Published in final edited form as:

Urol Oncol. 2017 January ; 35(1): 35.e1–35.e5. doi:10.1016/j.urolonc.2016.07.010.

The Difficulty in Selecting Patients for Cytoreductive Nephrectomy: An Evaluation of Previously Described Predictive Models

Brandon J Manley¹, Daniel M. Tennenbaum¹, Emily A. Vertosick², James J. Hsieh³, Daniel D. Sjoberg², Melissa Assel², Nicole E. Benfante¹, Seth A Strobe⁴, Eric Kim⁴, Jozefina Casuscelli¹, Maria F Becerra¹, Jonathan A. Coleman¹, A. Ari Hakimi¹, and Paul Russo¹

¹Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY

²Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY

³Department of Medical Oncology, Memorial Sloan Kettering Cancer Center, New York, NY

⁴Department of Surgery, Division of Urology, Washington University in St. Louis, St. Louis, MO

Abstract

Purpose—To externally evaluate a preoperative points system and a preoperative nomogram, both created to assess time to death after cytoreductive nephrectomy.

Materials and Methods—We identified 298 patients who underwent cytoreductive nephrectomy at our institution, a tertiary cancer center, between 1989 and 2015. To validate the points system, we compared reported overall survival (OS) for each criterion to observed OS in our cohort. To evaluate the nomogram, we prognosticated risk of death at 6 months after surgery for 280 patients with sufficient follow up in our cohort and evaluated discrimination using area under the curve (AUC) and calibration. Decision curve analysis was performed to assess clinical utility of the nomogram.

Results—Significant differences in OS were observed between patients with and without five of seven criteria on univariate analysis: low albumin ($p < 0.0001$), high lactate dehydrogenase ($p = 0.002$), liver metastasis ($p = 0.004$), retroperitoneal lymphadenopathy ($p = 0.002$) and supradiaphragmatic lymphadenopathy ($p = 0.019$). Discrimination from the preoperative model predicting death within 6 months of surgery was lower in our cohort (AUC 0.65; 95% CI 0.52, 0.79), than the original publication (AUC 0.76). Decision curve analysis demonstrated little benefit for applicability.

*Corresponding author: Paul Russo, MD, Department of Surgery, Urology Service, Sidney Kimmel Center for Prostate and Urologic Cancers, Memorial Sloan Kettering Cancer Center, 353 East 68th Street, New York, NY 10065, Phone: 1-646-422-4432, russop@mskcc.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—Five previously defined risk factors are predictive of decreased OS after cytoreductive nephrectomy in our cohort. We found lower discrimination using the preoperative model and minimal clinical utility according to decision analysis in our study cohort. These findings suggest the need for improved models to aid patient stratification and consequent treatment choice.

Keywords

Renal Cell Carcinoma; Neoplasm metastasis; Nephrectomy; Prognosis; Mortality

Introduction

On the basis of two landmark randomized controlled trials, most patients with metastatic renal cell carcinoma (mRCC) are advised to have a cytoreductive nephrectomy (CN) as part of their treatment.^{1–3} Despite this level of evidence, a recent multicenter randomized trial of sunitinib for advanced renal cell carcinoma demonstrated that 21% of participants had not had a prior nephrectomy.⁴ This apparent discrepancy in clinical practice suggests that uncertainty exists among urologists and medical oncologists regarding the continued benefit of CN in the era of targeted therapy.

To better differentiate the potential benefits of CN from more current therapies used for mRCC, two large Phase III randomized clinical trials were opened (CARMENA and SURTIME) and are currently recruiting participants. Unfortunately, these studies have had difficulty with accrual and the results will only apply to patients meeting the inclusion criteria which tend to favor healthy patients with no laboratory abnormalities.^{5, 6} As many patients fail to meet the eligibility criteria, questions regarding their optimal treatment remain.

Multiple studies have attempted to identify clinical characteristics or develop predictive algorithms to better select patients who will benefit from CN.^{7–10} In 2010 Culp et al described seven pretreatment risk factors that were negatively prognostic of outcome in patients being considered for CN: serum albumin below the lower limit of normal, serum lactate dehydrogenase (LDH) above the upper limit of normal, radiographic evidence of retroperitoneal or supradiaphragmatic lymphadenopathy, liver metastasis, symptomatic metastasis on presentation, and clinical T3 disease or greater (in which the tumor is growing into a major vein or into tissue around the kidney or Gerota's fascia).^{7, 11} Later, Margulis et al expanded the study cohort to include more recently treated patients and then used those data to generate two predictive models to aid clinicians and patients in deciding whether to include CN in their treatment.¹²

These models are based on readily available clinical parameters; however, these risk factors and proposed models have never been validated in an independent cohort of patients. To evaluate the clinical utility of the seven risk factors described by Culp and the preoperative predictive model described by Margulis, we applied these factors and the model to our own institutional cohort of mRCC patients who underwent CN.

Patients and Methods

Study Population

After institutional review board approval of our study protocol, we identified 363 patients who had undergone CN at our institution between 1989 and 2015. At our institution, Memorial Sloan Kettering Cancer Center (MSKCC), almost all mRCC patients, except those deemed unfit for surgery or with widely metastatic disease, are considered for CN; the decision to perform CN is made jointly by the individual surgeon, the medical oncologist, and the patient. Patients were excluded from our study if they had received systemic therapy before surgery (N=21), did not have available radiology or clinical data (N=19), had a prior nephrectomy (N=12), had a history of another metastatic cancer (N=7), were younger than 18 years (N=3), had a non-RCC kidney cancer (N=2), or had cryoablation (N=1), leaving 298 patients in our cohort.

Study Design

The primary aim of our study was to independently evaluate previously described predictive risk factors and a perioperative model created using a cohort of patients from MD Anderson Cancer Center (MDACC) to identify those at an increased risk of death after CN. A points system based on seven preoperative risk criteria has been proposed,⁷ as well as a preoperative nomogram for risk of death by 6 mo after surgery.¹² The cohort used to create the points system included all MDACC patients undergoing cytoreductive nephrectomy between 1991 and 2007, with the cohort expanded to include patients treated in 2008 for the preoperative model. Patients with a previous history of treated RCC were excluded, and for the predictive models those patients without sufficient follow-up were excluded.

We first evaluated the points system created to identify patients suitable for CN. In this system, patients are assigned one point for each of the following seven criteria: clinical T3 disease or higher; symptoms of metastasis at presentation; presence of liver metastasis; retroperitoneal lymphadenopathy; supradiaphragmatic lymphadenopathy; LDH > 618; and albumin < 3.5. We assigned a point total to 298 evaluable patients and calculated the median overall survival (OS) for each point total (0–7), then compared these survival estimates to the predictions reported in the original publication and created a calibration plot (Fig. 1).⁷

Statistical Analysis

A number of patients in our cohort were missing data on LDH, clinical T classification, and albumin. Patients who were evaluated preoperatively only by a surgeon were often missing LDH data, as surgeons at our institution order this test less commonly than medical oncologists do. Patients who had surgery before 1996 were often missing albumin measurements. We used multiple imputation by chained equations to impute values for the following variables: elevated lactate dehydrogenase (LDH) (N=131), clinical T3 disease or higher (N=70), low albumin (N=42), pathologic N classification (N=28), undefined race (N=8), liver metastasis (N=5), retroperitoneal lymphadenopathy (N=3), supradiaphragmatic lymphadenopathy (N=3), metastatic symptoms at presentation (N=3), and pathologic classification T3 or higher (N=2). This method of multiple imputation fills in missing values in multiple variables iteratively by using chained equations and accommodates arbitrary

missing-value patterns¹³. Statistical analyses were performed utilizing the measured and imputed values combined across 10 imputations using Rubin's rules. As a sensitivity analysis, we repeated these analyses using an additional imputed dataset and the original, non-imputed data. All analyses were conducted using Stata 13 (Stata Corporation, College Station, TX).

We then evaluated the preoperative nomogram predicting the risk of death at 6 and 12 mo, respectively, after CN. In our study, 280 patients were included in the analysis of the preoperative model as some patients were lost to follow-up before the endpoint could be determined. We reported model discrimination using the area under the curve (AUC) and created calibration plots (Fig. 2). Since there have not been previous studies regarding an optimal risk cut-point for CN, after discussion among several surgical experts we chose a 20% threshold for risk of death after CN, meaning that patients with a greater than 20% risk of death would not be recommended for surgery. This cut-point was used to assess clinical utility of the preoperative model, as this model could be used for counseling about whether or not to perform surgery. We recorded the number of patients who had a risk of death at 6 mo after surgery of less than 20% but died during that time period. We also used decision curve analysis to assess the clinical utility of the model for risk thresholds up to 30%.

Results

Our cohort consisted of 298 patients who underwent CN for mRCC at MSKCC with a median followup of 1.9 years (interquartile range [IQR] 0.6 – 5.5 years) between 1989 and 2015. Patient demographics and disease characteristics are reported in Table 1. We first sought to determine the validity of the points system based on seven preoperative prognostic risk factors. The prevalence of these seven risk factors in our cohort is reported in Table 2.

Using a univariate Cox proportional hazard model, a significant difference in OS was observed for five of the seven risk factors on univariate analysis: low albumin (hazard ratio [HR] 3.83, 95% confidence interval [CI] 2.15–6.81, $p < 0.0001$), high LDH (HR 23.76, 95% CI 3.09–182.8, $p = 0.002$), liver metastasis (HR 1.98, 95% CI 1.26–3.11, $p = 0.003$), retroperitoneal lymphadenopathy (HR 1.59, 95% CI 1.20–2.11, $p = 0.001$) and supradiaphragmatic lymphadenopathy (HR 1.41, 95% CI 1.03–1.94, $p = 0.032$). There was some evidence of difference in OS between patients with clinical T3 or higher disease and those without, but this did not reach conventional levels of statistical significance (HR 1.31, 95% CI 0.98–1.77, $p = 0.072$). We found no evidence of difference in OS between patients who presented with symptomatic metastases and those who were asymptomatic (HR 1.15, 95% CI 0.85–1.55, $p = 0.4$). Calibration between the median OS in our cohort and the median OS reported in the Culp cohort was generally good (Fig 1). OS was slightly higher in the MSKCC cohort, with the exception of 10 patients who had each scored 4 points whose actual OS was notably lower than predicted (Fig. 1). When non-imputed data were used (data not shown), the points system had similar calibration with the exception of the 20 patients (17%) with 0 points, who were found to have worse median OS (30.8 mo, 95% CI 27.8–50.6) than the OS predicted by the MDACC cohort (40.6 mo).

Discrimination from the preoperative model predicting death within 6 mo of surgery was much lower in this cohort (AUC 0.65, 95% CI 0.55–0.77), than in the MDACC cohort (AUC 0.76). In our cohort, applying the preoperative model underestimated the true risk of death (Fig. 2). Using a risk threshold of 20%, 277 patients (99%) were identified as low risk; 28 of these “low-risk” patients died within 6 mo of surgery, for a misclassification rate of 10%. Based on the decision curve analysis, we found little benefit in using the preoperative model, as net benefit is only higher than alternatives for a very small range of threshold probabilities (Fig. 3).

Given the amount of missing data in our cohort, we did two sensitivity analyses to assess whether our results were affected by this missing data or the imputations performed. We first used a different type of imputation, imputing only LDH without chained equations, since this variable had the highest number of missing values. Results were similar to the original imputed dataset. We then repeated the analysis of these models using only available, non-imputed patient data. When evaluating the criteria for the points system, results were consistent with the analysis of imputed data. The discrimination of the preoperative model was similar when using non-imputed data (AUC 0.66, 95% CI 0.46–0.86), although confidence intervals were wider due to limited sample size. While calibration of the preoperative model was fairly better using the non-imputed data, this did not improve the clinical utility of the model.

Discussion

In our analysis of the points system, we were able to demonstrate that five of the seven risk factors were significantly associated with OS in our patient cohort. We could also validate that increasing the number of risk factors for a patient correlated with a decreasing rate of estimated OS. Two risk factors, clinical T3 disease or higher and metastatic symptoms at presentation, were found to be less predictive in our cohort compared to the original study.

In regard to the predictive models, we found a lower discrimination (AUC 0.65 vs 0.76) for the preoperative model in our cohort compared to original reported cohorts. Using decision curve analysis we found that the preoperative model generally added little clinical utility for patient management.

Several reasons may account for the differences between the results of our study and the results reported by Culp et al and Margulis et al. A higher percentage of patients in our cohort were treated postoperatively with targeted therapies, which became first-line standard of care for mRCC in 2006. The Culp cohort underwent CN between 1991 and 2007, the Margulis cohort between 1991 and 2008; our study covers 1989–2015. Moreover, in our cohort a higher percentage of patients presented with 3 or fewer risk factors than in Culp et al’s (96.3% vs. 89.2%), while fewer of our patients had four or more risk factors (3.7% vs. 10.8%). Similarly, rates of transfusion were lower in our cohort (95 of 298, 31.9%) than the cohort reported by Margulis et al (289 of 601, 48.1%). The median LDH in the Margulis et al cohort was 460 U/L(interquartile range [IQR] 392–586) with a range of values from 190 to 7797 U/L. In our cohort, median LDH was 175 U/L(IQR 143–202) with a range of values from 90 to 702 U/L. The smaller range of LDH values may have limited our ability to

accurately predict differences among patients in our cohort and led to the lower discrimination using both the preoperative and postoperative models in our cohort.

Our study is not without limitations. We excluded patients with insufficient data or follow-up and this restricted cohort of patients may limit our ability to generalize our results to our target population, i.e. all patients with mRCC. Our results are from a tertiary care center, and may not reflect the full spectrum of mRCC patients seen in community practice. Also, the treating physician's perception of the patient's prognosis rather than the actual severity of their disease may have limited what treatment patients were offered. The subjectivity of what is deemed "unresectable" will vary to some extent by surgeon, and the judgment by medical oncologist of "widely metastatic" or "rapidly progressive" also could fluctuate between physicians and even institutions. These variables are difficult to quantify but doubtless do have an influence on which patients undergo CN. The developments of predictive nomograms are heavily dependent on the initial cohort in which it was developed. Even when applying a nomogram to a seemingly similar external patient cohort, it is fairly common to find the nomogram has inferior performance compared to the developmental cohort. Lastly, confounding by indication is commonly present in most retrospective studies that center on the role of CN and may make interpretation and applicability of study results arduous.

The difficulty in our attempts as clinicians to better select patients for CN likely stems from several issues. First, as mentioned earlier, the nature of mRCC as a terminal illness can bring about an extreme variance in the treatment of patients. Younger, healthier patients may follow a more aggressive treatment algorithm even in the face of clinical parameters pointing towards a poor outcome, and patients with a large disease burden and medical comorbidities may pursue more conservative therapy despite having favorable disease characteristics. These choices are dependent on patient and physician factors, many of which are not able to be accurately captured or measured.

Another issue centers on the rapidly expanding armamentarium of systemic treatments available to patients with mRCC. Since the introduction of targeted therapy, the median overall survival of mRCC patients has improved dramatically but the performance of CN also seems to be declining according to a recent analysis of the SEER registry.^{14, 15} As we stand at the dawn of personalized medicine, the genomic sequencing of tumors has created a pipeline that promises to bring with it many more therapeutic options for patients.¹⁶ The recent addition of a new class of systemic therapy, immune checkpoint inhibitors, seems primed to not only improve OS in mRCC patients but also add to the complexity in the ways we try to optimally manage the disease.^{17, 18} This evolution in medical treatment for mRCC has quickly outpaced our ability to organize and accrue patients for randomized controlled trials, making even successful trials quickly obsolete. In trying to place the appropriate use of CN in the treatment paradigm, we are looking to study a rapidly moving target. We believe improved incorporation of real-time genetic and immunological data on patients and in clinical trials will be needed to best define the role of CN in the future.

Conclusions

Five previously defined risk factors were found to be predictive of decreased OS after CN in our study. We found lower discrimination using the preoperative model and minimal clinical utility according to decision analysis in our study cohort. These findings suggest the need for improved models to aid patient stratification and consequent treatment choice among mRCC patients.

Acknowledgments

Supported by the Sidney Kimmel Center for Prostate and Urologic Cancers and the NIH/NCI Cancer Center Support Grant P30 CA008748 and Ruth L. Kirschstein National Research Service Award T32CA082088.

References

1. Flanigan RC, Mickisch G, Sylvester R, et al. Cytoreductive nephrectomy in patients with metastatic renal cancer: a combined analysis. *J Urol*. 2004; 171:1071. [PubMed: 14767273]
2. Flanigan RC, Salmon SE, Blumenstein BA, et al. Nephrectomy followed by interferon alfa-2b compared with interferon alfa-2b alone for metastatic renal-cell cancer. *N Engl J Med*. 2001; 345:1655. [PubMed: 11759643]
3. Mickisch GH, Garin A, van Poppel H, et al. Radical nephrectomy plus interferon-alfa-based immunotherapy compared with interferon alfa alone in metastatic renal-cell carcinoma: a randomised trial. *Lancet*. 2001; 358:966. [PubMed: 11583750]
4. Motzer RJ, Hutson TE, Olsen MR, et al. Randomized phase II trial of sunitinib on an intermittent versus continuous dosing schedule as first-line therapy for advanced renal cell carcinoma. *J Clin Oncol*. 2012; 30:1371. [PubMed: 22430274]
5. Clinical Trial to Assess the Importance of Nephrectomy (CARMENA). US National Institute of Health website. 2015; 2015
6. Immediate Surgery or Surgery After Sunitinib Malate in Treating Patients With Metastatic Kidney Cancer (SURTIME); U.S. National Institutes of Health website. 2015; 2015
7. Culp SH, Tannir NM, Abel EJ, et al. Can we better select patients with metastatic renal cell carcinoma for cytoreductive nephrectomy? *Cancer*. 2010; 116:3378. [PubMed: 20564061]
8. Leibovich BC, Cheville JC, Lohse CM, et al. A scoring algorithm to predict survival for patients with metastatic clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *J Urol*. 2005; 174:1759. [PubMed: 16217278]
9. Sakai I, Miyake H, Hinata N, et al. Improved survival in patients with metastatic renal cell carcinoma undergoing cytoreductive nephrectomy in the era of targeted therapy. *Int J Clin Oncol*. 2014; 19:674. [PubMed: 23989537]
10. You D, Jeong IG, Ahn JH, et al. The value of cytoreductive nephrectomy for metastatic renal cell carcinoma in the era of targeted therapy. *J Urol*. 2011; 185:54. [PubMed: 21074811]
11. Edge, SB.; B, D.; Compton, CC., et al., editors. *AJCC Cancer Staging Manual*. 7th. New York, NY: Springer; 2010.
12. Margulis V, Shariat SF, Rapoport Y, et al. Development of accurate models for individualized prediction of survival after cytoreductive nephrectomy for metastatic renal cell carcinoma. *Eur Urol*. 2013; 63:947. [PubMed: 23273681]
13. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007; 16:219. [PubMed: 17621469]
14. Choueiri TK, Motzer RJ. The Right Drug for the Right Patient: Navigating Systemic Therapy Options in Metastatic Renal Cell Carcinoma and Future Directions. *J Natl Compr Canc Netw*. 2015; 13:1168. [PubMed: 26483056]
15. Conti SL, Thomas IC, Hagedorn JC, et al. Utilization of cytoreductive nephrectomy and patient survival in the targeted therapy era. *Int J Cancer*. 2014; 134:2245. [PubMed: 24135850]

16. Wei EY, Hsieh JJ. A river model to map convergent cancer evolution and guide therapy in RCC. *Nat Rev Urol*. 2015
17. Choueiri TK, Escudier B, Powles T, et al. Cabozantinib versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med*. 2015
18. Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med*. 2015

Standard Abbreviations

OS	Overall survival
AUC	Area under the curve
RCC	Renal cell carcinoma
mRCC	Metastatic renal cell carcinoma
CN	Cytoreductive nephrectomy
LDH	Lactate dehydrogenase
MDACC	MD Anderson Cancer Center
MSKCC	Memorial Sloan Kettering Cancer Center
IQR	Interquartile range
HR	Hazard ratio
CI	Confidence interval

Highlights

1. 5 of 7 preoperative risk factors were predictive of overall survival
2. The complete model had an AUC of AUC 0.65; 95% (CI 0.52, 0.79) in our cohort
3. Decision curve analysis demonstrated little benefit for applicability
4. Differences in patient selection and pathology may explain differences

Figure 1.

Calibration between overall survival in months for MSKCC patients and the original MDACC cohort, by imputed point total (N=298). For example, the median observed overall survival among MSKCC patients with 0 points was 45 mo (95% CI 28–55), compared to the predicted median overall survival in MDACC patients, reported as 41 mo.

Figure 2. Calibration plot for preoperative nomogram using imputed data to predict overall survival at 6 mo after surgery (N=280).

Figure 3.

Decision curve for preoperative model predicting death within 6 mo after surgery. The blue line represents treating all patients. The red line represents treating no patients. The green line represents treating patients based on their predicted risk of death within 6 mo.

Table 1

Demographic and clinical characteristics of eligible patients, N=298. Data are presented as median (IQR) or frequency (percent).

Characteristic	No. of patients
Male	210 (70%)
Age at surgery, yr	61 (53, 69)
White race (n=290)	266 (92%)
<i>Imputed</i>	274 (92%)
Disease presentation	
Incidental	130 (44%)
Local	112 (38%)
Systemic	53 (18%)
Unknown	3 (1%)
ASA III/IV	163 (55%)
High grade disease	231 (88%)
Preoperative albumin (n=256)	4.2 (4.0, 4.5)
<i>Imputed</i>	4.2 (4.0, 4.4)
Preoperative LDH (n=167)	177 (156, 195)
<i>Imputed</i>	177 (155, 198)

IQR=interquartile range, LDH=lactate dehydrogenase.

Table 2

Imputed criteria for preoperative points system and total points, N=298.

Variables	No. of patients (%)
Clinical T3 disease or higher	130 (44%)
Metastasis symptoms at presentation	98 (33%)
Presence of liver metastasis	28 (9.4%)
Retroperitoneal lymphadenopathy	97 (33%)
Supradiaphragmatic lymphadenopathy	74 (25%)
High LDH	1 (0.3%)
Low albumin	14 (4.7%)
Total points	
0	62 (21%)
1	119 (40%)
2	75 (25%)
3	31 (10%)
4	10 (3.4%)
5	1 (0.3%)
6	0(0.0%)
7	0 (0.0%)

IQR=interquartile range, LDH=lactate dehydrogenase.