Introduction

Population-based observational studies allow investigators to evaluate the adoption of new therapies and to determine whether benefits seen in clinical trials are realized in the real world. These studies also provide rich insights into factors that affect access and quality of cancer care in the general population. However, except in a few specific circumstances, real-world data are much less useful in establishing treatment efficacy. Studies that compare outcomes between nonrandomized groups of patients are fundamentally problematic because the patients may also differ with respect to other prognostic factors. This is just as true for modern population-based studies of electronic records as it was for the traditional nonrandomized institution-based reviews of paper charts that were popular 50 years ago before the widespread adoption of hierarchies of evidence raised awareness of the relative weakness of observational study designs. Population-based studies may have greater external validity than institution-based studies, but there is no reason to believe they have any greater internal validity; both are therefore classified as level-3 evidence in Sackett’s original hierarchy.

In evaluating claims of treatment effectiveness from observational data, we believe that clinicians should emulate the approach used by epidemiologists for evaluating claims that a given exposure causes a particular disease. Bradford Hill’s widely used framework for establishing causation is helpful in this context. Hill’s framework provides a set of criteria for causation that can be applied to the totality of available evidence, including strength of the association and its consistency across studies, presence of a biologic gradient, specificity (the exposure is associated with the expected outcome and not with other outcomes), and plausibility of the effect and its coherence with related scientific knowledge.

We have been following with some alarm the recent series of articles that explore the role of aggressive local therapy (LT) in patients with metastatic urologic cancer. Although it is standard practice to consider cytoreductive nephrectomy in metastatic renal cell carcinoma, this is based on level-1 evidence from two randomized controlled trials. Investigators in breast and colorectal cancer were equally enthusiastic about this phenomenon, but recent studies have not shown benefit. In this commentary, we propose that recent findings in bladder cancer are subject to such fundamental methodologic shortcomings that the purported findings of benefit to aggressive therapy are likely false and potentially dangerous to patients.

Key points

- Population-based observational studies allow investigators to evaluate the adoption of new therapies and to determine whether benefits seen in clinical trials are realized in the real world.
- Population-based studies may have greater external validity than institution-based studies, but there is no reason to believe they have any greater internal validity; both are therefore classified as level-3 evidence in Sackett’s original hierarchy.
- We have been following with some alarm the recent series of articles that explore the role of aggressive local therapy (LT) in patients with metastatic urologic cancer.
- Although it is standard practice to consider cytoreductive nephrectomy in metastatic renal cell carcinoma, this is based on level-1 evidence from two randomized controlled trials.
Metastatic UCB was defined as the presence of extrapelvic positive lymph nodes or bone and/or visceral involvement. High-intensity LT was defined as radical cystectomy (RC) or high-dose radiotherapy (RT).

The overall study population (N = 3,753) and the 297 patients with high-intensity LT represent 0.6% and 0.05% of all patients with UCB, respectively.

The authors report that high-intensity LT was associated with improved survival (medial survival, 15 vs. 10 months; P < .001), and this difference persisted on adjusted analysis using propensity scores (hazard ratio, 0.56; 95% CI, 0.48 to 0.65).

The authors concluded that there was an overall survival benefit for individuals with metastatic UCB treated with high-intensity versus conservative LT.

We believe it is more likely that many patients underwent high-intensity LT because it was initially assumed that they did not have metastatic disease.

For example, a patient who undergoes RC for apparently localized bladder cancer but is subsequently found to have low-volume bone metastases when staging is completed before adjuvant chemotherapy.

Seisen et al14 have recently used the National Cancer Database to evaluate the relationship between high-intensity LT and survival in patients with metastatic urothelial carcinoma of the bladder (UCB). Metastatic UCB was defined as the presence of extrapelvic positive lymph nodes or bone and/or visceral involvement. High-intensity LT was defined as radical cystectomy (RC) or high-dose radiotherapy (RT). From a population of 603,298 patients diagnosed with UCB, they identified 3,753 patients with metastatic disease treated with systemic chemotherapy from 1998 to 2012; 297 (8%) had high-intensity LT (248 had RC and 49 had RT). The overall study population (N = 3,753) and the 297 patients with high-intensity LT represent 0.6% and 0.05% of all patients with UCB, respectively. The authors report that high-intensity LT was associated with improved survival (medial survival, 15 vs. 10 months; P < .001), and this difference persisted on adjusted analysis using propensity scores (hazard ratio, 0.56; 95% CI, 0.48 to 0.65).

The authors concluded that there was an overall survival benefit for individuals with metastatic UCB treated with high-intensity versus conservative LT.

The first thing to consider is whether patients who received high-intensity LT are comparable to patients who did not receive high-intensity LT. It may be useful to consider why any patients actually received high-intensity LT before there was any suggestion in the literature that this might prolong survival. In that era, the only conceivable reason for deliberately treating a patient who had metastatic UCB with RC or RT would be to alleviate refractory pelvic symptoms. In general, this would be performed only if the patient had good performance status, minimal burden of disease, and a reasonable life expectancy.

In our experience, those circumstances are rare and unlikely to account for the almost 1 in 10 patients with metastatic UCB treated with chemotherapy who received high-intensity LT in this cohort. We believe it is more likely that many patients underwent high-intensity LT because it was initially assumed that they did not have metastatic disease. Consider, for example, a patient with normal preoperative imaging who is unexpectedly found to have extrapelvic lymph node metastases at the time of cystectomy. Consider a patient who undergoes RC for apparently localized bladder cancer but is subsequently found to have low-volume bone metastases when staging is completed before adjuvant chemotherapy. Both of these patients would be staged as having metastatic disease in the National Cancer Database. Collaborative staging methodology, which was used during the study period,15 factors into the stage assignment any additional information about the extent of the disease obtained at surgery and any evidence found on imaging within 4 months after surgery. Whether the patients in the high-intensity LT group were initially treated with palliative or curative intent, it is probable that they all had good performance status, relatively low or equivocal burden of metastatic disease, and better life expectancy than patients who received conservative LT. It would therefore have been astonishing if the high-intensity LT group had not experienced much better overall survival than the conservative LT group. Finally, it is possible that some patients in the LT cohort were subject to misclassification bias. This would be the case for a patient who presents with UCB and small lung lesions that are presumed metastases. After initiating chemotherapy, if repeat imaging shows an excellent response in the pelvis and no change in the lung nodules, these might now be considered benign. This misclassified patient would subsequently have high-intensity LT and be destined to have better outcomes than patients without high-intensity LT.

There is another important methodologic reason why the two treatment groups would not be expected to experience similar survival. Patients could not be included in the high-intensity LT group unless they survived long enough to undergo RC or
RT, patients in this group had a zero risk of dying during the period between diagnosis and treatment. Conversely, patients who did not undergo RC or RT were automatically included in the conservative LT group and were exposed to the risk of death from date of diagnosis. Thus, the high-intensity LT group would be expected to live longer from date of diagnosis than the conservative LT group, even if LT had no effect on survival. This well-known problem of immortal time bias is particularly relevant in metastatic UCB in which one would expect a not-insignificant number of deaths shortly after diagnosis. Those patients who died soon after diagnosis would never be eligible for high-intensity LT; they would all be included in the conservative LT group and thereby worsen its prognosis. The effect of immortal time bias may be mitigated by doing a landmark analysis in which the survival analysis is restricted to patients who have already survived for a prespecified time since diagnosis; this does not seem to have been performed by Seisen et al. A secondary analysis by Seisen et al reports that patients with consolidative high-intensity LT (ie, after chemotherapy) have improved survival compared with those with cytoreductive high-intensity LT (ie, before chemotherapy). This in itself strongly suggests the presence of immortal time bias because the consolidative patients have to survive even longer from diagnosis to surgery than the cytoreductive patients.

Recognizing then that the two treatment groups in the Seisen et al study almost certainly differed considerably with respect to important prognostic factors, the next question is whether the investigators adequately controlled for these differences in their analysis. The authors performed a propensity score analysis to adjust for differences in the distribution of prognostic factors between the high-intensity LT and conservative LT groups. This enabled them to control for covariates, including age, comorbidity, and T and N stage. However, they were unable to control for performance status or the presence of visceral metastases, which are far more important prognostic factors in this context. Given that these two factors would have been expected to play an important role in the decision of whether to provide high-intensity LT, it would be naïve to believe that controlling for other, less important factors would be sufficient to overcome the problem of treatment selection bias in that study.

Recognizing that propensity score methodology will balance only observed covariates, Seisen et al performed a sensitivity analysis without assumptions as proposed by Ding and VanderWeele. Seisen et al use the results of that analysis to justify their study results. However, in the context of survival analysis, the approach used by Ding and VanderWeele is applicable only for rare event scenarios. Because death is not a rare event in metastatic bladder cancer, we do not think this statistical approach should have been used to justify the study’s primary analysis.

In summary, there are strong a priori reasons to suspect large inherent differences in prognosis between the two treatment groups compared in the study by Seisen et al, and the investigators were unable to control for the most important prognostic factors. We therefore believe that the difference in survival between the high-intensity LT and conservative LT groups in their study is almost certainly the result of residual confounding. Although Seisen et al acknowledge this, they still conclude that they have found evidence of a real treatment effect, based on what other reviewers refer to as fancy statistics. However, we have evidence to believe that the proposed sensitivity analysis is unable to protect their findings from fundamental methodologic flaws.

In addition to these methodologic concerns, the potential association described by Seisen et al does not satisfy several of Hill’s criteria for causation, including consistency (this is the first such report in bladder cancer), plausibility (there is little underlying correlative evidence to substantiate that removal of the primary tumor would lead to a large improvement in survival analysis is restricted to patients who have already survived for a prespecified time since diagnosis.

• A secondary analysis by Seisen et al reports that patients with consolidative high-intensity LT (ie, after chemotherapy) have improved survival compared with those with cytoreductive high-intensity LT (ie, before chemotherapy).

• Recognizing then that the two treatment groups in the Seisen et al study almost certainly differed considerably with respect to important prognostic factors, the next question is whether the investigators adequately controlled for these differences in their analysis.

• The authors performed a propensity score analysis to adjust for differences in the distribution of prognostic factors between the high-intensity LT and conservative LT groups.

• Recognizing that propensity score methodology will balance only observed covariates, Seisen et al performed a sensitivity analysis without assumptions as proposed by Ding and VanderWeele. Seisen et al use the results of that analysis to justify their study results.
Key points

- We disagree with the stated recommendation by Seisen et al that this question now be tested in a randomized controlled trial.

- It is critical that study methodology and interpretation of results be held to the same scientific standards as a clinical trial.

- Although the focus of this commentary has been metastatic bladder cancer, there is currently even greater interest in the role of aggressive LT for metastatic prostate cancer.

The way forward

Although observational research using real-world data can offer important insights into care and outcome of patients in the general population, it is critical that study methodology and interpretation of results be held to the same scientific standards as a clinical trial. Real-world data can provide insight into measures of comparative effectiveness; however, these analyses are complex and fraught with methodologic shortcomings. Although the focus of this commentary has been metastatic bladder cancer, there is currently even greater interest in the role of aggressive LT for metastatic prostate cancer. A detailed review is beyond the scope of this commentary, but we are concerned that many of the existing studies in prostate cancer are vulnerable to the same methodologic shortcomings as the report by Seisen et al. Although Seisen et al appropriately caution that their results should not change practice, we are concerned that the clinical community may read these complex analyses in major journals and move directly to clinical application. It behooves all of us to be cautious in interpreting the results of studies such as these. The risks of getting this wrong are substantial; the oncology community will be doing a tremendous disservice to our patients if we generate data that are used to justify major surgical procedures with significant morbidity and no real benefit in patients with terminal cancer.

References