Methodological Review by Joseph V. Terza, PhD

Article Reviewed


Review

In this critical review I focus on the study design, methodology and statistical analysis and address the following three key questions: 1) Are the results valid? 2) What are the results? 3) Can they be applied to patient care? I will begin by describing the objectives of the study and answering the second of these questions by summarizing the results vis-à-vis these objectives. Next I will critique the methods and in that context address the first question regarding the validity of the results. Finally, I will suggest how results from models of this type can be applied to patient care.

Objectives of the study

The overarching objective of the study was to compare the effectiveness of radical cystectomy as a treatment for invasive bladder cancer to two alternative therapeutic approaches – chemotherapy/radiation (separately or in combination) or surveillance. The study finds that survival, as measured by the hazard ratio of death and 2- and 5-year survival, was better for those who underwent cystectomy compared with those who underwent the other treatments. There were two secondary goals of the research. The first of these was to document the marked underuse of guideline-recommended radical cystectomy for invasive bladder cancer. Only 21% of their sample underwent radical cystectomy indicating vast underuse. The authors also sought to identify the factors that are significantly (in a statistical sense) associated with treatment choice. Age (-), comorbidity (-), cancer grade (+), and travel distance to cystectomy provider (-) were all found to be important factors associated with treatment choice.

Methods

As best I can tell, the method used by the authors comprises four components:

1) Apply a two-level (logistic?) regression model, in which level 1 is person-specific and level 2 corresponds to the diagnosing physician, and perform a stepwise analysis to determine: a) covariate effects on whether or not the patient undergoes radical cystectomy surgery; and b) the variables to be included in the multinomial first stage of the two-stage residual inclusion (2SRI) analysis of the effect of bladder cancer treatment on survival.

2) Use a trinomial (logit?) model to estimate the determinants of treatment choice (chemo/radiation vs. cystectomy vs. surveillance). The results of this analysis are used to compute the residuals for the first-stage of the 2SRI estimator. The 2SRI estimator requires that at least one instrumental variable (IV) be included as a regressor in this component of the model. The authors used travel distance to the cystectomy provider as the IV.

3) Use a Cox proportional hazards (PH) model to perform a stepwise analysis to determine: a) the effect of treatment on patient survival (death hazard); and b) the variables to be included in the second stage
2SRI survival analysis of patient treatment outcomes. In this Cox proportional hazards analysis, there is no account taken of unobserved confounding (endogeneity) -- i.e. this is conventional survival analysis without the inclusion of an IV.

4) As the second stage of the 2SRI protocol, use a Weibull survival model to estimate the effect of treatment on patient survival (death hazard) that is adjusted for both observable and unobservable confounders. At this stage, the residuals from the multinomial model (component 3 above) are included as additional regressors in the survival model.

Queries

1) Why the four-component approach? In particular, why not combine components (1) and (2), as they are redundant? Because, ultimately, in the 2SRI context, the treatment choice analysis will be conducted using the trinomial (logit?) model, there is no need for a separate multilevel binary (logistic?) regression. In fact, such an analysis ignores the fact that treatment choice is trichotomous (not binary) -- a feature that trinomial logit fully accommodates. Moreover, the stepwise procedures implemented in component (2) above, could also be applied in the multinomial regression context (although I would recommend against stepwise analysis in this context -- see (4) below).

2) Similarly, why not combine components (3) and (4), as they too are redundant. There is no need to use a different survival model to achieve the goals of component (3). These could have been accomplished using the Weibull 2SRI second-stage results. Alternatively, Cox proportional hazard analysis could have been used in the second stage of 2SRI. To discern the differences in the treatment effect estimates due to accounting for unobservable confounding (via 2SRI), you could simply estimate the survival model (whichever you chose -- Cox PH or Weibull) without residual inclusion. Use Cox PH or Weibull but not both.

3) Why aren’t the trinomial (logit?) results reported and discussed? It would seem that these results are more relevant than the binary odds-ratio results obtained from the multi-level model. The trinomial (logit?) results take full account of the trichotomous nature of the treatment choice. Moreover, it is the trinomial results that are relevant in the determination of the “strength” of the IV. A Wald test of the joint significance of the estimated coefficients of the distance variable in the multinomial model should be used as a check of the strength of the IV [see (2) under “Critiques” below].

4) I am not a fan of stepwise regression because it typically makes inference (both finite and asymptotic) difficult. Difficulty arises because pre-test outcomes must be taken into account in the derivation of the relevant sampling distribution to be used for inference (i.e., hypothesis testing and confidence interval estimation). This will typically complicate such derivations. That’s why this issue is usually ignored, which leads to biased inference. Do you really need stepwise analysis here? It seems that the regression specifications are based on sound conceptual modeling. Is it really necessary to winnow the set of variables to avoid increased regression variance? Are there that many potentially relevant candidate variables? What happened when you tried to estimate the model with all of them included? Do the results (particularly the standard errors) differ very much?

5) Why the focus on hazard ratios? The main outcome of interest here is survival post diagnosis. Why not base all treatment effect measures on expected survival [see (1) under “Critiques” below]?
6) How exactly were the hazards in Table 3 computed? Are they averages across the sample at the estimated parameter values or are they estimated at the average values of the covariates? This should be made clear.

Critiques

1) The authors do not define a specific evidence-based analytic objective at the outset. This is a very important first step in any empirical study. For example, it would have been reasonable if they had specified the estimation of the treatment effects of the various therapeutic options on average patient survival time as the study objective. Such population average treatment effects are rigorously defined (mathematically) and have relatively easy to implement (and easy to interpret) corresponding estimators. The authors discuss estimated hazard rates but do not specifically define the underlying analytic objective of interest. This makes interpretation of the tabled results difficult. Clear definition of the estimation objectives and explicit discussion of how the survival regression results would be implemented in the estimation of these objectives would solve this problem.

2) The odds ratio analysis (multilevel logistic regression?) described above in (1) under “Methods” should be eliminated and replaced by trinomial logit analysis which will also serve as the first stage estimator for 2SRI (this may be the method used by the authors for the first stage of the 2SRI estimator; this is not made clear in the paper). Two important points support this approach. First, the trinomial logit framework affords a direct way to test the “strength” of the IV (travel distance to cystectomy provider) [see Stock and Watson, 2003, p. 349 for an introductory discussion of IV strength]. Secondly, the “effects” of the remaining covariates are easy to characterize and statistically test (an important secondary goal of the study) in the trinomial logit framework.

Validity of the results

Given the lack of specificity regarding the estimation objectives and the methods used, it is difficult to assess the validity of the results. The main stated objective of the paper was to compare the survival outcomes of patients with invasive bladder cancer by the treatment they received. To this end, the authors report two sets of results. In Figure 1, they plot hazard functions that are adjusted neither for observed nor unobserved covariates. It is not clear what should be concluded from these results. In Table 3, presumably, covariate adjusted hazard rates are reported for both Cox proportional hazards (adjusted for observed covariates) and Weibull 2SRI (adjusted for observed and unobserved covariates) along with similarly adjusted estimates of 2-year and 5-year survival probabilities. There are, however, no details given regarding exactly what version of the hazard rate is being estimated and how these hazard rates were adjusted for the covariates. A similar criticism of the estimated survival rates holds. The results show that death hazards are smaller and survival probabilities are larger for cystectomy vs. the other two treatments. Without the aforementioned modeling details, however, the validity of these results is difficult to evaluate.

The secondary goal of the study was to identify the statistical determinants of the use of cystectomy as a treatment for invasive bladder cancer. A multi-level odds ratio (logistic regression?) analysis was implemented for this purpose. Predictive margins were also estimated. The results in Table 2 appear to be valid but more detail on the underlying analyses would have been helpful—especially with regard to the definition and estimation of the predictive margins. As I discussed earlier, this binary analysis is redundant and a bit incongruent with the ultimate (and in my opinion appropriate) trinomial analysis
implemented in the first stage of the 2SRI estimation. All forms of information gleaned from the binary model underlying Table 2 could have been obtained from the more appropriately modeled trinomial logit framework.

Applicability of results to patient care

The results obtained from the study are applicable to patient care insofar as they: 1) demonstrate population wide under-utilization of cystectomy vis-à-vis guideline-recommended care for invasive bladder cancer; and 2) are indicative of improved outcomes at the population level for those who have undergone cystectomy compared to the alternative treatments – chemotherapy/radiation and surveillance. It would have greatly enhanced the applicability of the analyses to patient care if the authors could have shown how the results might be used to predict outcomes under the various treatments for a particular covariate profile. Such capability would be useful for taking account of patient heterogeneity in a clinical setting. For example, in the average treatment effects approach, suggested above in (1) under “Critiques”, the estimation results can be used to predict differential survival durations for a particular patient profile. One might then rank the treatments with respect to expected survival for the patient profile in question as

   cystectomy > chemo/radiation > surveillance.

In this way, the model and estimation results might be used as an aid in patient/physician decision making, while taking account of patient heterogeneity. Without more methodological detail, however, it is impossible to see how the models and results put forth in the paper might be implemented in this way.

References

Clinical Review by Brent Hollenbeck, MD

Article Reviewed


Review

Although clinical trials comparing treatments are lacking, guidelines favor radical cystectomy with urinary diversion as the preferred therapy for muscle invasive bladder cancer. However, prior work suggests that this modality may be underused. This study uses data from a clinical cancer registry (SEER) linked to billing claims (Medicare data) to identify factors associated with the use of cystectomy as well as to compare survival outcomes of the various treatment approaches (cystectomy vs. chemotherapy and/or radiation vs. surveillance). Given its nature, this study was limited to those 66 years and older with insurance coverage and who were diagnosed between 1992 and 2002.

There are several notable findings:

- Only 678 (21%) of 3262 patients with muscle invasive (stage II) bladder cancer underwent cystectomy. Perhaps even more importantly, 2132 (65%) of patients with a Charlson score of 0 or 1 (who have an approximately 90% overall survival at 12 months) underwent chemotherapy/radiation or surveillance in lieu of the guideline-preferred therapy.

- Among those registries with at least 50 patients, the relative use of cystectomy varied by approximately two-fold—13% in Iowa to 28% in Seattle—confirming the uncertainty surrounding the effectiveness of treatments for the disease.

- Older patients (especially those 75 and older), those with more comorbidity (Charlson score of 2 or higher) and those residing 50 miles or more away from a cystectomy provider were less likely to undergo radical cystectomy. Among these, the strongest determinant of cystectomy use was advanced age. Patients aged 80 years and older were 90% less likely to undergo cystectomy compared to those aged 66 – 69 years.

- Using both traditional and econometric methods, radical cystectomy conferred a significant survival advantage compared to chemotherapy/radiation and surveillance. Among the marginal population (from the 2-stage residual inclusion estimation from the instrumental variables analysis), the 5-year overall survival was 42%, 21% and 15% for patients managed with cystectomy, chemotherapy/radiation and surveillance, respectively.

In terms of assessing the validity of findings, this is a relatively straightforward study:

- The external validity of the findings is limited by the restriction of SEER-Medicare data to patients 66 years of age and older with muscle invasive bladder cancer. However, because nearly three-quarters of bladder cancer cases occur within the Medicare population, these results are generalizable to the population at highest risk. Furthermore, the age threshold of 66 years is unlikely to be a biologically or clinically meaningful inflection point for the surgical management of patients with bladder cancer and, therefore, systematic differences between Medicare beneficiaries and younger populations are not anticipated.

- As the authors allude to in the manuscript, the use of multilevel models to assess determinants of radical cystectomy provides less biased estimates than traditional multivariable modeling and account for the clustered nature of the data (patients within providers). This is particularly important given that the surgeon explained nearly a third of the variation in the use of
cystectomy (i.e., who you see matters more than the nature of your disease in terms of determining your treatment).

- As with any observational study, results may be (and almost certainly are) confounded by indication—healthier patients by unmeasured variables are more likely to undergo cystectomy. For this reason, traditional methods of assessing treatment effects (in this case, Cox proportional hazards models) are biased. As such, the authors implement an instrumental variables analysis using travel distance as the instrument. This approach represents an econometric of purging models of measured and unmeasured confounding, and has been used to mimic randomization in observational studies. To be valid, the instrumental variable must meet two assumptions: (1) be sufficiently correlated with treatment (i.e., avoid a weak instrument), and (2) affect the outcome (i.e., survival) only through the endogenous regressor (i.e., the instrument is exogenous). While the first assumption is testable and appears to have been met, the second cannot directly be tested empirically. The results from the instrumental variables analysis are generalizable to the marginal population defined as the group of patients whose receipt of treatment depends entirely on the instrument.

The principal limitations of this study are:

- Due to the nature of the data, patient preference for treatment is not captured. This is not the first study to describe the relative underuse of cystectomy for invasive bladder cancer using population-based data. Because of its implications for perioperative morbidity (approximately 50%) and mortality (3-6% in this population), some patients may opt for less effective treatments. Thus, while “21%” seems low, the right rate cannot be accurately assessed.
- Similarly, the nature of the data precludes assessing outcomes beyond mortality. Patient reported quality-of-life is directly impacted by treatment choice and almost certainly varies widely depending on a variety of contextual factors (e.g., surgical expertise, clinical perioperative care, ability to “rescue”, etc.).

These limitations notwithstanding, these data are provocative and suggest ample room for improving the care of muscle invasive bladder cancer patients. In particular, these data can be used to provide patients with some idea of the relative effectiveness of treatment approaches for invasive bladder cancer. In this context, these data can help both patients and providers when contemplating treatment selection. As the physician is the preeminent determinant of the ultimate treatment, patients with invasive cancer should contemplate a second opinion so as to limit the implications of physician practice style.