



Published in final edited form as:

Cancer. 2009 December 1; 115(23): 5460–5469. doi:10.1002/cncr.24615.

## Clinical benefits of a multivariable prediction model for bladder cancer: a decision analytic approach

Andrew J Vickers, PhD<sup>1</sup>, Angel M Cronin, MS<sup>1</sup>, Michael W Kattan, PhD<sup>2</sup>, Mithat Gonen, PhD<sup>1</sup>, Peter T Scardino, MD<sup>3</sup>, Matthew I. Milowsky, MD<sup>4</sup>, Guido Dalbagni, MD<sup>3</sup>, and Bernard H. Bochner, MD<sup>3</sup> for The International Bladder Cancer Nomogram Consortium

<sup>1</sup>Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center 1275 York Ave, New York, NY 10021

<sup>2</sup>Quantitative Health Sciences, Cleveland Clinic 9500 Euclid Avenue Cleveland, OH 44195

<sup>3</sup>Department of Surgery, Memorial Sloan-Kettering Cancer Center 1275 York Ave, New York, NY 10021

<sup>4</sup>Department of Medicine, Memorial Sloan-Kettering Cancer Center 1275 York Ave, New York, NY 10021

### Abstract

**Background**—Multivariable prediction models have been shown to predict cancer outcomes more accurately than cancer stage. The effects on clinical management are unclear. We aimed to determine whether a published multivariable prediction model for bladder cancer (“bladder nomogram”) improves medical decision making, using referral for adjuvant chemotherapy as a model.

**Methods**—We analyzed data from an international cohort study of 4462 patients undergoing cystectomy without chemotherapy 1969 – 2004. The number of patients eligible for chemotherapy was determined using pathologic stage criteria (lymph node positive or stage pT3 or pT4), and for three cut-offs on the bladder nomogram (10%, 25% and 70% risk of recurrence with surgery alone). The number of recurrences was calculated by applying a relative risk reduction to eligible patients' baseline risk. Clinical net benefit was then calculated by combining recurrences and treatments, weighting the latter by a factor related to drug tolerability.

**Results**—A nomogram cut-off outperformed pathologic stage for chemotherapy for every scenario of drug effectiveness and tolerability. For a drug with a relative risk of 0.80, where clinicians would treat no more than 20 patients to prevent one recurrence, use of the nomogram was equivalent to a strategy that resulted in 60 fewer chemotherapy treatments per 1000 patients without any increase in recurrence rates.

---

**Corresponding Author:** Andrew Vickers, PhD, 307 E 63<sup>rd</sup> Street New York, NY 10065, vickersa@mskcc.org, Phone: (646)-735-8142 Fax: (646)-735-0011.

#### Contributors

A J Vickers took part in the study design, study funding, and revising of the manuscript. Both A J Vickers and A M Cronin took part in the statistical analysis and drafting of the manuscript. B H Bochner took part in data acquisition, study design, and drafting and revising the manuscript. M W Kattan, M Gonen, P T Scardino, M I Milowsky, and G Dalbagni took part in drafting and revising the manuscript. All authors have seen and approved the final version of the manuscript.

#### Conflict of interest statement

Andrew Vickers had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. None of the authors have any relevant conflicts of interest.

**Conclusions**—Referring cystectomy patients to adjuvant chemotherapy on the basis of a multivariable model is likely to lead to better patient outcomes than the use of pathological stage. Further research is warranted to evaluate the clinical effects of multivariable prediction models.

### Keywords

ladder cancer; adjuvant chemotherapy; prognosis; decision support; outcomes

## Background

Many decisions in oncology depend, implicitly or explicitly, on predictions. These predictions are normally thought of in terms of “risk”: typically, we act when a patient is deemed at sufficiently high risk that the benefits of intervention, in terms of reducing risk, outweigh the harms, in terms of toxicities. Most commonly, decisions that involve prediction are based on risk categories. The most common system for risk categorization in cancer is cancer stage, with more aggressive treatments reserved for patients with higher stage disease. Stage can influence the extent of surgery, such as in breast or bladder cancer; the intensity of chemotherapy, such as in lymphoma; and whether adjuvant therapy is indicated, such as in bladder or colon cancer.

Recent years have seen an upsurge of interest in multivariable prediction models. Typically, these models provide a numerical estimate of risk in the form of a probability on the basis of several tumor and patient characteristics. The well-known “Kattan nomogram”, for example, provides the probability of prostate cancer recurrence after radical prostatectomy on the basis of stage, grade and prostate-specific antigen level<sup>1</sup>. Numerous similar models have been published for a variety of different cancers<sup>2–6</sup> and for specific treatment decisions, such as adjuvant chemotherapy after breast cancer<sup>7</sup>. It seems reasonable that such models might predict more accurately than simple staging systems because they include additional prognostic information. For instance, a patient with high-grade organ-confined prostate cancer is likely at a higher risk of recurrence than a patient at a similar stage, but with low-grade disease. Empirical studies have confirmed that multivariable models provide more accurate predictions than American Joint Committee on Cancer (AJCC) staging, or other simple risk groupings, in a wide variety of cancers including melanoma<sup>8</sup>, gastric cancer<sup>9</sup>, pancreatic cancer<sup>10</sup> and prostate cancer<sup>11</sup>. In the case of bladder cancer, the topic of the current paper, the predictive accuracy of a multivariable model for recurrence after radical cystectomy (the “bladder cancer nomogram”) was a concordance index of 0.75 compared to only 0.68 for the AJCC TNM staging and 0.62 for standard pathologic stage<sup>12</sup>.

Nonetheless, the clinical implications of an improved concordance index are not immediately obvious: we may be able to predict better who is at high risk of death from bladder cancer, however, this may make little practical difference to patient care. We are interested in better defining whether use of the bladder cancer nomogram would improve clinical decision making, such as whether a patient should be administered adjuvant chemotherapy. To address this question we examined a recent multivariable prediction model, the bladder cancer nomogram, using decision analytic methods.

## Methods

### Patients

Collection of data for the International Bladder Cancer Nomogram Consortium has been previously described<sup>12</sup>. In total, data on 9064 patients who underwent radical cystectomy were collected from 13 institutions in 6 countries. Since we wished to estimate the risk of recurrence following radical cystectomy alone, we excluded patients who had received

systemic adjuvant chemotherapy, neoadjuvant chemotherapy, or definitive pelvic radiotherapy (n=2001), and patients for whom pelvic radiotherapy was unknown (n=319). Of the remaining patients, 871 were not followed for recurrence, and 1411 were excluded due to missing data on variables included in the nomogram. Our sample therefore consisted of 4462 patients who were managed by radical cystectomy only, were followed for recurrence and had complete data for all predictors: sex, age, pathologic stage, histology, nodal status, grade and time between diagnosis and cystectomy. The proportion of patients by institution is similar to that previously reported<sup>12</sup>.

### Adjuvant chemotherapy for bladder cancer

Our interest in this paper is adjuvant rather than salvage chemotherapy. Adjuvant chemotherapy is commonly given to high-risk bladder cancer patients after radical cystectomy. Two separate groups have published meta-analyses of randomized trials and reported similar findings suggesting a survival benefit to adjuvant chemotherapy. Ruggeri et al. pooled data from five Phase III trials and reported a statistically significant improvement in disease-free survival in patients receiving adjuvant chemotherapy (hazard ratio of 0.65; 95% C.I. 0.54, 0.78) for disease-free survival<sup>13</sup>. The advanced bladder cancer meta-analysis collaboration<sup>14</sup> included an additional trial and performed individual patient data analysis: they report a hazard ratio of 0.68 (95% C.I. 0.53, 0.89). These numbers are approximately equivalent of a relative risk of recurrence at 5 years of approximately 0.75. However, given the relatively limited number of patients included in these meta-analyses, fewer than 500, and the accordingly wide confidence intervals, we planned to use a variety of different estimates of relative risks in our analyses.

### Statistical analysis

Eligibility for adjuvant chemotherapy therefore depends on a decision rule for identification of high risk patients. With respect to bladder cancer, patients with stage pT3 or pT4 disease, or those with positive nodes, are generally considered at high risk for recurrence and therefore recommended by standard guidelines as eligible for chemotherapy; node-negative patients with organ confined disease (less than pT3) are followed by observation only after surgery.<sup>15</sup> We wished to compare this eligibility criterion with one of three prespecified rules based on the nomogram predicted risk of recurrence at five years: 10%, 25% and 70%.

Our overall statistical approach follows a previously published methodology<sup>16</sup>. In brief, to calculate sensitivity and specificity for survival time data, we first define  $x=1$  if the patient is classified as being at high risk and  $x=0$  otherwise;  $s(t)$  is the Kaplan-Meier survival probability at time  $t$ , predefined as five years. Following Begg et al.<sup>17</sup> we use the following formula for sensitivity and specificity:

$$\begin{aligned} \text{Sensitivity} &= [1 - (s(t)|_{x=1})] \times P(x=1) \div [1 - s(t)] \\ \text{Specificity} &= (s(t)|_{x=0}) \times P(x=0) \div s(t) \end{aligned}$$

Assuming a constant relative risk ( $RR$ ), the proportion of patients who recur with a particular intervention scenario  $i$  can be given as:

$$\text{Sensitivity}_i \times [1 - s(t)] \times RR + (1 - \text{Specificity}_i) \times [1 - s(t)]$$

As we are treating recurrence as a binary event, relative risk is defined as the risk of recurrence at five years with treatment divided by five-year recurrence risk without treatment. Our primary analyses assumed that relative risk was constant across risk groups

so that, for example, given a chemotherapy regimen associated with a relative risk of 0.9, a patient with a 50% probability of recurrence without treatment would have a risk of 45% on chemotherapy and patient with a 10% baseline risk would have a 9% risk.

As a sensitivity analysis, we considered three scenarios, where the treatment was most effective for patients at low, average, and high risk. A patient's risk was estimated using the nomogram probability, which is justified because the nomogram has been shown to be well calibrated<sup>12</sup>. Figure 1 gives an illustration of how relative risk was varied by absolute risk for each scenario. The mean relative risk was kept constant between scenarios, such that the total number of recurrences was identical regardless of the presumed relationship between absolute and relative risk.

The proportion of patients who would be treated under each strategy is estimated counting the proportion of patients in our data set meeting each criterion. Knowing the treatment and recurrence rates for each strategy does not necessarily identify the optimal approach. Often, when comparing two strategies, one will be associated with a lower rate of treatment but also a higher recurrence rate. To calculate whether the reduction in the number of patients receiving chemotherapy offsets the increase in recurrence rates, we need to consider the maximum number of patients a clinician would consider treating to prevent one recurrence. This is known as the “number-needed-to-treat threshold” ( $NNT_T$ ) and is a clinical judgment that can vary from clinician to clinician, and from patient to patient<sup>18</sup>. The  $NNT_T$  is reciprocal of the minimum, clinically significant difference, a concept necessary to design and interpret randomized trials<sup>19, 20</sup>.  $NNT_T$  is a measure of drug tolerability: an agent that is easy to take, inexpensive and associated with low toxicity would have a high  $NNT_T$ ; a toxic, inconvenient or expensive drug would have a low  $NNT_T$ . Note that the number-needed-to-treat threshold ( $NNT_T$ ) is different from the usually reported number-needed-to-treat (NNT): NNT is calculated from the results of a study and tells us how many patients would need to be treated to prevent one event;  $NNT_T$  is a clinical consideration independent of the results of any study, and tells us how a physician weights the harms of treatment against the benefits of avoiding an event.

We can then define “clinical net benefit”<sup>16</sup> as follows:

$$\text{Clinical net benefit} = [\text{reduction in event rate}] - [\text{treatment rate} \div NNT_T].$$

Because net benefit includes both treatments and recurrences, the optimal treatment strategy is the one with the highest net benefit, irrespective of the size of difference between strategies. Statistical analyses were conducted using Stata 9.2 (Stata Corp., College Station, TX).

## Findings

Baseline characteristics of the 4462 patients in our sample are summarized in Table 1. The median age at cystectomy was 62 years (interquartile range 51, 70) and the majority of patients were male (77%). Approximately half were either lymph node positive or had pathologic stage  $\geq$  pT3 (n=2466, 55%); 4081 (91%), 1835 (41%), 365 (8%) patients had a  $\geq$ 10%,  $\geq$ 25%, and  $\geq$ 70% 5-year nomogram probability of recurrence, respectively.

There were 1069 recurrence events. The median follow-up for recurrence-free patients was 3.8 years. The Kaplan-Meier estimate of the 5-year probability of recurrence for the entire sample was 28% (95% C.I. 26%, 29%).

Table 2 shows the number of patients treated using each treatment strategy, and the expected number of recurrences given various levels of treatment effectiveness. Compared to the standard approach, using a nomogram probability of 25% as eligibility for postoperative chemotherapy reduces the number of patients treated by nearly a quarter, at the cost of a slight increase in recurrence rates; using a 10% nomogram probability increases the number of patients treated by about 65%, but this is associated with a large reduction in recurrence rates. Table 3 shows the net benefit for each benefit given a range of  $NNT_T$  and relative risks. The strategy with the highest net benefit will have the optimal clinical results. Where drugs are very effective (relative risk of 0.6 or 0.7) or tolerable ( $NNT_T$  of 35 or 50), the highest net benefit is for a nomogram cut-off of 10%. Where drugs are of more marginal benefit and poor tolerability (relative risk of 0.9;  $NNT_T$  of 20), the nomogram cut-off of 70% is optimal. For the remaining scenarios a nomogram cut-off of 25% provides the highest net benefit. The differences between strategies are not trivial. For example, at a relative risk of 0.80 and an  $NNT_T$  of 20, use of a 25% threshold has a net benefit 0.003 higher than the standard pathological groups. This is equivalent to a strategy that led to 3 fewer recurrences per 1000 patients without any change in the number of patients given chemotherapy, or to one associated with 60 fewer treatments per 1000 patients without any increase in recurrence rates.

To explain these findings, figure 2 gives a distribution of predicted probabilities within each pathological grouping. A nomogram cut-off of 25% includes all patients with node-positive disease. However, a little more than a quarter of patients with pT3 or pT4, node-negative disease are at low risk (<25%) according to the nomogram and would therefore not be treated by chemotherapy; conversely, just under a quarter of patients with less than pT3 disease (organ confined tumors) are considered high risk by the nomogram.

We then repeated our analysis for all possible combinations of relative risk (0.50 to 0.99) and  $NNT_T$  (1 to 250) and recorded which strategy had the highest net benefit. The results are shown in figure 3. For example, table 3 shows that for a  $NNT_T$  of 20 and a relative risk of 0.8, the highest net benefit is obtained by treating only patients with at least a 25% risk of recurrence; a point corresponding to 0.8 on the  $x$  axis and 20 on the  $y$  axis of figure 3 is medium gray, indicating the 25% threshold as the optimal strategy. As expected, where adjuvant therapy is highly effective or tolerable, either all patients, or all patients except those with very low risk (nomogram probability less than 10%) should be treated; where therapy is of moderate effectiveness or tolerability, either no patients or only those at the highest risk (nomogram probability  $\geq 70\%$ ) should be treated. For the remaining cases, the 25% cut-off is optimal. Of note is that for no combination of effectiveness or tolerability is the highest net benefit associated with the standard eligibility for chemotherapy. In other words, the nomogram outperforms the standard guideline for every plausible scenario of drug effect and tolerability: whatever a clinician believes about a particular drug, he or she should use a nomogram cut-off in order to decide which patients to treat.

We excluded 2320 patients receiving chemotherapy or radiotherapy, or where other treatments were unknown; 19% (679/3551) with low stage ( $\leq$  pT2 and node negative); 32% (1536/4875) with high stage ( $\geq$  pT3 or node positive); and 16% (105/638) with unknown stage. Although a higher proportion of patients with high stage were excluded, our cohort comprised the vast majority of patients with both high and low pathologic stages. Accordingly, we see no reason to believe that treatment selection would have an important impact on our findings.

Our primary analyses involved an assumption that the relative risk reduction associated with treatment was constant across risk groups. We therefore performed sensitivity analyses to check that our results were robust to this assumption. Results of these sensitivity analyses

are displayed in table 4. Although it initially appears that conclusions about the optimal cut-point sometimes depend on our assumptions of how relative risk varies with absolute risk, the assumption of constant relative risk analyses gives good results: generally there were only small differences in net benefit between the optimal strategy under constant relative risk and the optimal strategy where relative risk varies by baseline risk. Most importantly, the strategy chosen under the assumption of constant relative risk was superior to standard pathological risk groups across almost but two of sensitivity analyses conducted, and in both cases the advantage was trivial (1 or 3 recurrences per 10,000 patients). Thus our conclusion that the nomogram improves clinical outcomes holds regardless of the relationship between absolute risk and relative risk reduction.

The nomogram was developed excluding patients who had received chemotherapy or radiotherapy – patients likely having the most unfavorable tumor characteristics. As a result, the predictions from the nomogram might underestimate the overall risk of recurrence. Therefore, as a second sensitivity analysis, we added a constant to the nomogram prediction for every patient (varying the constant from 2–5%) and repeated all analyses: none of our results were changed.

As a third sensitivity analysis, we repeated all analyses using only patients with transitional cell carcinoma, the predominant histology for bladder cancer in Europe and North America (see figure 3). Although there were some differences in results, our key finding was unaffected: one or other nomogram cut-off was superior to the standard eligibility criteria for chemotherapy for every combination of drug tolerability and effectiveness.

Our final sensitivity analysis accounted for the competing risk of death, that is, we defined the probability of recurrence by the cumulative incidence function instead of the Kaplan-Meier estimate<sup>21</sup>. The optimal treatment strategy under the competing risk analysis was the same as that under the primary analysis for all but one combination of NNT<sub>T</sub> and relative risk shown in Table 3.

## Interpretation

Adjuvant chemotherapy after bladder cancer is subject to some debate, in particular, data from randomized trials, though positive<sup>13, 14</sup>, is limited by inadequate number of patients. Nonetheless, adjuvant therapy is often given after radical cystectomy (it is recommended in standard guidelines<sup>15</sup>), and pathologic stage is the most common criterion to determine which patients receive it. Our analyses suggest that determining eligibility for adjuvant chemotherapy after radical cystectomy on the basis of a multivariable model will give superior clinical results compared to determining eligibility on the basis of pathologic characteristics alone. In a typical comparison, use of the nomogram would reduce the number of patients subjected to chemotherapy by 14%, with only a small increase in recurrences (0.4%). This improvement in outcome is obtained purely by changing a decision rule: there are no additional tests, procedures or treatments. The superior clinical performance of the nomogram appears to result from reclassification of an important proportion of patients: some patients with pT3 or pT4 disease would not be eligible for chemotherapy on the basis of the nomogram due to the absence of any other risk factor (such as a long time from diagnosis to treatment); comparably, some patients with <pT3 disease are eligible for chemotherapy by nomogram because they are defined as high risk for a reason other than pathologic stage.

Although previous studies have shown that multivariable models improve predictive accuracy compared to staging systems<sup>8–11</sup>, we believe that we are the first to show that use of such models would have beneficial effects with respect to a therapeutic decision. This

finding has important consequences for cancer care. An enormous range of decisions about the care of the cancer patient are based on risk, with patients thought to be a higher risk subject to more intensive treatment or monitoring. At the current time, however, nearly all such decisions are currently based on simple risk stratifications, such as stage. We hypothesize that multivariable risk prediction models could be used to replace many of the decisions currently made on the basis of stage, including whether a patient receives surgery, chemotherapy and radiotherapy; how aggressively he or she is treated; the intensity of post-treatment follow-up and eligibility for clinical trials. We have shown that changing the criteria we use to make decisions has important clinical consequences and so we further hypothesize that use of such models would improve cancer care either by decreasing the number of patients subject to unnecessary treatment or by decreasing event rates, as a result of better identification of patients requiring intervention.

Multivariable models have two additional advantages over crude risk stratifications based on criteria such as stage. First, multivariable models allow for individualization of care. Patients may differ with respect to the relative value they place on treatment toxicities and disease recurrence, and a model allows patients to vary with respect to the thresholds they use for treatment. Second, multivariable models allow for the addition of prognostic markers when and if they are shown to be of benefit. In the case of bladder cancer, for example, it has been suggested that markers such as cyclin E1, p53, p21, pRB, and p27 status might distinguish more from less aggressive tumors<sup>22, 23</sup>; similarly, genomic analyses have indicated that certain patterns of gene expression may be associated with cancer outcome<sup>24</sup>. Were these markers to be validated, it is unclear how they could be incorporated into a staging system without an unmanageable expansion of multiple categories (high stage / node negative / low p53 / low genomic risk; high stage / node negative / high p53 / low genomic risk; and so on). Conversely, such markers can be easily incorporated into multivariable models.

There are several possible limitations of our study. First, the same set of data was used to generate the predictive model as to assess it. That said, we do not believe that this results in statistical overfit, on the grounds that there are a very large number of events. Indeed, we conducted some preliminary analyses to estimate statistical “optimism”<sup>25</sup> and found that this was close to zero (e.g. using a cutoff probability of 25% and a relative risk of 0.8, the optimism for the net benefit from the nomogram was 0.0001). It may also be that patients treated in the community differ systematically from the patients treated at the academic centers contributing data to the nomogram. But even if this was true, it is not clear that this would favor either the nomogram or the standard criteria for chemotherapy. For example, if stage was poorly assessed in a community setting, or if node removal was less extensive, this would reduce the predictive accuracy of the nomogram, but it would also affect the predictiveness of a decision rule based on stage and nodal status alone.

As discussed, a clear limitation of this study is that, despite being NCCN guideline standard of care, adjuvant therapy for bladder cancer is not unequivocally considered to be of benefit. However, it can be argued that the quality of evidence for adjuvant therapy is not of direct relevance to our findings. This is unless one takes the position that adjuvant therapy does not and cannot possibly work for bladder cancer, and should never be subject to clinical trial. In the absence of such a position, a decision will have to be made about which patients receive adjuvant therapy, whether as a clinical decision rule or the eligibility criteria for a trial. Our figure 3 shows that irrespective of the characteristics of the adjuvant agent – whether its toxicity is high or low; whether it is of great or only moderate effectiveness – the predicted clinical results will be superior if the nomogram rather than standard pathologic criteria are used to determine which patients receive treatment.

Although desirable in principle, subjecting a nomogram-based strategy for chemotherapy referral to prospective experimental trial is of doubtful feasibility. For example, imagine that we wanted to show that a nomogram-based strategy would lead to a similar recurrence rate as the standard nomogram, but required fewer patients to be subjected to chemotherapy. A non-inferiority trial wishing to test that the nomogram did not increase recurrence rates by more than 1% might well need more than 25,000 patients.

This paper is presented in the spirit of “proof-of-principle”. We believe that we are the first to show clearly that use of a prediction model to inform decisions about chemotherapy would improve clinical outcome. That said, there are several steps that would need to be taken before the prediction model could be used in the clinic. First, we believe that the prediction model itself should be updated. The original bladder nomogram was developed for general use, rather than for the specific purpose of chemotherapy referral. As a result, the data set used to develop the nomogram includes patients that would not be considered for adjuvant chemotherapy, such as those of advanced age, or those diagnosed with squamous cell carcinoma. Second, in this paper we chose illustrative cut-offs for the nomogram (10%, 25% and 70%). The alternative would be choose optimal cut-offs for each of several scenarios of drug effectiveness and tolerability. Third, any statistical model would have to be implemented in a user-friendly format, perhaps in web version similar to Adjuvantonline7.

We have shown referring patients to chemotherapy on the basis of a multivariable model is likely to lead to better patient outcomes than use of pathological groups. Given the importance of this finding – that we can improve outcome merely by changing the basis on which we make decisions – we recommend research on other multivariable prediction models to determine their clinical effects.

## Acknowledgments

The following are members of the International Bladder Cancer Nomogram Consortium: Director, Bernard H. Bochner, MD; Co-Director, Guido Dalbagni, MD. Statistical Group: Michael W. Kattan, PhD (Director), Paul Fearn (bioinformatics coordinator), Kinjal Vora, Hee Song Seo, Lauren Zoref; Mansura University: Hassan Abol-Enein, Mohamed A. Ghoneim; Memorial Sloan-Kettering Cancer Center: Bernard H. Bochner, Guido Dalbagni, Peter T. Scardino, Dean Bajorin; University of Southern California: Donald G. Skinner, John P. Stein, Gus Miranda; Ulm University: Jürgen E. Gschwend, MD, Bjoern G. Volkmer, MD, Richard E. Hautmann, MD; Vanderbilt University: Sam Chang, Michael Cookson, Joseph A. Smith; University of Bern: George Thalman, Urs E. Studer; University of Michigan: Cheryl T. Lee, James Montie; David Wood; Fundació Puigvert: Juan Palou, Humberto Villavicencio, Antonio Rosales; Laval University: Yves Fradet, Louis LaCombe, Pierre Simard; Johns Hopkins Medical Center: Mark P. Schoenberg; Baylor College of Medicine: Seth Lerner, Amnon Vazina; University of Padova Medical School: PierFrancesco Bassi; Keio University: Masaru Murai, Eiji Kikuchi.

This research was funded by a P50-CA92629 SPORE from the National Cancer Institute. The funding source had no role in study design, data collection, analysis, interpretation, writing or the decision to submit. The authors have no financial conflicts of interest

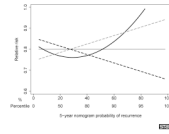
## References

1. Stephenson AJ, Scardino PT, Eastham JA, Bianco FJ Jr, Dotan ZA, DiBlasio CJ, et al. Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Clin Oncol.* 2005; 23(28):7005–7012. [PubMed: 16192588]
2. Kanemitsu Y, Kato T, Hirai T, Yasui K. Preoperative probability model for predicting overall survival after resection of pulmonary metastases from colorectal cancer. *Br J Surg.* 2004; 91(1): 112–120. [PubMed: 14716804]
3. Kattan MW, Karpeh MS, Mazumdar M, Brennan MF. Postoperative nomogram for disease-specific survival after an R0 resection for gastric carcinoma. *J Clin Oncol.* 2003; 21(19):3647–3650. [PubMed: 14512396]



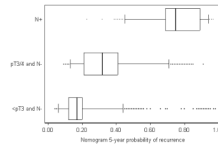
4. Van Zee KJ, Manasseh DM, Bevilacqua JL, Boolbol SK, Fey JV, Tan LK, et al. A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy. *Ann Surg Oncol*. 2003; 10(10):1140–1151. [PubMed: 14654469]
5. Karakiewicz PI, Briganti A, Chun FK, Trinh QD, Perrotte P, Ficarra V, et al. Multi-institutional validation of a new renal cancer-specific survival nomogram. *J Clin Oncol*. 2007; 25(11):1316–1322. [PubMed: 17416852]
6. Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Prognostic nomogram and index for overall survival in previously untreated patients with chronic lymphocytic leukemia. *Blood*. 2007; 109(11):4679–4685. [PubMed: 17299097]
7. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001; 19(4):980–991. [PubMed: 11181660]
8. Wong SL, Kattan MW, McMasters KM, Coit DG. A nomogram that predicts the presence of sentinel node metastasis in melanoma with better discrimination than the American Joint Committee on Cancer staging system. *Ann Surg Oncol*. 2005; 12(4):282–288. [PubMed: 15827679]
9. Peeters KC, Kattan MW, Hartgrink HH, Kranenbarg EK, Karpeh MS, Brennan MF, et al. Validation of a nomogram for predicting disease-specific survival after an R0 resection for gastric carcinoma. *Cancer*. 2005; 103(4):702–707. [PubMed: 15641033]
10. Ferrone CR, Kattan MW, Tomlinson JS, Thayer SP, Brennan MF, Warshaw AL. Validation of a postresection pancreatic adenocarcinoma nomogram for disease-specific survival. *J Clin Oncol*. 2005; 23(30):7529–7535. [PubMed: 16234519]
11. Kattan MW, Zelefsky MJ, Kupelian PA, Scardino PT, Fuks Z, Leibel SA. Pretreatment nomogram for predicting the outcome of three-dimensional conformal radiotherapy in prostate cancer. *J Clin Oncol*. 2000; 18(19):3352–3359. [PubMed: 11013275]
12. Bochner BH, Kattan MW, Vora KC. Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. *J Clin Oncol*. 2006; 24(24):3967–3972. [PubMed: 16864855]
13. Ruggeri EM, Giannarelli D, Bria E, Carlini P, Felici A, Nelli F, et al. Adjuvant chemotherapy in muscle-invasive bladder carcinoma: a pooled analysis from phase III studies. *Cancer*. 2006; 106(4):783–788. [PubMed: 16419069]
14. Adjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis of individual patient data Advanced Bladder Cancer (ABC) Meta-analysis Collaboration. *Eur Urol*. 2005; 48(2):189–199. discussion 99–201. [PubMed: 15939530]
15. Bennicke P. [Treatment of lumbar disc prolapse]. *Ugeskr Laeger*. 2008; 170(36):2804. author reply 04. [PubMed: 18780480]
16. Vickers AJ, Kramer BS, Baker SG. Selecting patients for randomized trials: a systematic approach based on risk group. *Trials*. 2006; 7:30. [PubMed: 17022818]
17. Begg CB, Cramer LD, Venkatraman ES, Rosai J. Comparing tumour staging and grading systems: a case study and a review of the issues, using thymoma as a model. *Stat Med*. 2000; 19(15):1997–2014. [PubMed: 10900448]
18. Sinclair JC, Cook RJ, Guyatt GH, Pauker SG, Cook DJ. When should an effective treatment be used? Derivation of the threshold number needed to treat and the minimum event rate for treatment. *J Clin Epidemiol*. 2001; 54(3):253–262. [PubMed: 11223323]
19. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *Bmj*. 1995; 311(7013):1145–1148. [PubMed: 7580713]
20. Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference: expanding the framework of clinical significance. *Med Decis Making*. 2005; 25(3):250–261. [PubMed: 15951453]
21. Gray R. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics*. 1988; 16:1141–1154.
22. Esrig D, Elmajian D, Groshen S, Freeman JA, Stein JP, Chen SC, et al. Accumulation of nuclear p53 and tumor progression in bladder cancer. *N Engl J Med*. 1994; 331(19):1259–1264. [PubMed: 7935683]

23. Shariat SF, Karakiewicz PI, Ashfaq R, Lerner SP, Palapattu GS, Cote RJ, et al. Multiple biomarkers improve prediction of bladder cancer recurrence and mortality in patients undergoing cystectomy. *Cancer*. 2008; 112(2):315–325. [PubMed: 18008359]
24. Blaveri E, Brewer JL, Roydasgupta R, Fridlyand J, DeVries S, Koppie T, et al. Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clin Cancer Res*. 2005; 11(19 Pt 1):7012–7022. [PubMed: 16203795]
25. Harrell, FE. *Regression Modelling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY: Springer; 2001.

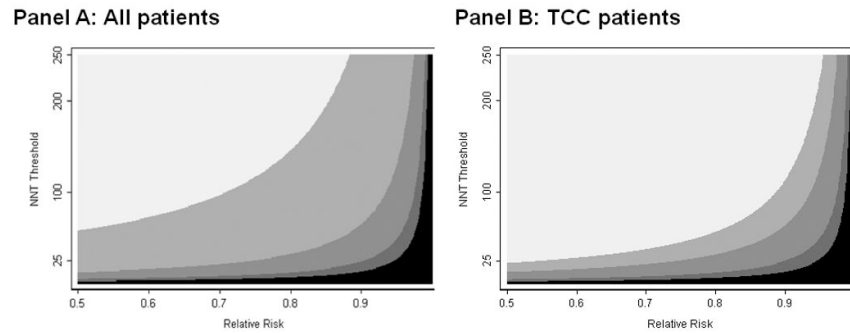


**Figure 1. Sensitivity analyses**

The figure illustrates how relative risk was varied under various assumptions for the relationship between absolute and relative risk. Gray solid line: constant relative risk; gray dashed line: treatment most effective for low risk patients; black dashed line: treatment most effective for high risk patients; black solid line: treatment most effective for average risk patients.



**Figure 2. Box plot of nomogram probabilities by pathologic grouping**  
The box gives the 25<sup>th</sup> – 75<sup>th</sup> centile, with the median given by the line inside the box. The outside lines give the 5<sup>th</sup> and 95<sup>th</sup> centiles, with outliers given by dots.



**Figure 3. Choice of optimal strategy for administering adjuvant chemotherapy to radical cystectomy patients. Panel A: All patients; Panel B: TCC patients**

The shaded areas identify the optimal strategy for each combination of NNT threshold and relative risk. White: administer to all radical cystectomy patients. Light grey: administer to patients with a 5-year probability of recurrence  $\geq 10\%$ . Grey: administer to patients with a 5-year probability of recurrence  $\geq 25\%$ . Dark grey: administer to patients with a 5-year probability of recurrence  $\geq 70\%$ . Black: administer to no patients (intervention does more harm than good). The specificity of the optimal strategy increases from top left to bottom right. Note that the conventional definition was inferior to all other strategies for every combination of NNT threshold and relative risk and therefore has no shaded region on the figure.

**Table 1**  
**Patient characteristics**

Characteristic	Frequency (%) or median (interquartile range)
Age at cystectomy (years)	62 (51, 70)
Elapsed time from diagnosis to cystectomy (months)	3.4 (1.2, 9.5)
5-year nomogram probability of recurrence (%)	21 (15, 36)
Gender	
Male	3457 (77%)
Female	1005 (23%)
Pathologic stage	
pT0	72 (2%)
pTis	124 (3%)
pTa	111 (2%)
pT1	765 (17%)
pT2	1035 (23%)
pT3	1878 (42%)
pT4	477 (11%)
Pathologic grade	
Low	706 (16%)
High	3676 (82%)
Unknown	80 (2%)
Histology	
TCC	3429 (77%)
SCC	853 (19%)
Adenocarcinoma	180 (4%)
Nodal status	
Negative	3668 (82%)
Positive	542 (12%)
Unknown	252 (6%)
Year of cystectomy	
Prior to 1985	359 (8%)
1985–1989	814 (18%)
1990–1994	999 (22%)
1995–1999	1401 (31%)
2000–2004	889 (20%)

**Table 2**  
**Percentage of patients treated under each treatment strategy, and the associated reduction in 5-year recurrence for various relative risks**

Sensitivity is the proportion of patients who recur by five years who would be treated by a particular treatment strategy; specificity is the proportion of patients recurrence free by five years who would be spared treatment by a particular treatment strategy.

Treatment strategy	Percentage of patients treated	Sensitivity	Specificity	Proportion of patients with event (reduction in event rate from treating none)		
				0.60	0.70	0.80
All	100%	100%	0%	16.8% (11.2%)	19.5% (8.4%)	22.3% (5.6%)
Standard	56%	78%	53%	19.2% (8.7%)	21.4% (6.5%)	23.6% (4.4%)
Nomogram cutoff 10%	92%	99%	11%	16.9% (11.1%)	19.6% (8.3%)	22.4% (5.5%)
Nomogram cutoff 25%	42%	72%	70%	19.9% (8.0%)	21.9% (6.0%)	23.9% (4.0%)
Nomogram cutoff 70%	9%	23%	97%	25.4% (2.6%)	26.0% (1.9%)	26.6% (1.3%)
None	0%	0%	100%	27.9%	27.9%	27.9%

**Table 3**  
**Net benefit of each treatment strategy per 1000 patients treated for various combinations**  
**of number-needed-to-treat threshold and relative risk**

Treatment Strategy	Net Benefit			
	Relative Risk			
	0.60	0.70	0.80	0.90
<b>NNT<sub>t</sub> 20</b>				
All	61.7	33.8	5.9	-22.1
Standard	59.3	37.5	15.7	-6.0
Nomogram cutoff 10%	64.7	37.1	9.4	-18.2
Nomogram cutoff 25%	59.6	39.5	19.4	-0.8
Nomogram cutoff 70%	21.4	15.0	8.6	2.1
<b>NNT<sub>t</sub> 35</b>				
All	83.1	55.2	27.3	-0.6
Standard	71.2	49.4	27.7	5.9
Nomogram cutoff 10%	84.4	56.7	29.1	1.4
Nomogram cutoff 25%	68.5	48.4	28.3	8.2
Nomogram cutoff 70%	23.2	16.8	10.4	4.0
<b>NNT<sub>t</sub> 50</b>				
All	91.7	63.8	35.9	7.9
Standard	76.0	54.2	32.4	10.7
Nomogram cutoff 10%	92.2	64.6	36.9	9.3
Nomogram cutoff 25%	72.1	52.0	31.9	11.8
Nomogram cutoff 70%	24.0	17.6	11.1	4.7



**Table 4**  
**Results of sensitivity analyses**

The different scenarios for how the effectiveness of treatment varies with baseline risk are illustrated in figure 1. Numbers given below the optimal treatment strategy represent the improvement in net benefit per 1000 patients treated if the optimal treatment strategy under the main analysis had been implemented instead of the conventional strategy under the respective sensitivity analysis.

Treatment Strategy	Optimal treatment strategy			
	Relative Risk			
	0.60	0.70	0.80	0.90
<b>NNT<sub>1</sub> 20</b>				
Main analysis – constant relative risk	10% Cutoff 5.4	25% Cutoff 1.9	25% Cutoff 3.6	70% Cutoff 8.2
Sensitivity analysis 1 – treatment most effective for those at average risk	All 3.7	25% Cutoff 2.4	25% Cutoff 4.0	70% Cutoff 11.9
Sensitivity analysis 2 – treatment most effective for those at low risk	All 4.0	25% Cutoff 2.4	25% Cutoff 4.1	70% Cutoff 13.1
Sensitivity analysis 3 – treatment most effective for those at high risk	10% Cutoff 6.8	25% Cutoff 1.5	25% Cutoff 3.1	25% Cutoff 3.4
<b>NNT<sub>1</sub> 35</b>				
Main analysis – constant relative risk	10% Cutoff 13.1	10% Cutoff 7.3	10% Cutoff 1.4	25% Cutoff 2.3
Sensitivity analysis 1 – treatment most effective for those at average risk	All 11.5	All 5.7	All -0.1	25% Cutoff 2.7
Sensitivity analysis 2 – treatment most effective for those at low risk	All 11.8	All 5.7	All -0.3	70% Cutoff 2.8
Sensitivity analysis 3 – treatment most effective for those at high risk	10% Cutoff 14.5	10% Cutoff 8.8	10% Cutoff 3.1	25% Cutoff 1.8
<b>NNT<sub>1</sub> 50</b>				
Main analysis – constant relative risk	10% Cutoff 16.2	10% Cutoff 10.4	10% Cutoff 4.5	25% Cutoff 1.1
Sensitivity analysis 1 – treatment most effective for those at average risk	All 14.6	All 8.8	All 3.0	All 1.5
Sensitivity analysis 2 – treatment most effective for those at low risk	All 14.9	All 8.8	All 2.8	All 1.6
Sensitivity analysis 3 – treatment most effective for those at high risk	10% Cutoff 17.6	10% Cutoff 11.9	10% Cutoff 6.2	25% Cutoff 0.6