Clinical raters of life-expectancy before radical prostatectomy or definitive radiotherapy for localized prostate cancer

Jochen Walz*†, Andrea Gallina*‡, Paul Perrotte¶, Claudio Jeldres*, Quoc-Dien Trinh¶, Georg C. Hutterer*, Miriam Traumann*, Alvaro Ramirez¶, Shahrokh F. Shariat§, Michael McCormack¶, Jean-Paul Perreault¶, Francois Bénard¶, Luc Valiquette¶, Fred Saad¶ and Pierre I. Karakiewicz*

*Cancer Prognostics and Health Outcomes Unit, University of Montreal, Montreal, Quebec, Canada, and Departments of Urology, †University Medical Centre Eppendorf, Hamburg, Germany, ‡Vital-Salute University, Milan, Italy, ¶University of Montreal, Montreal, Quebec, Canada, and §University of Texas South-western Medical Center, Dallas, TX, USA

Accepted for publication 1 June 2007

OBJECTIVE
To test the accuracy of predicting life-expectancy (LE) among 19 raters, as the accurate prediction of LE in candidates for definitive therapy for localized prostate cancer is crucial, and little is known of the ability of clinicians to predict LE.

SUBJECTS AND METHODS
We randomly selected the case-vignettes of 50 patients treated with either radical prostatectomy (RP, 25) or external beam radiotherapy (EBRT, 25) for prostate cancer, and who either survived for >10 years or died earlier with no evidence of disease relapse. The median age at treatment was 67 years and the median Charlson Comorbidity Index (CCI) was 2. The raters consisted of urology staff (six), urology residents (10) and medical students (three). The case-vignettes included patient age, comorbidities and CCI score, and raters were asked to predict the survival at 10 years (yes vs no), assuming no disease relapse.

RESULTS
Of the 50 cases, 20 (40%) did not survive for >10 years; clinicians estimated a mean (range) of 23 (10–35) deaths before 10 years. The mean (95% confidence interval) overall predictive accuracy (0.5 = chance, 1.0 = perfect prediction) of LE predictions was 0.68 (0.64–0.71). Individual accuracy ranged from 0.52 (staff) to 0.78 (staff). There were no important differences among the rater groups (residents 0.69 vs staff 0.67 vs medical students 0.67).

CONCLUSIONS
Clinicians are relatively poor at predicting LE; tools to predict LE might be able to improve clinicians’ performance in this important part of decision-making about prostate cancer treatment. It remains to be determined whether this limitation exclusively applies to prostate cancer or also to other malignancies.

KEYWORDS
prostate cancer, prediction, radical prostatectomy, radiotherapy, life expectancy

INTRODUCTION
The 10-year rule is the most frequently cited life-expectancy (LE) benchmark for delivering definitive therapy to patients with localized prostate cancer. This 10-year rule has been adopted by several professional associations and appears in their guidelines [1,2]. It is based on the assumption that delivery of definitive therapy to individuals whose tumour characteristics are too indolent to threaten their LE represents over-treatment. The latter might unnecessarily add to cost, complications, side-effects, early and late morbidities, and treatment-related mortality [3,4].

In Canada, the average LE of men aged 65 years is 17.0 years, vs 10.3 for those aged 75 years and 7.7 for those aged 80 years, vs respectively, 18.4, 11.8 and 9.0 years in the USA [5,6]. Based on a 10-year LE, men aged <75 years qualify for definitive therapy for prostate cancer. This was substantiated by recent reports suggesting definitive therapy up to the age of 80 years, especially in men with high-grade prostate cancer [7]. However, in septa- and octogenarians, comorbidities and prostate cancer compete for LE [3]. Indeed, Albertsen et al. [3], in their series of men treated with watchful-waiting and delayed intervention, showed that 63% of men aged 65–69 years with intermediate grade prostate cancer (Gleason sum 6), die from causes other than cancer. Therefore, favouring definitive therapy in older men might result in significant over-treatment.

The accurate estimation of LE could possibly decrease the rate of over-treatment. Unfortunately, reports suggest that clinicians are poor judges of LE [8,9]. Extrapolating these data to patients with prostate cancer suggests that an equally poor performance might be expected, if LE is predicted for these patients. As assessing LE is particularly important in deciding treatments for prostate cancer, we explored the ability of clinicians to predict the 10-year LE in patients who were...
either survived for external beam radiotherapy (EBRT), who radical prostatectomy (RP) or definitive vignettes of patients treated with either purpose we used 50 randomly selected case-
candidates for definitive therapy. For this restriction resulted in 9131 patients remaining; of these, 5955 were treated with RP and 3176 received definitive EBRT. Finally, to use only uncensored observations for the analysis, we identified all patients who died or had a follow-up of >10 years (4724); of these, we randomly selected 25 patients each who had RP or EBRT.

The case-vignettes of these 50 patients were presented in the form of a questionnaire to a panel of 19 clinicians, and included age and CCI score at the time of treatment, as well as each individual comorbidity. The raters remained unaware of the previously chosen treatment type. Of the 19 clinicians, six were urology staff, 10 were residents and three were medical students in their last year of training. The clinicians were asked to rate whether each of the 50 patients survived 10 years after definitive therapy (yes vs no), provided there was no disease relapse.

Statistical analyses consisted of independent sample t-tests, Kaplan-Meier, life-table and predictive accuracy (PA) analyses. The latter consisted of receiver operating characteristics curve (ROC) analysis, a tool to graphically explore the compromise between the sensitivity and the specificity of a diagnostic test. The ROC-derived accuracy is quantified by the area under the curve (AUC), where a value of 0.5 represents equal chance and 1.0 represents perfect prediction. As a binary outcome was assessed (survival at 10 years, yes vs no), no adjustment for censored data was necessary. PA expressed by the AUC was calculated for each clinician and for the three groups defined as staff, residents and medical students; all tests were two-sided with a significance level set at 0.05.

RESULTS

The characteristics of the 50 patients included in the study are given in Table 1; the mean (median, range) follow-up was 8.1 (9.8, 0.1–15.4) years, the age 65.8 (67, 55–83) years and the median CCI 2 (0–9). During follow-up there were 21 deaths. The median actuarial

<table>
<thead>
<tr>
<th>Variables</th>
<th>All</th>
<th>RP</th>
<th>EBRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at treatment, years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean (median)</td>
<td>65.8 (67)</td>
<td>61.4 (61)</td>
<td>70.2 (71)</td>
</tr>
<tr>
<td>range</td>
<td>55–83</td>
<td>55–70</td>
<td>55–83</td>
</tr>
<tr>
<td>n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;60</td>
<td>12 (24)</td>
<td>9 (36)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>60–69</td>
<td>20 (40)</td>
<td>15 (60)</td>
<td>5 (20)</td>
</tr>
<tr>
<td>≥70</td>
<td>18 (36)</td>
<td>1 (4)</td>
<td>17 (68)</td>
</tr>
<tr>
<td>Follow-up, years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>median (mean)</td>
<td>8.1 (9.8)</td>
<td>10.8 (10.2)</td>
<td>5.4 (3.5)</td>
</tr>
<tr>
<td>range</td>
<td>0.1–15.4</td>
<td>4.2–15.4</td>
<td>0.1–14.6</td>
</tr>
<tr>
<td>Deaths at 10 years</td>
<td>20 (40)</td>
<td>4 (16)</td>
<td>16 (64)</td>
</tr>
<tr>
<td>Actuarial survival, years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>median</td>
<td>nr</td>
<td>nr</td>
<td>3.5</td>
</tr>
<tr>
<td>mean</td>
<td>10.3</td>
<td>14.0</td>
<td>6.4</td>
</tr>
<tr>
<td>Overall survival probabilities, %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at 5 years</td>
<td>64.0</td>
<td>92.0</td>
<td>36.0</td>
</tr>
<tr>
<td>at 10 years</td>
<td>57.3</td>
<td>84.0</td>
<td>28.8</td>
</tr>
</tbody>
</table>

TABLE 1

The descriptive characteristics of the 50 men who had no secondary therapy for prostate cancer after either RP or EBRT.

candidates for definitive therapy. For this purpose we used 50 randomly selected case-vignettes of patients treated with either radical prostatectomy (RP) or definitive external beam radiotherapy (EBRT), who either survived for >10 years or died with no evidence of disease relapse before then.

SUBJECTS AND METHODS

In Canada, healthcare is under independent administrative control by its 11 provinces and territories. In the Province of Quebec, the Quebec Health Plan represents the exclusive insurer. Its database is used for billing purpose and allows virtually complete ascertainment of all health services and medications covered by the Plan and provided in the Province of Quebec. These also contain all treatment methods for prostate cancer and include RP, EBRT, and all types of hormonal manipulation. Moreover, the Health Plan relies on the 9th Version of the International Classification of Diseases (ICD-9) and the respective dates of any secondary therapy, age, ICD-9 codes of individual comorbidities and CCI score before therapy. The Health Plan records all disease codes since 1 June 1983. These allow the defining of the Charlson Comorbidity Index (CCI) scores of all disease codes since 1 June 1983.

The Health Plan database allowed us to identify all men diagnosed with prostate cancer (ICD-9185). We used the respective RP and EBRT billing codes to identify patients who received one or both treatments. Analyses were restricted to men treated between 1 January 1989 and 31 December 2000. Each record included the type of treatment and was accompanied by the date of RP or the starting date of EBRT, types and dates of any secondary therapy, age, ICD-9 codes of individual comorbidities and CCI score before therapy. The Health Plan records contain no information on cancer stage, grade, preoperative serum PSA level or specific cause of death.

Overall, 17 570 patients were identified; of these, we selected only men who receive no secondary therapy for cancer, to exclude the effect of cancer-specific mortality on the overall survival. This was done because no information on cause specific mortality was available. Secondary therapy was defined as either EBRT after RP or RP after EBRT, as well as hormonal manipulation at any time before or after either RP or definitive EBRT. This restriction resulted in 9131 patients remaining; of these, 5955 were treated with RP and 3176 received definitive EBRT. Finally, to use only uncensored observations for the analysis, we identified all patients who died or had a follow-up of >10 years (4724); of these, we randomly selected 25 patients each who had RP or EBRT.
FIG. 1. The overall survival after RP or EBRT in the 50 men included in the survey.

![Graph showing overall survival over time for men treated with RP or EBRT]

Survival was not reached (mean 10.3), and the 5- and 10-year actuarial survival probabilities were, respectively, 64.0% and 57.3% (Fig. 1).

Between the 25 patients treated with RP or EBRT there was a statistically significant difference in age ($P < 0.001$), but not for CCI score ($P = 0.2$, Table 1). There was a statistically significant difference in actuarial survival between the groups, as shown by the 10-year survival rates of 84% vs 29% for each, respectively (log-rank $P < 0.001$).

Of the 50 cases included in the study, 20 (40%) did not survive for $>10$ years after definitive therapy, whereas the clinicians estimated that a mean (range) of 23 (10–35) men would die before 10 years after definitive therapy. The overall mean (95% CI) AUC for predicting the 10-year LE of all 19 participants was 0.68 (0.64–0.71). Figure 2 shows the ROCs for all participants; the highest individual AUC was 0.78 (0.64–0.92; urology staff) and the lowest individual AUC was 0.52 (0.50–0.68; urology staff). As a group, the residents had the highest overall AUC, which was 0.69 (0.64–0.74). Individual AUCs of residents ranged from 0.61 (0.50–0.77) to 0.75 (0.61–0.89). The overall AUCs achieved by the group of medical students and by the urology staff were both 0.67 (0.58–0.76 and 0.60–0.72, respectively). The individual AUCs of the medical students ranged from 0.58 (0.50–0.75) to 0.74 (0.60–0.87) and the individual AUCs of the urology staff from 0.52 (0.50–0.68) to 0.78 (0.64–0.92). There was no statistically significant difference in PAs among these three groups ($P \geq 0.7$).

DISCUSSION

Prostate cancer is an important clinical and public health issue, based on its incidence and mortality [13]. It was estimated that in 2006, there were 234,460 new cases and 27,350 prostate cancer-related deaths in the USA [13]. It accounts for 33% of all male cancers and 9% of male cancer-related deaths [13]. Canadian and European statistics corroborate these data [14,15].

These values show the importance of the burden of prostate cancer; consequently prostate cancer-related health expenditures represent an important part of the total health budget. The delivery of definitive therapy to men whose tumour characteristics are too indolent to threaten their LE represents over-treatment, which might unnecessarily add to cost, complications, side-effects, early and late morbidities, and treatment-related mortality [3,4]. Thus, from a societal and individual perspective, men with a suboptimal LE should not be considered for definitive therapy, but should be offered expectant management with delayed intervention [3]. These considerations clearly indicate the importance of LE-based treatment-decision making. However, the accurate prediction of LE in patients with prostate cancer is a challenge.

Our study showed that clinicians’ ability to predict the 10-year LE was moderate at best. The mean AUC of all 19 participants was only 0.68, where an AUC of 0.5 would imply a prediction no better than equal chance, and an AUC of 1.0 would imply a perfect prediction. The highest individual AUC was 0.78, which represented a reasonably good ability to predict this outcome. Conversely, the lowest AUC was 0.52, which corresponds to total inability to predict this outcome. Interestingly, both values were achieved by urology staff. Therefore, urology staff might be very good or very poor raters of LE. The AUC achieved by the residents as a group was the highest (0.69); it exceeded that of the urology staff (0.67) and that of medical students (0.67). The lack of meaningful differences among these three groups, that had different levels of clinical experience and exposure to patients with prostate cancer, suggests that neither expertise nor exposure time are important in predicting LE. However, there were too few participants and case-vignettes to attain meaningful and/or statistically significant differences among the three groups or among individual raters. It might be postulated that several hundreds of case-vignettes would have been required to accurately discern individual intergroup or inter-rater differences. Therefore, it can only be stated that all three groups performed equally poorly.

The analysis of the case-vignettes showed a discrepancy between the survival pattern of patients treated with RP or EBRT. The survival rates of the latter were lower, which might be due to numerous baseline comorbidities, as well as advanced age in this group. Previous reports indicated similar survival rates in patients treated with definitive RT [16–18]; e.g. Fowler et al. [16] (in 138 patients) reported a 10-year actuarial survival of 27% for those treated with RT and who had a CCI of 1. More recent studies from the USA and
Europe showed that only 39–40% of men treated with RT were alive 10 years afterward [17,18]. These and the present survival data indicate that the suboptimal survival of patients after RT can be recorded throughout North America and Europe. However, this difference in survival pattern between RP and RT adds to the difficulty in predicting LE and might represent one of the main reasons accounting for the poor ability to predict LE. To avoid biased rater opinions related to treatment type, raters were not informed about the type of treatment that was offered to each patient.

We are not the first to report the phenomenon of poor LE predictions; we corroborate previous reports where clinicians were judged to have a poor ability to predict survival in patients [6,9]. For example, Chow et al. [8] showed that clinicians overestimated the LE of patients with metastatic cancers who were referred for palliative RT by nearly 100%. Conversely, we could not corroborate the findings of Krahn et al. [19] who reported 82% accuracy for predicting the 10-year LE in a survey of 191 urologists and radiation oncologists. In that study, the investigators used 18 patient scenarios and provided age and comorbidities as predictors. The main caveat of that study was that real patient survival data were not used. Instead, estimates of patient survival derived from a Markov model were considered as the reference standard; unfortunately, the accuracy of the Markov model was not reported. Therefore, the reference standard of LE might differ substantially from the true survival of these patients. This methodological weakness undermines the validity of the reported 82% accuracy and caution is required when generalizing those results.

Our results question the ability of clinicians to predict LE, but several tools can be used to improve the prediction of LE; these consist of life tables, comorbidity indices and three prognostic models [5,17,20,21]. Life tables represent an average prediction of the remaining life-years based on gender and age characteristics. Life-table predictions can be undermined when grouping patients according to age strata [22]. Several individuals might have a LE below or above the average, and this will be missed by life tables. Moreover, population averages might not apply to patients with localized prostate cancer, who might be healthier than the general population [17]. Finally, to the best of our knowledge, the accuracy of life tables in predicting LE has not been formally tested in patients with prostate cancer.

Various comorbidity indices can also assist in estimating LE; of those, the CCI is the most widely reported [17,20,21,23]. Unfortunately, the PA of the available comorbidity indices has never been quantified in patients with prostate cancer. Albertsen et al. [21] tested three different comorbidity indices in patients with prostate cancer, i.e. the Kaplan–Feinstein Index (KFI), the CCI and the Index of Coexisting Disease (ICED). All indices were highly significant predictors of mortality for patients dying of causes other than prostate cancer, but no PA for these indices was provided [21]. Boulos et al. [24] compared five different comorbidity indices in patients with prostate cancer, i.e. the Chronic Disease Score (CDS), the ICED, the Cumulative Illness Rating Scale (CIRS), the KFI and the CCI. They found that the CDS and the ICED were the best age-independent predictors of survival. The CIRS, the KFI and the CCI followed in terms of their ability to predict survival. However, again no PA was provided for any of the tested indices; instead, the authors reported the percentage of explained variance [24]. This accuracy measure might overestimate the model's performance and is not directly amenable to correction for overfit bias.

Several investigators combined comorbidities with age and prostate cancer characteristics in an attempt to predict LE [17,20,21]. Cowen et al. [20] used 506 patients to develop a nomogram predicting the 5-, 10- and 15-year LE in patients treated with either RP, RT or conservative management. Predictors included age, CCI, general performance, angina history, blood pressure, body mass index, tobacco use, marital status, PSA level, Gleason sum, clinical stage and treatment type (AUC 0.73). Tewari et al. [17] developed a similar tool in 1611 patients, which relied on race, CCI, age, biopsy Gleason score, PSA level and treatment type (AUC 0.69). Finally, Albertsen et al. [20,21] developed a model (in 451 men), which relied on comorbidity, age and Gleason sum, and an AUC of 0.71 was reported in an external validation. The limitation of these tools resides in their modest PA for the 10-year LE [0.69–0.73] [17,20,21]. The overall PA of all participants in the present study was 0.68. Unfortunately, we were unable to externally validate the above comorbidity indices or predictive tools in our dataset, as it did not include all the necessary predictors. Lack of direct external validation undermines the validity of direct comparisons, and therefore no conclusion can be drawn on either superiority or inferiority of the tools over clinicians.

None of the reported available predictive tools, devised to predict various prostate cancer outcomes, is capable of predicting perfectly; the PA of these tools is 0.65–0.84 [25,26]. Therefore, clinicians should not be expected to predict perfectly either. However, an AUC of 0.68, attained by the present clinicians, represents low to moderate accuracy. Thus, caution is recommended when clinician-derived LE estimates are used for deciding treatments for prostate cancer. Lack of a sufficient PA calls for tools that could assist clinicians in predicting LE, to provide greater accuracy.

The present study has some limitations; the lack of cause-specific mortality is the main one. Instead of restricting the group to patients with no secondary therapy, we could have excluded those who died from prostate cancer, if that information had been available. Furthermore, the design of the study, which relied on case-vignettes, might be regarded as artificial. However, it is impossible to prospectively present sufficiently many clinical patients to sufficiently many raters at the same time. Therefore, this design based on case-vignettes is shared with several other studies that explored the ability of clinicians to judge patient outcomes in daily practice [19,27,28]. These studies are also limited by either the number of raters [17–25] or by the number of patients [10–33] [19,27,28]. Moreover, we are well aware that selecting treatment for prostate cancer is not solely based on LE and cancer characteristics. Quality-of-life considerations, patient and physician preferences and treatment availability all add to the complexity of treatment selection. Therefore, LE represents only one component in the complex process of treatment decision-making. More studies will be needed to find tools with a better performance in predicting LE. Moreover, the appropriateness of the 10-year rule in treatment for prostate cancer remains to be determined, as no study so far has shown that this benchmark allows differentiation between possible over-treatment and the need for definitive therapy.

In conclusion, clinicians are relatively poor at predicting LE in patients with prostate cancer.
Individual predictions can vary substantially and seem to be independent of professional experience. The inability to accurately predict LE calls for tools that could better quantify this outcome. It remains to be determined whether this limitation exclusively applies to prostate cancer or also to other malignancies.

CONFLICT OF INTEREST

None declared.

REFERENCES


Correspondence: Pierre I. Karakiewicz, Cancer Prognostics and Health Outcomes Unit, University of Montreal Health Center (CHUM), 1058, rue St-Denis, Montréal, Québec, Canada, H2X3J4. e-mail: pierre.karakiewicz@umontreal.ca

Abbreviations: LE, life-expectancy; RP, radical prostatectomy; EB(RT), external beam (radiotherapy); ROC, receiver operating characteristics (curve); AUC, area under the curve; CCI, Charlson Comorbidity Index; ICD, International Classification of Diseases; PA, predictive accuracy; KFI, Kaplan-Feinstein Index; ICED, Index of Coexisting Disease; CDS, Chronic Disease Score; CIRS, Cumulative Illness Rating Scale.