The Comparison of Intravesical Therapy and Surgery as Treatment Options (CISTO) for Bladder Cancer Study

A pragmatic observational study of 572 patients with non-muscle invasive bladder cancer undergoing treatment with either radical cystectomy or bladder sparing therapy

ClinicalTrials.gov ID
NCT03933826

Statistical Analysis Plan

Version 1.0, March 14, 2024

TABLE OF CONTENTS

1. General Design Considerations

1.1 Aims and Hypotheses

Aim 1: To compare patient-reported and patient-centered clinical outcomes between patients undergoing radical cystectomy and those receiving bladder-sparing therapies (BST) for NMIBC that have failed first-line BCG.

Hypothesis 1.1: Patients undergoing radical cystectomy will have worse generic health-related QOL within 12 months of bladder removal surgery compared with BST patients.

Hypothesis 1.2: 12-month disease-free survival and metastasis-free survival will be better among radical cystectomy patients than among BST patients.

Hypothesis 1.3: Patient-reported and clinical outcome differences will vary within important subgroups including women, non-white patients, elderly patients, patients with multiple comorbid health conditions, patients with poor urinary function scores at the time of BCG failure, patients without caregivers, patients with atypical cancer histologies, and patients receiving investigational agents.

Aim 2: To characterize the heterogeneity of treatments received and corresponding patient and caregiver preferences for NMIBC that have failed first-line BCG.

Hypothesis 2.1: Patients who choose radical cystectomy for NMIBC with BCG failure will do so more often because they feel this offers their best chance for survival. Patients who choose BST will do so more often because they are concerned about the complications and QOL detriments associated with radical cystectomy.

Hypothesis 2.2: Patient and caregiver preferences will correlate strongly with observed treatments received and health-related QOL outcomes. We anticipate that QOL will vary by concordance of treatments received with treatment preferences and health state utilities.

1.2 Design

The CISTO Study is a pragmatic, prospective observational cohort study of patients with recurrent high-grade NMIBC who have selected management with BST or radical cystectomy. This multicenter study has an allocation ratio of 2:1 across treatment arms (BST to radical cystectomy, respectively) by site. A prospective observational cohort study design was chosen for the CISTO Study in response to critical input from the BCAN Patient Survey Network. Among 291 respondents with NMIBC, only 11% reported being willing to consent to randomization for a study of BST versus radical cystectomy. Therefore, a prospective observational cohort study was selected as the highest quality study design for addressing the research questions.

1.3 Treatment Group Selection

Patients are enrolled in either of two arms, based on the patient's individual treatment decision: participants undergoing radical cystectomy (any surgical approach including open or robotic surgery) and those receiving BST (additional BST, including BCG, or intravenous immunotherapy) for recurrent high-grade NMIBC.

1.4 Enrollment Timing

A participant's study clock begins on the date of completion of the baseline survey entry date.

1.5 Analyses

Primary analyses will be based on an Intention to Treat (ITT) approach. Patient data will be analyzed according to the treatment arm decided upon at enrollment regardless of whether that treatment was initiated. Due to the 2:1 allocation of treatment, statisticians and analysts are not blinded to the study treatment assignment when conducting statistical analyses. Patients that withdraw consent after enrollment and before the beginning treatment will be excluded from the ITT analyses. Patients will also be excluded from the study cohort if found to meet exclusion criteria upon review of EMR data entered into the study data management system.

1.6 Analysis Timing

To facilitate and expedite publication of the primary results within the allotted window of funding for CISTO, an analytic data set will be frozen on March 14, 2024. All baseline information will be included in the analytic data set, as well as all longitudinal follow-up data for patients were enrolled in CISTO on February 14, 2023 or earlier. Statistical analysis, tables, and figures will be generated by the data coordinating center and manuscript development will begin in conjunction with the CISTO Writing Committee who are not involved with patient care or outcome collection. Once data have been collected and cleaned for the final participant (estimated January 2025), a final analytic data set will be generated to refresh tables, figures, and associated details with reporting.

2. Primary Outcome

The *primary outcome* is patient-reported health-related quality of life as measured with the EORTC QLQ-C30,[1] which has been previously used in bladder cancer populations. The *primary endpoint* is the physical functioning domain score of the EORTC QLQ-C30 at 12-months.

3. Secondary Outcomes

For patients, secondary outcomes through 12 months include self-reported urinary, bowel, and sexual function and bothersomeness as measured with the Bladder Cancer Index,[2] PROMIS depression and anxiety, EuroQoL 5D, NMIBC treatment preferences, decisional regret, financial distress measured with the Comprehensive Score for Financial Toxicity (COST),[3] healthcare utilization measured as 12-month hospital and urology clinic days, return to work/normal activities, 12-month cancer-specific survival, metastasis-free survival, bladder cancer-specific survival, and progression to muscle-invasive bladder cancer. Among patients with available data, we will assess health-related quality of life, health literacy, urinary function, sexual function, financial distress, healthcare utilization, disease-free and metastasis-free survival, bladder cancer-specific survival, and progression at up to 48 months, and up to 6 additional years for extended follow-up.

4. Data Analyses

We will use descriptive statistics to characterize the treatments received, PROs, and clinical outcomes. Prior to statistical analysis, we will first compare continuous demographic and clinical characteristics by treatment group with Wilcoxon rank-sum tests to protect against violations of normality assumptions. P-values from Exact Conditional Tests, such as Fisher's exact test and its multi-degree of freedom extensions, will be used to compare categorical data.

To address treatment selection bias and potential confounding by indication in the comparison of BST vs. radical cystectomy, we will utilize targeted maximum likelihood estimation (TMLE) as the primary analytic approach for causal effect estimation using observational data.[4-6] Unlike approaches that focus on creating

carefully matched treatment and control groups, TMLE allows for the inclusion of all patient participants and their reported outcome measures.[4-6] TMLE first uses covariate-adjusted regression models to generate an initial estimate of the treatment effect of radical cystectomy through the creation of *potential outcomes*: two predicted outcomes for each individual patient participant, under the hypothetical assumption that they had been treated with either radical cystectomy or BST. Next, as with a propensity score approach, we will estimate the probability of treatment with radical cystectomy using SuperLearner. The SuperLearner procedure selects the best weighted combination of prediction models from among candidate learners such as logistic regression, stepwise logistic regression, generalized additive models, Bayesian GLM, LASSO, random forest, gradient descent boosting, support vector machine, and the sample mean. Finally, we will use the probabilities obtained in the second step to update the initial estimate of each patient participant's pair of potential outcomes and the TMLE estimate is interpreted as the difference in outcomes if all patients had been treated with BST versus having been treated with radical cystectomy.[5]

4.1 Primary Outcome

Patient data will be analyzed according to an intention-to-treat (ITT) framework for the primary analysis, where the treatment arm is decided upon at enrollment regardless of whether that treatment was initiated. Patients that withdraw consent after enrollment and before the beginning treatment will be excluded from the ITT. Longitudinal trajectories of the effect of treatment on patient QOL over time will be modeled and characterized using standard linear mixed effects models for the primary analytic approach.

4.1.1 Subgroups

Because each patient has a pair of potential outcomes under each treatment, we will include covariates in the TMLE and longitudinal data models to provide treatment effect estimates for important subgroups of interest stratified by age (75 or older vs. under 75), gender (males vs. females), patients who have a caregiver (yes vs. no), and cancer severity (carcinoma in situ vs not).

4.2 Secondary Outcomes

Other patient reported outcomes will be evaluated using a similar analytic framework to the primary outcome. To test the hypothesis that cancer-specific survival will be better in patients undergoing radical cystectomy, we will use a superiority testing framework. For assessment of time-to-event outcomes we will use inverse probability weighted Cox proportional hazards models to estimate relative risks. Survival outcomes will be censored at the time of the last study contact or clinic visit for patients who are lost to follow-up and have no recent EHR-verified clinic visits. Since time-to-event outcomes (cancer-specific survival and metastasis-free survival) will be highly correlated with one another, two-sided 95% confidence intervals will be used for inference without adjustment for multiple survival endpoints.

4.3 As-Treated Analyses

Patients who choose a treatment arm upon enrollment may ultimately be treated as intended or may have first received treatment of the other arm, or treatment may be delayed or postponed. For as-treated analyses, we will classify participants' treatment arm according to the first treatment observed within six months of enrollment (BST, radical cystectomy, or none observed within six months). Patients with no bladder cancer treatment observed within six months will be excluded from as-treated analyses directly comparing BST to radical cystectomy. Outcomes and characteristics of patients with no treatment observed will be presented and informally compared to the as-treated groups. The statistical

approach with the as-treated groups will otherwise mimic ITT analyses. Results from the ITT analysis will be considered primary, with as-treated analyses considered secondary.

4.4 Sensitivity Analyses

As a complementary sensitivity analysis to the primary analytic approach of TMLE, we will use alternative approaches such as propensity score matching and G-computation methods to deal with potential treatment selection bias. Any potential differences in results would be investigated and substantive departures from the TMLE results will be reported. We will also use generalized estimating equations with robust standard errors as a secondary approach towards analyzing longitudinal patient reported outcomes. For patients that die, in sensitivity analyses we will impute the floor of each QOL instrument for all subsequent time points to provide death-adjusted treatment effect estimates.

Despite our intent to be comprehensive in identifying important variables that we can measure in comparing BST and radical cystectomy, it remains possible that unknown unmeasured confounders may remain. For example, beyond collecting tumor stage classification, histology, multifocality, and tumor size, there may be residual cancer severity characteristics that are difficult to extract from health record data. For binary or time-to-event endpoints, we will calculate E-values to evaluate the minimum strength of association an unmeasured confounder would need to have to fully explain away any observed treatment effect.[7]

5. Compliance, Retention, and Missing Data

5.1 Adherence and Retention

Participant follow-up assessments are completed in-person, online, by phone, or by mail. Study staff from the University of Washington Clinical Coordinating Center will contact the participant to complete the follow-up assessments. Outreach will include a combination of phone, mail, and email as determined by the contact information provided by the participant. If the contact protocol is exhausted and the assessment has not been completed, a research coordinator at the participant's enrollment site will be notified. Research personnel will review the EMR and provide any site-specific or EMR-specific updates that may be helpful, for example, updating a telephone number. Additional contact attempts (calls, emails, texts, etc.) may be attempted from the local site coordinators.

Participants may experience research assessment burnout due to the frequency and number of questions asked of them. To optimize complete data collection, research personnel may instead complete a subset of survey responses (minimal research assessment). If a participant requests to complete a minimal research assessment, research personnel will prioritize asking research participants the primary and secondary outcome measures.

5.2 Missing Data and Dropouts

We will strive to sustain excellent participant involvement throughout the study, and we have achieved 90%+ follow-up rates in numerous prior studies. The Data Coordinating Center at the University of Washington will generate automated nightly reports, available to staff at the study sites, identifying these fields with a request to discuss and prevent further missingness. Important data elements will be prospectively monitored to examine patterns of missingness.

5.3 Handling

In final manuscripts and analyses, the number of non-responders will be enumerated by study arm according to CONSORT guidelines. We will conduct a missing data analysis to describe and characterize enrolled participants who do not provide further response due to *attrition* or *dropout*. As a part of treatment effect estimation, the TMLE algorithm provides an integrated approach to incorporating uncertainty that arises due to missing longitudinal data. The mean outcome conditional on observing the outcome may be a biased estimate when missingness is informative. TMLE can reduce this bias when missingness is a function of measured baseline demographic and clinical characteristics. Missing outcome measurements will be accounted for by nonparametrically estimating the missing data mechanism to produce a matrix of missing data probabilities conditioned on baseline demographic and clinical characteristics. The missing data conditional probabilities will then be incorporated into the TMLE estimation procedure during the third step of updating potential outcomes.

5.4 Sensitivity Analysis

When missing data are suspected to be *missing not at random*, any standard analysis of the available cases is likely biased. The bias cannot be corrected since missingness depends on unobserved data and therefore cannot be empirically modelled reliably. We will therefore assess the sensitivity of inferences made from missing data by imputing missing data under both pessimistic and optimistic scenarios and repeating the TMLE algorithm to provide bounds on the statistical uncertainty. The characteristics of non-responders will be summarized, and we will present the heterogeneity of the treatment effect due to missing data.

6. Sample Size and Accrual

With recruitment of 572 participants, this study has >0.80 statistical power to detect small differences in QOL between treatment approaches (Cohen's d = 0.24, or 5.5 points on the physical function scale of EORTC-QLQ-C30).[8] Power analyses conservatively allowed for 10% missing data and assumed a correlation between repeated QOL measurements of 0.3. Assuming a similar balance of treatments within subgroups, this study also has >0.80 power to detect moderate but clinically important treatment effects (Cohen's d = 0.43, or 9.9 points on the physical function scale of EORTC-QLQ-C30) within subgroups as small as 30% of the study cohort. Anticipated subgroups of this size include patients aged 75 years or older, women, and patients with multiple comorbid health conditions. For the secondary outcome of bladder cancer-specific mortality, data from a review combining multiple trials demonstrated an anticipated 1-year bladder cancer-specific mortality rate of 4.8% when treated with medical management.[9] Assuming a 4.8% mortality rate in the medical management arm, this study has power to detect differences in 1-year bladder cancer-specific mortality rates of $\leq$1% in the radical cystectomy arm (power > 80%).

7. Data Monitoring

We will monitor the accuracy of data entry by the sites both internally and externally. We will review study data on arrival for completeness. We will then subject each submitted data set to a set of preliminary checks to search for values that are out-of-range or otherwise inappropriate. The DCC will set up in-line data quality checks within the data entry system, which will check for missing information, inconsistencies, and logic errors. The CCC and DCC will monitor this information on a regular basis. In addition, the DCC will perform quality assurance checks by running regular data quality reports. These reports include missing values for required fields, incorrect data type, range checks, outliers, hidden fields that contain values, and multiple-choice fields with invalid values. Values that need to be corrected will be brought to the attention of the research staff at

that site. The DCC and CCC will investigate any discrepancies or unexpected values that were discovered through the regular report or through visual inspection.

8. Data and Safety Monitoring Plan (DSMP)

The CISTO Study adheres to a Data Safety Monitoring Plan. Given the minimal risk nature of this observational cohort study, monitoring is conducted by the study's Executive Committee. On a monthly basis, the Executive Committee reviews any data and safety events and procedures and determine recommendations for these events and procedures as appropriate, including identifying, reviewing, and reporting adverse events (AEs) and serious adverse events (SAEs) and unanticipated problems to the applicable Institutional Review Boards (IRBs) or other monitoring bodies. SAEs are defined as 1) death during the study period; 2) life-threatening event related to the treatment or significant disability/incapacity related to the treatment; 3) inpatient hospitalization (other than for cystectomy); and 4) prolonged hospitalization following cystectomy (14 days or more). The number of adverse events and related unexpected serious adverse events (SAEs) will be summarized by arm, by recruitment site, by grade, and by the National Cancer Institute's Common Terminology Criteria for Adverse Events (CTCAE) system organ class. In addition, for each toxicity, the proportion of affected participants overall and by arm will be summarized by the maximum CTCAE grade experienced. The interim review of SAEs will not lead to recommendations towards study recruitment or discontinuation.

9. References

1. Gilbert SM, Wood DP, Dunn RL, et al. Measuring health-related quality of life outcomes in bladder cancer patients using the Bladder Cancer Index (BCI). *Cancer*. 2007;109(9):1756-1762. doi:10.1002/cncr.22556
2. Rosenblum M, van der Laan MJ. Targeted maximum likelihood estimation of the parameter of a marginal structural model. Int J Biostat. 2010 Apr 15;6(2):Article 19.
3. Souza JA, Yap BJ, Wroblewski K, et al. Measuring financial toxicity as a clinically relevant patient-reported outcome: The validation of the COmprehensive Score for financial Toxicity (COST). *Cancer*. 2017;123(3):476-484.
4. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. Am J Epidemiol. 2017;185(1):65–73.
5. Gruber S, van der Laan MJ. A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. Int J Biostat. 2010 Aug;6(1):Article 26.
6. Pirracchio R, Petersen ML, van der Laan M. Improving propensity score estimators' robustness to model misspecification using super learner. Am J Epidemiol. 2015 Jan 15;181(2):108–19.
7. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017 Aug 15;167(4):268–274. PMID: 28693043.
8. Sawilowsky SS. New Effect Size Rules of Thumb. J Mod Appl Stat Methods. 2009 Nov 1;8(2):597–9.
9. Cambier S, Sylvester RJ, Collette L, Gontero P, Brausi MA, van Andel G, et al. EORTC Nomograms and Risk Groups for Predicting Recurrence, Progression, and Disease-specific and Overall Survival in Non–Muscle-invasive Stage Ta–T1 Urothelial Bladder Cancer Patients Treated with 1–3 Years of Maintenance Bacillus Calmette-Guérin. Eur Urol. 2016 Jan 1;69(1):60–9.