15 and 30, corresponding to an average 3.3–6.7% absolute risk reduction; 18% of men had a predicted absolute risk reduction <1% or a number needed to treat >100, and 16% had a predicted risk reduction >10% or a number needed to treat <10.

The results for metastasis followed the same general trends. In total there were 163 metastasis events. Due to the increased incidence of metastasis, the net benefit associated with surgery was slightly larger and appeared to peak and decline at a slightly older age compared with the outcome of death from prostate cancer (Fig. 2). The overall difference in cumulative incidence of metastasis between men treated with radical prostatectomy and watchful waiting was 5.4% at 10 yr, corresponding to a number needed to treat of 19. Again, most of the patients were not at “average” risk: Only 32% of men had a predicted risk between 4% and 10% or a number needed to treat between 10 and 30.

Lastly, we evaluated how sensitive our results were to the choice of PSA level. We found essentially no difference in the estimated benefit of surgery for a given stage and...

---

**Table 2 – The 10-yr predicted cumulative incidence of metastasis or death from prostate cancer for men treated by radical prostatectomy versus watchful waiting**

<table>
<thead>
<tr>
<th>Age at diagnosis, yr</th>
<th>55</th>
<th>60</th>
<th>65</th>
<th>70</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Death from prostate cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gleason ≤6</td>
<td>RP</td>
<td>WW</td>
<td>RP</td>
<td>WW</td>
<td>RP</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>6</td>
<td>16</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
<td>11</td>
<td>28</td>
<td>14</td>
<td>27</td>
</tr>
<tr>
<td>Stage T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gleason ≤6</td>
<td>5</td>
<td>11</td>
<td>4</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>9</td>
<td>24</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>19</td>
<td>37</td>
<td>18</td>
<td>40</td>
<td>22</td>
<td>40</td>
</tr>
<tr>
<td>Gleason ≥8</td>
<td>8</td>
<td>15</td>
<td>5</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>14</td>
<td>32</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>13</td>
<td>34</td>
<td>21</td>
<td>42</td>
<td>21</td>
<td>41</td>
</tr>
<tr>
<td><strong>Stage T2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gleason ≤6</td>
<td>5</td>
<td>15</td>
<td>7</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>19</td>
<td>40</td>
<td>20</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>28</td>
<td>52</td>
<td>29</td>
<td>50</td>
</tr>
<tr>
<td>Gleason ≥8</td>
<td>8</td>
<td>15</td>
<td>5</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>14</td>
<td>32</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>13</td>
<td>34</td>
<td>21</td>
<td>42</td>
<td>21</td>
<td>41</td>
</tr>
</tbody>
</table>

**RP** = radical prostatectomy; **WW** = watchful waiting. *Death from other causes was treated as a competing risk.*
surgery in men finding as hypothesis generating, is that there is no effect of
tion of the SPCG-4 data, despite the authors presenting this
also in subgroup analyses. One widely quoted interpreta-
suggest heterogeneity not just in the trial as a whole but
predicted benefit of radical prostatectomy varies dramati-
Our statistical modeling approach indicated that the
absolute risk reduction suggests that about 21 patients
averages that are applicable only to a minority of patients.
reported, such as a 4.6% reduction in risk of death, are
projected benefit of radical prostatectomy varies dramati-
benefit for patients who have Gleason 8, or Gleason 7, stage
Gleason 6, T1 disease. Such patients would appear appropri-
treatment considered for stage or grade progression. It is
also difficult to justify surgery in patients much older than
Gleason 6 in 1995 might often be graded as Gleason 7 today.
This leads to a “Will Rogers” phenomenon [12], with risk falling for patients in both grading categories. This effect is exacerbated by the use of limited core biopsies in SPCG-4, which further underestimates grade.
That said, we would also caution against any oversimplis-
tic dismissal of SPCG-4 as irrelevant to contemporary
patients with screen-detected tumors [13]. Whatever the
difficulties in applying SPCG-4 results to, say, a contemporary
setting with prevalent PSA testing, we find randomized trial
data superior to what has become a type of academic nihilism, with conclusions such as “assessment is difficult”
for the US population [2]. Indeed, a recent high-profile
decision analysis of prostate cancer treatment, published in the
Journal of the American Medical Association [14], took a pragmatic approach in assuming that the benefit of
treatment in SPCG-4, a relative risk reduction of about
35%, would be halved. Using this type of approach on our data
demonstrates that any conclusions from our findings would be relatively robust to various assumptions about the
reduction in treatment effect that might be seen in a screened population subject to stage shift. For example,
whether the attenuation of treatment benefit is 33%, 50%, or
67%, the absolute risk reductions for a 65-yr-old with T1
Gleason 6 disease would be 4%, 3%, or 2%, which are of
questionable clinical significance given that both the side
effects of surgery and urinary and erectile dysfunction are persistent.
In the context of the side effects of contemporary radical
prostatectomy, a simple interpretation of our results is therefore as follows. First, it is hard to justify surgery in
Gleason 6, T1 disease. Such patients would appear appropri-
ate for an active surveillance program, with curative
treatment considered for stage or grade progression. It is
also difficult to justify surgery in patients much older than
age 70. At the other extreme, surgery seems unequivocally of
benefit for patients who have Gleason 8, or Gleason 7, stage
T2 disease. For Gleason 6 T2 and Gleason 7 T1, treatment is

---

**Fig. 2 – The 10-yr predicted risk reduction for metastasis among men treated by radical prostatectomy versus watchful waiting in the presence of a competing risk (death from other causes) for men with Gleason 6 (orange line), Gleason 7 (blue dashed line), or Gleason 8 (green line) disease. Figures are stratified by stage and adjusted to a prostate-specific antigen level of 10 ng/ml.**
more of a judgment call, depending on the patient’s tolerance to the consequences of surgery, such as possible urinary dysfunction, versus anxiety associated with active surveillance. Clinical judgment based on considerations such as the number of positive biopsy cores and the health of the patient will also play a role: Age is an imperfect predictor of death from other causes, and comorbidities should play an important role in the clinical evaluation of the individual patient. Note that these findings are highly comparable with those of the Prostate Cancer Intervention Versus Observation Trial (PIVOT), which have been reported at the American and European urologic association meetings. In brief, PIVOT finds little if any benefit for the treatment of low-risk prostate cancer, with correspondingly larger differences between groups for intermediate- and high-risk disease.

A contemporary patient opting for conservative management will likely be followed on an active surveillance protocol, with crossover to active treatment on signs of disease progression. This reduces the risk in control groups and thus the benefit of immediate surgery. Focal therapy is also becoming increasingly available for low-risk patients, again likely reducing the relative benefit of surgery. A countervailing influence is that contemporary surgery, where practiced at high-volume hospitals, may also be associated with lower mortality than reported in SPCG-4. First, in contemporary practice, patients with positive lymph nodes remain eligible for radical prostatectomy and, when treated, they have a 5-yr probability of remaining recurrence free close to 50% [15]. These patients tend to be at above-average risk and have the most to gain from radical prostatectomy. Because they were not treated in SPCG-4, although still analyzed in the treatment group, the SPCG-4 data do not fully reflect current policy in many centers and may not entirely capture the benefit of surgery as it is practiced today compared with conservative management. Second, strong evidence has emerged that surgical outcomes are related to surgical experience [16,17]; it seems likely that a patient presenting at a high-volume center in the 2010s will be treated by a surgeon with, on average, greater surgical experience than the average SPCG-4 patient.

The confidence intervals associated with the estimates reported here are wide. For example, the confidence intervals around the predicted risk of dying from prostate cancer within 10 yr for a 65-yr-old man with Gleason 7, stage T1 disease treated with radical prostatectomy ranged from 3% to 14%. Given the relatively small sample size—the SPCG-4 study was powered for confidence intervals around a group estimate, not individualized estimates of benefit—these findings are not unexpected. Importantly, the confidence intervals reflect statistical precision, but clearly greater imprecision is associated with the application of the results to populations subject to the stage shift.

Our conclusions are reasonably consistent with the clinical practice and the work of prior authors. Few doubt that Gleason 8 is an indication for curative treatment in a man with a reasonable life expectancy; similarly, there is increasing consensus that patients with Gleason 6, T1 disease can be followed by active surveillance [18]. This report provides randomized data in support of these positions. Our finding of questionable benefit in Gleason 6 T2 and Gleason 7 T1 is more novel. There has been limited work on active surveillance for men with intermediate-risk disease [19]; we hope our work prompts further investigation of whether the current rather restrictive criteria for active surveillance [20] might be safely liberalized.

Our results call for changes in the way in which randomized trials are interpreted. The traditional way to use randomized trial evidence in the clinical consultation is to cite the overall results of a study, such as a 5% reduction in the absolute risk of death, and then discuss with the patient whether the benefit of treatment outweighs its harms. However, where it is possible to predict risk, it can be clearly seen that overall results of a trial are an average of very different levels of patient benefit. Such variations in risk reduction constitute the rationale for models such as Adjuvant! Online that provide patients with the expected risk of cancer recurrence with and without adjuvant therapy [21].

5. Conclusions

We have shown using statistical modeling that expected patient benefit from radical prostatectomy varies enormously around the central estimate from a randomized trial. Our findings can be used to counsel individual patients who are considering treatment options for localized prostate cancer. Our work should also prompt further statistical prediction modeling of randomized trial data.

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, Holmberg.

Acquisition of data: Vickers, Bennette, Steineck, Adami, Johansson, Bill-Axelson, Palmgren, Garmo, Holmberg.

Analysis and interpretation of data: Vickers, Bennette, Steineck, Adami, Johansson, Bill-Axelson, Palmgren, Garmo, Holmberg.

Drafting of the manuscript: Vickers, Bennette, Holmberg.

Critical revision of the manuscript for important intellectual content: Vickers, Bennette, Steineck, Adami, Johansson, Bill-Axelson, Palmgren, Garmo, Holmberg.

Statistical analysis: Vickers, Bennette, Garmo.

Obtaining funding: None.

Administrative, technical, or material support: None.

Supervision: None.

Other (specify): None.

Financial disclosures: Andrew Vickers certifies 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: The Swedish Cancer Society (07 0512 to Jan-Erik Johansson) and the National Institutes of Health in the United States (1R01 CA 108746-01A1 to Gunnar Steineck). 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 P50-CA92629 SPORE grant from the National Cancer Institute to Dr. P.T. Scardino. The funders had no role in the design and
References


