Janssen Research & Development

Statistical Analysis Plan

Phase 2, Open-Label, Multi-Center, Randomized Study of TAR-200 in Combination with Cetrelimab and Cetrelimab Alone in Participants with Muscle-Invasive Urothelial Carcinoma of the Bladder who are Scheduled for Radical Cystectomy and are Ineligible for or Refusing Platinum-Based Neoadjuvant Chemotherapy

SunRISe-4

Protocol 17000139BLC2002; Phase 2
Version: Amendment 2

JNJ-17000139 (TAR-200)
JNJ-63723283 (Cetrelimab)

Status:	Approved
Date:	09 June 2025
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-RIM-887278, V4.0

GCP Compliance: This study will be conducted in compliance with Good Clinical Practice, and applicable regulatory requirements.

TABLE OF CONTENTS

VERSION HISTORY

Table 1: SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1.0	11 Nov 2022	Not Applicable	Initial release
2.0	18 Apr 2025	1. Clarified and added new analysis sets. 2. Updated TSP definition 3. Added additional sensitivity analyses. 4. Provided clarification throughout the document	To incorporate changes from protocol amendment and to provide more clarification of texts.
3.0	06 May 2025	Corrected format errors	To correct formatting error
4.0	09 June 2025	Added additional sensitivity analysis	To address FDA comments

1. INTRODUCTION

This document presents the Statistical Analysis Plan (SAP) for protocol 17000139BLC2002. This SAP contains definitions of analysis sets, derived variables, and statistical methods for the planned analyses for this study.

This SAP follows guidelines provided in the International Conference on Harmonization (ICH) Topic E9 Statistical Principles for Clinical Trials. In the event of future amendments to the protocol, this SAP may be modified as necessary to account for changes relevant to the statistical analysis.

1.1. Objectives and endpoints

Primary Objective

To determine the anti-tumor effects, as assessed by the pathological complete response (pCR) at Radical Cystectomy, of TAR-200 + IV cetrelimab (Cohort 1) and IV cetrelimab alone (Cohort 2) in participants with Muscle-Invasive Urothelial Carcinoma of the Bladder who are scheduled for Radical Cystectomy and are ineligible for or refusing Platinum-Based Neoadjuvant Chemotherapy.

Secondary Objectives

The secondary objectives are:

- To evaluate the safety and tolerability of up to 4 dosing cycles of TAR-200 + IV cetrelimab (Cohort 1) and IV cetrelimab alone (Cohort 2) prior to radical cystectomy (RC).

- To determine the recurrence-free survival (RFS) in participants receiving TAR-200 + IV cetrelimab (Cohort 1) and IV cetrelimab alone (Cohort 2).

Exploratory Objectives

The exploratory objectives are:

- To assess the participant's cancer-related QoL using the FACT-Bl (Functional Assessment of Cancer Therapy – Bladder).
- To evaluate changes in gene or protein expression in the tumor, blood, and urine with the correlation of other endpoints.
- To determine whether the baseline immune status, molecular subtype or mutational status influences treatment response.
- To determine the overall survival (OS) in participants receiving TAR-200 + IV cetrelimab (Cohort 1) and IV cetrelimab alone (Cohort 2).
- To determine the pathologic overall response (pOR) rate at RC in TAR-200 + IV cetrelimab (Cohort 1) and IV cetrelimab alone (Cohort 2).
- To determine time to symptomatic progression (TSP) in participants receiving TAR-200 + IV cetrelimab (Cohort 1) and IV cetrelimab alone (Cohort 2).

- To explore the relationships between PK, pharmacodynamic, adverse event profile, and antitumor activity.
- To evaluate the PK of gemcitabine (TAR-200) and major metabolite dFdU in urine and plasma following administration of TAR-200 in combination with IV cetrelimab (Cohort 1).
- To evaluate the PK and immunogenicity of cetrelimab in serum following administration of TAR-200 in combination with IV cetrelimab (Cohort 1) and IV cetrelimab alone (Cohort 2).

1.2. Study Design

This is a randomized, open-label, multicenter, Phase 2 clinical study of intravesical gemcitabine delivered via the TAR-200 in combination with IV cetrelimab and neoadjuvant IV cetrelimab alone in participants with muscle-invasive urothelial carcinoma of the bladder (clinically referred to as MIBC) who are scheduled for RC and are ineligible for or refusing platinum-based neoadjuvant chemotherapy. The initial, diagnostic TURBT confirming a pathologic diagnosis of cT2-T4a MIBC must have been completed within 120 days of randomization (90 days in France). All participants who have confirmed MIBC at screening and have met all eligibility criteria will be randomly assigned (5:3) to receive intravesical TAR-200 in combination with IV cetrelimab (Cohort 1) or IV cetrelimab alone (Cohort 2) neoadjuvant to scheduled RC and urinary diversion. Two stratification factors for analysis will be the completeness of TURBT (visibly complete vs incomplete and ⩽3 cm) and tumor stage (cT2 vs. cT3-4a) at initial diagnosis. Participants must have tumor volume ⩽3 cm prior to randomization. Participants in screening must undergo a cardiovascular risk assessment based on available guidelines for cardiac and surgical risk assessment and must not be enrolled if not considered eligible for RC.

A target of 160 participants will be enrolled and randomly assigned in a 5:3 ratio in this study with 100 participants planned to be randomized into Cohort 1 and 60 participants into Cohort 2.

For Cohort 1, at the initial treatment visit, TAR-200 will be placed intravesically via the Urinary Placement Catheter (UPC) and cetrelimab will be dosed intravenously (360 mg CCI [redacted] IV infusion). The initial TAR-200 will be removed via flexible or rigid cystoscopy at Week 3 (±3 days), and then the second TAR-200 will be placed via a UPC. This removal/replacement procedure will be repeated for a third and fourth dosing cycle at Week 6 (±3 days) and Week 9 (±3 days), respectively. The fourth TAR-200 will be removed at Week 12 (±3 days). A cystoscopy is not required at these visits if TAR-200 removal is not needed.

For Cohort 1 and Cohort 2, IV cetrelimab will be dosed approximately every 21 days for 4 consecutive cycles, with cycles starting at Week 0, Day 1. Subsequent doses will be administered at Weeks 3 (±3 days), 6 (±3 days), and 9 (±3 days). IV cetrelimab is not required to be administered on the same day as the placement of TAR-200, but treatments should occur within 24 hours and no more than 72 hours of each other. On days in which both TAR-200 and IV cetrelimab are administered on the same day, TAR-200 placement should be separated by a minimum of 45 minutes from the start of IV cetrelimab IV infusion to allow for discrete evaluation of potential adverse events of each treatment.

For both cohorts, the RC should be performed at the designated RC visit and within 3 weeks of the Week 12 visit. During Week 6, participants in both cohorts will undergo an imaging (CT/MRI) assessment. If metastatic progression is noted on the local report and confirmed by central radiology review, the participant will be discontinued from the efficacy (imaging) follow-up and continued to be followed for survival status, subsequent anti-cancer therapies, and other malignancies. If local progression is suspected or clinical symptoms warrant, the participant will proceed with immediate RC and urinary diversion, at the local Investigator's discretion. If the RC is delayed, from the Week 12 visit, the date of the RC surgery will become the protocol-specified 'RC visit' (per Schedule of Activities, Section 1.3 in the protocol).

Following the RC and urinary diversion, all participants will have a safety follow-up visit 4 weeks following the RC surgery. Additionally, participants will have follow-up study visits at Weeks 8, 12, 24, 36, 48, 60, 72, 84, 96, and 108 (End of Study) post RC. The post RC follow-up visits (e.g., Weeks 4, 8, 12, 24, etc.) are derived from the date of the RC surgery. No restrictions are placed on adjuvant therapy after RC.

Site visits may be replaced with telephone visits during a national disaster, with site visits resuming as soon as possible thereafter. Please refer to Section 10.11 in the protocol for additional details.

An IDMC will be commissioned for this study. Refer to Committees Structure in Section 10.3, Appendix 3, in the protocol for details.

A diagram of the study design is provided.

Figure 1: Schematic Overview of the Study

2. STATISTICAL HYPOTHESES

The statistical hypothesis is that TAR-200 + IV cetrelimab (Cohort 1) will lead to a pCR rate CCI or greater (against a hurdle rate of CCI), and the IV cetrelimab alone (Cohort 2) will lead to a pCR rate CCI or greater (against a hurdle rate of CCI). There is no formal statistical hypothesis testing for the comparison between the 2 cohorts.

3. SAMPLE SIZE DETERMINATION

This study will randomize approximately 160 participants in a 5:3 randomization ratio to receive TAR-200 in combination with IV cetrelimab (n=100) or IV cetrelimab alone (n=60).

This sample size determination utilizes a Bayesian approach with Beta(0.5, 0.5) as the prior distribution and Beta(0.5+m, 0.5+n-m) as the posterior distribution for the pCR rate, where m is the observed number of participants with pCR and n is the total number of participants. Assuming a pCR rate of CCI for TAR-200 in combination with IV cetrelimab (Cohort 1), 100 participants will provide over 90% probability to have the lower limit of the 80% credible interval exceeding CCI pCR rate (or about 80% probability of exceeding CCI pCR rate). Assuming a pCR rate of CCI for the IV cetrelimab alone (Cohort 2), 60 participants will provide about 65% probability to have the lower limit of the 80% credible interval exceeding CCI pCR rate.

4. POPULATIONS FOR ANALYSIS SETS

For purposes of analysis, participants will be classified into the following analysis sets. For each analysis set described below participants who received an incorrect study intervention will be analyzed under the planned study intervention.

Analysis Sets	Description
Screened	All participants who signed Informed Consent Form (ICF)
Randomized	All participants who were randomized in the study
Full	All participants who receive at least 1 dose of any study treatment
Safety	All participants who receive at least 1 dose of any study treatment
Efficacy Evaluable	All participants who have adequate RC results
Immunogenicity Analysis	All participants who receive at least 1 dose of IV cetrelimab and have at least 1 posttreatment immunogenicity sample
PK analysis	All participants who received at least 1 dose of TAR-200 or IV cetrelimab and had at least 1 evaluable PK sample (plasma/serum sample or urine PK sample) obtained posttreatment

5. STATISTICAL ANALYSES

5.1. General Considerations

This study is being performed to assess the efficacy, safety, and tolerability of CCI [redacted] of TAR-200 in combination with IV cetrelimab and IV cetrelimab alone neoadjuvant to radical cystectomy. A side-by-side descriptive summary of efficacy will be provided to illustrate the contribution of TAR-200 to the efficacy of the combination therapy.

5.1.1. Study Phases

Screening Phase

The screening phase is only defined for screened participants who were not screen failures and were assigned to an intervention cohort. The reference start date of the screening phase is the day

when participants signed the ICF, and the reference end date is the day before the reference start date of the treatment phase.

Treatment Phase

The treatment phase is only defined for participants who received at least 1 dose of the study intervention. The treatment phase is defined to be between the date of the first dose of study intervention and the date of the RC visit. If the date of the end of treatment or RC visit is not available, the date of the last dose of study medication + 21 days will be used. The assessments performed during the RC visit will be included in this phase.

Follow-up Phase

The follow-up phase is only defined for participants who entered the follow-up phase. The reference start date of the follow-up phase is the day after the reference end date of the treatment phase and the reference end date of the follow-up phase is the date of death, withdrawal of consent, or the end of the study, whichever occurs first.

5.2. Participant Dispositions

The number of participants in the following disposition categories will be summarized using tables throughout the study by cohort and overall:

- Participants randomized
- Participants who received study drug
- Participants who completed the study
- Participants who discontinued study drug
- Reasons for discontinuation of study drug
- Participants who did not complete study (see definition below)
- Reasons for not completing study

Listings of participant's dispositions will be provided for the following categories:

- Participants who discontinued study drug
- Participants who did not complete study
- Participants who were randomized yet did not receive study intervention
- Listing of overall study start and end dates

The following definitions will be used in the determination of participant disposition:

Participant Completion of Treatment Definition

A participant will be considered to have completed treatment if, they have completed treatment through [redacted]-week visit including RC visit (may coincide with TAR-200 removal) for both cohort 1 and cohort 2. Participants who prematurely discontinue study treatment for any reason before completion of the treatment phase will not be considered to have completed the study treatment.

Participant Completion of Study Definition

A participant will be considered to have completed the study if they have completed the treatment and follow-up phases or died before the end of the study treatment or follow-up phases. Participants who have been lost to follow-up or have withdrawn consent for study participation before the end of the follow-up phase will not be considered to have completed the study.

Participants who prematurely discontinue study treatment for any reason will remain in the study and all efforts will be made to continue to follow those participants for assessments/procedures outlined in the Follow-Up Phase. However, if a participant develops metastatic disease progression, they should be discontinued from the efficacy (imaging) follow up portion of the study and continued to be followed for survival status, subsequent anti-cancer therapies, and other malignancies per Section 1.3. The only exception to this is when a participant specifically withdraws consent for any further contact with him/her or persons previously authorized by participant to provide this information. All reasonable efforts should be made to maintain the participant on study to ensure accurate assessment of all protocol-specified endpoints.

5.3. Primary Endpoint Analysis

5.3.1. Definition of Endpoint

Pathologic complete response (pCR) rate at RC is defined as the proportion of participants with a pCR or ypT0N0 on RC specimen. pCR will be derived from the analysis of the RC bladder specimen based on central histopathologic review.

5.3.2. Primary Estimand

The primary estimand, the main clinical quantity of interest to be estimated in the study, is defined by the following five components:

- Population: participants with MIBC who are scheduled for RC and are ineligible for or refusing platinum-based neoadjuvant chemotherapy
- Variable: pCR or ypT0N0, at the radical cystectomy
- Study Treatment: TAR-200 + IV cetrelimab, and IV cetrelimab alone
- Intercurrent event: subsequent therapy. Hypothetical Strategy: Response after this intercurrent event will not be considered as a response.
- Summary: pCR rate at the radical cystectomy

5.3.3. Analysis Methods

The Efficacy Evaluable analysis set will be used for the analysis of the primary endpoint pCR. The point estimate of pCR rate and the corresponding 2-sided 80% credible interval (CrI) will be calculated for each cohort. The 80% CrI will be symmetric (i.e., 10% for each tail) and be calculated utilizing a Bayesian approach with Beta(0.5, 0.5) as prior distribution and Beta(0.5+m, 0.5+n-m) as the posterior distribution for the pCR rate, where m is the observed number of participants with pCR and n is the total number of participants.

Additionally, 95% exact confidence interval (CI) will also be calculated for each cohort based on binomial distribution. Summary statistics-frequency counts and percentages by intervention groups will be provided. A side-by-side descriptive summary of efficacy will be provided to illustrate the contribution of TAR-200 to the efficacy of the combination therapy.

The primary analysis will be performed approximately [redacted] months after the 160th participant is treated.

5.3.4. Borrowing Information via Regularized Bayes Method

Two external data from similar PD-1/PD-L1 treatments: CCI [redacted] in a similar disease population will be explored for dynamically borrow information to support the primary analysis of this study. Given the similarities in safety and efficacy of IV cetrelimab with other PD-1 inhibitors, as noted from prior phase 1 and 2 studies, IV cetrelimab was estimated to provide comparable pT0 rates as noted in the CCI [redacted] studies.

The CCI [redacted] are neoadjuvant studies and were designed to assess the efficacy of neoadjuvant anti-PD-1 antibody administration in patients with MIBC who were scheduled for RC and unable to receive platinum-based chemotherapy. Following dosing of systemic checkpoint blockade in the neoadjuvant period CCI [redacted] both studies demonstrated improvements in pathologic downstaging rates at RC as compared to historical controls of transurethral resection of bladder tumor (TURBT) alone CCI [redacted] The overall response rates to these immunotherapies are provided in Table 2.

Table 2: External Data from other PD-1/PD-L1 Inhibitors

	CCI [redacted]
pCR	
80% Credible Interval	

To explore the efficiency of the estimates using borrowing information, Regularized Bayes (RB) method will be used. Regularized Bayes (RB) is a machine learning-based dynamic borrowing approach, which minimizes mean square error (MSE) to ensure efficient estimation of treatment effect by achieving an optimal balance between bias and uncertainty. A regularization term was used to link the current study and historical external data, and the degree of borrowing was dynamically calibrated. The final estimate based on RB is expected to improve the MSE based on independent analysis (IND) of current study data (i.e., no borrowing). Details about the RB method and simulation are provided in Appendix 6.13. The strategy for borrowing is outlined**Error! Reference source not found.** in **Figure 2**.

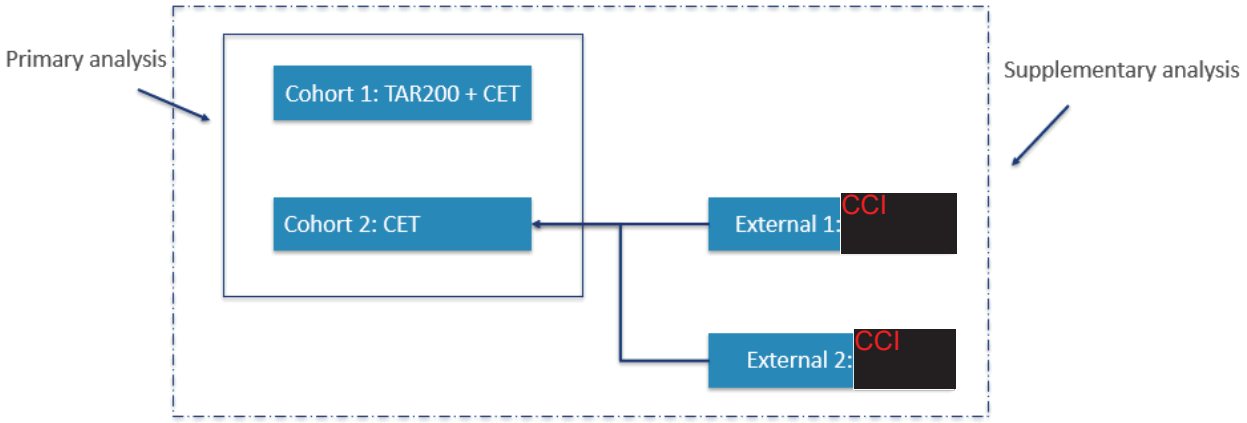


Figure 2: Borrowing strategy from CCI

5.3.5. Sensitivity Analyses

- A sensitivity analysis based on local pathology for each cohort will be performed similarly as described above.
- A sensitivity analysis based on number of TAR-200 insertion and Cetrelimab infusion.
- A sensitivity analysis will be based on participants who have adequate RC results or who have radiographic progression or death before RC
- A sensitivity analysis will be based on Full analysis set

5.4. Secondary Endpoint(s) Analysis

- Recurrence-free survival (RFS)

5.4.1. Definition of Endpoint

Recurrence-free survival (RFS) will be determined in participants receiving TAR-200 in combination with IV cetrelimab and IV cetrelimab alone. RFS is defined as the time from first dose of any study treatment to first radiologic evidence of nodal or metastatic disease that precludes RC, first radiologic evidence of nodal or metastatic disease after RC, or death due to any cause. For participants who are free from the RFS event, data will be censored at the last study disease assessment.

5.4.2. Analysis Methods

The Full Analysis Set will be used for the analysis of RFS. The distribution (median and Kaplan-Meier curves) of RFS for each treatment cohort will be provided using the Kaplan-Meier estimate. In addition, the RFS rate with 95% CI will be estimated by the Kaplan-Meier method and reported for each cohort.

5.5. Exploratory Endpoints/Analysis

Key exploratory analyses:

- Overall Survival (OS) will be determined in participants receiving TAR-200 in combination with IV cetrelimab and IV cetrelimab alone. OS is defined as the time from first dose of any study treatment to death. The Full Analysis Set will be used for the analysis of OS. The distribution of OS will be provided using the Kaplan-Meier estimate for each treatment cohort. In addition, the OS rate with 95% CI will be estimated by the Kaplan-Meier method and reported for each cohort.
- Pathologic overall response (pOR) will be determined at RC in TAR-200 in combination with IV cetrelimab and IV cetrelimab alone. Pathologic overall response (pOR) rate at RC is defined as the proportion of participants with a pCR or =<ypT1N0 on RC specimen. The pOR rate will be tabulated together with its 95% exact confidence interval. The Efficacy Evaluable Analysis Set will be used for the analysis of pOR. A sensitivity analysis will be based on Full analysis set.
- Time to symptomatic progression (TSP) will be determined in participants receiving TAR-200 in combination with IV cetrelimab and IV cetrelimab alone. TSP is defined as the time from first dose of any study treatment to documentation of any of the following (whichever occurred earlier):
    - Progression of pain or worsening of disease-related symptoms requiring initiation of a new systemic anticancer therapy
    - Development of clinically significant symptoms due to locoregional tumor progression requiring surgical treatment or radiation therapy.

The Full Analysis Set will be used for the analysis of TSP. The distribution of TSP will be provided using the Kaplan-Meier estimate for each treatment cohort. In addition, the TSP rate with 95% CI will be estimated by the Kaplan-Meier method and reported for each cohort.

5.6. Safety Analyses

All safety analyses will be based on the safety analysis set based on actual intervention received unless otherwise specified.

For all continuous safety variables, descriptive statistics by intervention group will include the N, mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized by intervention group using frequency counts and percentages.

5.6.1. Extent of Exposure

The number and percentage of participants who receive each study drug (TAR-200 + IV cetrelimab, IV cetrelimab alone) within a study intervention will be summarized.

Descriptive statistics for the duration of each study drug within a study intervention (N, mean, SD, median, and range [minimum, maximum]) will be summarized.

Duration of intervention will be summarized in the following duration categories: [<3 weeks, 3-<6 weeks, 6-<9 weeks, 9-<12 weeks, >=12 weeks] for each study drug within a study intervention and presented graphically in a histogram. Cumulative duration of intervention [>=3 weeks, >=6 weeks, >=9 weeks, >=12 weeks] will be summarized.

The total duration of the intervention is defined as (date of the last dose of study intervention – date of the first dose of study intervention) +1. Descriptive statistics (N, mean, SD, median, and range [minimum, maximum]) of the total duration of the intervention will be summarized for each study drug within a study intervention.

Total dose days of intervention are defined as the total number of days that study intervention was administered to the participant (excluding days off study intervention). Descriptive statistics (N, mean, SD, median, and range [minimum, maximum]) of total dose days of intervention will be summarized for each study agent within a study intervention. The number (%) of participants and the number of doses will be summarized. Study intervention compliance will be summarized descriptively. See Appendix 7 for details.Appendix 7

5.6.2. Adverse Events

The verbatim terms used in the CRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any new or worsening AEs with onset date and time after the first dose through 100 days after the last study drug administration or prior to the start of subsequent anticancer therapy, whichever is earlier, or the follow-up AE (linked to an existing TEAE) with onset date and time beyond 100 days after the last dose of study intervention but prior to the start of subsequent therapy or any AE that is considered treatment-related regardless of the start date of the event, is considered to be treatment emergent. If the event occurs on the day of the initial administration of study intervention, and either event time or time of administration are missing, then the event will be assumed to be treatment emergent. If the event date is recorded as partial or completely missing, then the event will be considered to be treatment emergent unless it is known to be prior to the first administration of study intervention or later than 100 days after last study drug administration based on partial onset date or resolution date. If the event is considered drug-related regardless of the start date of the event, or the event that is present at baseline but worsens in toxicity grade or is subsequently considered drug-related by the investigator, then this event will be assumed to be treatment emergent. All reported treatment-emergent adverse events will be included in the analysis. Immune-related AEs will be determined by the investigator only. For each adverse event, the number and percentage of participants who experience at least 1 occurrence of the given event will be summarized overall and by cohort.

Summary tables will be provided for treatment-emergent adverse events:

- AEs
- Serious AEs (SAEs)
- Grade 3 or higher AEs (G3+ AEs)
- AEs leading to discontinuation of each study drug within a study intervention
- AEs and SAEs based on NCI-CTCAE (version 5.0) toxicity grade
- AEs and SAEs by relationship to each study drug (TAR-200, IV cetrelimab) and to UPC or procedure (TAR-200 insertion or removal)
- AEs leading to dose interruption of each study agent within a study intervention
- AE due to TAR-200 removal, not per-protocol

- Other safety observations (e.g., immune-related AEs, infusion reactions) by preferred term (PT) and toxicity grade

In addition to the summary tables, listings will be provided for participants who:

- Had SAEs
- Had AEs leading to discontinuation of study intervention

Any AE occurring in the follow-up phase will also be summarized similarly

Deaths will be displayed by actual intervention received. Frequencies for the following parameters will be included in the summary table:

- Number of participants who died
- Cause of death
- Relationship to study intervention, drug, or procedure (yes/no)

A listing of participants who died will be provided.

5.6.3. Clinical Laboratory Tests

Descriptive statistics will be presented for select chemistry, hematology, and urinalysis laboratory tests at scheduled time points. Change from baseline to each scheduled time point will be summarized for select chemistry, hematology, and urinalysis tests by intervention group.

The number and percentage of participants who meet the abnormality based on normal ranges will be provided by treatment group at scheduled time points. Listing will also be provided.

Applicable laboratory results will be graded according to NCI-CTCAE version 5.0. The toxicity grades and the shift from baseline to post-baseline will be summarized at scheduled time points. The worst toxicity grades and the shift from baseline to the worst toxicity grades will also be provided. Listing will also be provided.

5.6.4. Electrocardiogram

Electrocardiogram (ECG) measurements are taken at baseline and end of treatment visits. ECG interpretations (clinically significant, not clinically significant) will be summarized at baseline and end of treatment visits by cohort.

Listings will be produced for abnormal or clinically significant ECG data including unscheduled visit data

5.6.5. Vital Signs

Vital sign parameters including heart rate, weight and blood pressure (systolic and diastolic) will be summarized descriptively at each assessment timepoint by treatment group. Descriptive statistics (mean, standard deviation, median, minimum and maximum) of vital signs and the change from baseline will be presented by treatment group.

Abnormality criteria (based on criteria defined in Table 3) will be applied to baseline and postbaseline values. For baseline values, increase or decrease criteria are not applied.

Postbaseline values will be considered treatment emergent (TE) if they meet both value and change criteria in the table below. For criteria that do not include an increase or decrease from baseline:

- TE will be concluded if the postbaseline value is above the upper limit and the baseline value is below the upper limit (e.g., Normal or Low). The same applies to the postbaseline value being below the lower limit with the baseline value being above the lower limit (e.g., Normal or High).

- If the baseline value is missing, a postbaseline abnormality will always be considered as TE.

Incidence of markedly abnormal vital signs during intervention, as defined in Table 3, will be summarized for participants who had a baseline assessment and at least 1 postbaseline assessment for that vital sign. A listing of participants with TE abnormal vital signs will be presented.

Table 3: Markedly Abnormal Vital Signs

Vital Sign	Criteria
Pulse	>110 bpm and with >30 bpm increases from baseline
	<50 bpm and with >20 bpm decreases from baseline
Systolic blood pressure	>180 mm Hg and with >40 mm Hg increase from baseline
	<90 mm Hg and with >30 mm Hg decrease from baseline
Diastolic blood pressure	>105 mm Hg and with >30 mm Hg increase from baseline
	<50 mm Hg and with >20 mm Hg decrease from baseline
Weight	Increase 10% kg from baseline
	Decrease 10% kg from baseline

5.7. Other Analyses (including other exploratory endpoints)

5.7.1. Pharmacokinetic (PK)

Sparse PK analysis set includes all randomized subjects who received at least 1 dose of TAR-200 or IV cetrelimab and had at least 1 evaluable PK sample (plasma/serum sample or sparse urine PK sample) obtained posttreatment in cohort 1 (TAR-200 + IV cetrelimab), and cohort 2 (IV cetrelimab).

Serial PK analysis set includes all randomized participants assigned into serial urine PK collection who received at least 1 dose of TAR-200 and had at least 1 evaluable PK sample obtained posttreatment in cohort 1 (TAR-200 + IV cetrelimab).

Sparse PK analysis

TAR-200

Gemcitabine, dFdU urine, as well as plasma concentration and gemcitabine and dFdU urine amount will be listed by group. All concentrations below the lowest quantifiable concentration or missing data will be labeled as such in the listing.

If feasible, population PK analysis of urine concentration-time data of TAR-200 may be performed using nonlinear mixed-effects modeling. Data may be combined with those of other selected studies to support a relevant structural model. Available baseline participant characteristics (demographics, laboratory variables, genotypes, race, etc.) will be tested as potential covariates affecting PK parameters. Details will be given in a population PK analysis plan and the results of the population PK analysis will be presented in a separate report.

Cetrelimab

Cetrelimab concentrations will be summarized at each timepoint using descriptive statistics (arithmetic mean, SD, coefficient of variation, geometric mean, geometric CV, median, minimum, and maximum) and figures. All concentrations below the lowest quantifiable concentration or missing data will be labeled as such in the concentration data presentation. Concentrations below the lowest quantifiable concentration will be treated as zero in the summary statistics and for the calculation of PK parameter(s). All participants and samples excluded from the analysis will be clearly documented in the study report.

If feasible, population PK analysis of serum concentration-time data of IV cetrelimab may be performed using nonlinear mixed-effects modeling. Data may be combined with those of other selected studies to support a relevant structural model. Available baseline participant characteristics (demographics, laboratory variables, genotypes, race, etc.) will be tested as potential covariates affecting PK parameters. Details will be given in a population PK analysis plan and the results of the population PK analysis will be presented in a separate report.

Serial PK analysis

This analysis is applicable to serial PK analysis set. For serial urine PK, PK parameters will be listed and summarized using descriptive statistics. Concentrations below the lowest quantifiable concentration will be treated as zero in the summary statistics and for the calculation of PK parameters. All participants and samples excluded from the analysis will be clearly documented in the study report. Details of the analyses will be described in the Clinical Pharmacology Analysis Plan.

5.7.2. Immunogenicity of Cetrelimab

The incidence of anti-cetrelimab antibodies will be summarized for all participants who receive at least 1 dose of cetrelimab and have appropriate samples for detection of antibodies to cetrelimab (ie, participants with at least 1 sample obtained after their first dose of cetrelimab). A listing of participants who are positive for antibodies to cetrelimab will be provided. The maximum titers of antibodies to cetrelimab will be summarized for participants who are positive for antibodies to cetrelimab by cohort. Other immunogenicity analyses may be performed to further characterize the immune responses that are generated.

5.7.3. Biomarkers

Expression of PD-L1 data from tissue samples that are associated with RFS and OS will be summarized by treatment groups. A subgroup summary may also be provided.

Expression of PD-L1 data will be listed for all subjects with available evaluable samples. All missing data will be labeled as such in the database. All subjects and samples excluded from the listing will be clearly documented in the study report.

Descriptive statistics frequency count and percentage will be provided. Graphical exploration of data may be performed as deemed useful.

5.7.4. Patient-Reported Outcomes (PRO)

Patient-reported outcome instruments measure changes in health-related quality of life (HRQoL) for participants treated with TAR-200 and IV cetrelimab compared to IV cetrelimab alone. The disease specific PRO captured in this clinical trial is the Functional Assessment of Cancer Therapy – Bladder (FACT-Bl). The Patient Global Impression of Severity (PGIS) and the Patient Global Impression of Change (PGIC) will also be administered to participants within the study to provide anchor-based assessment.

FACT-BI

Participant's cancer-related HRQoL will be assessed by using the FACT-Bl (Functional Assessment of Cancer Therapy – Bladder). The FACT-Bl consists of 36 items, with 5-point Likert scales, covering 5 primary domains: physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), functional well-being (FWB), and additional concerns for patients with bladder cancer (BLC). The response options range from "Not at all" to "very much."

PGIS

The PGIS is a single question regarding the patient report of disease severity: "Considering all aspects of your bladder cancer symptoms right now, would you say your bladder cancer symptoms are" (1) None, (2) Mild, (3) Moderate, (4) Severe, or (5) Very Severe.

PGIC

The PGIC is a single question regarding the patient report of disease change: "Compared to when you received the first treatment in this study, how has your cancer changed (1) A lot better now, (2) Moderately better now, (3) A little better now, (4) Neither better, nor worse (no change), (5) A little worse now, (6) Moderately worse now, or (7) A lot worse now?".

Analysis Method

PRO data will be collected as outlined in the SoA (Section 1.3) in the protocol. The analyses of PRO endpoints are considered exploratory and as such, any p-values reported are considered as nominal. No adjustments will be made for multiplicity. For each PRO assessment, scores will be summarized with descriptive statistics at each assessment time point and graphical summaries will be provided. Where variables are summarized in terms of change from baseline, the baseline mean score of those subjects with non-missing change scores will also be included.

For FACT-Bl scores, a mixed-effects model using repeated measures (MMRM) based on observed data will be performed to obtain the least squares (LS) mean change estimates from baseline and 95% CI at each assessment time point. Participants who have a baseline measurement and at least one post-baseline value will be included in the analysis. If necessary, truncation will be applied for all subsequent visits at the first visit where 90% or more of the subjects are missing the FACT-Bl assessment from either treatment group. Once the truncation cycle is determined, this cycle will be applied across both treatment groups.

Time to symptom worsening is the time from first dose until the date of the first clinically meaningful worsening, defined using the following change score thresholds derived from distribution-based methods (1/2 SD of baseline mean) and published literature: PWB=-3, SWB=-3, EWB=-2, FWB=-3, BLC=-4, TOI=-9, FACT-G=-8, FACT-Bl=-11. Time to worsening in FACT-Bl total score and individual scales will be analyzed using a Kaplan Meier method and stratified Cox proportional hazard model. Death due to disease progression will be considered as worsening. Participants who have not met the definition of worsening will be censored at the last PRO assessment. Participants without baseline assessment or post-baseline assessment will be censored at date of treatment start date.

Compliance rates for completion of PROs at each time point will be generated and will be represented as the actual number of assessments received over the number of expected assessments for each scheduled visit of PRO collection. Completion rates will be calculated for the FACT-Bl. Participants disposition for PRO completion will also be described.

5.7.5. Medical Resource Utilization (MRU) and Health Economics Analyses

To assess number of participants with type and length of inpatient stay and overall medical encounters, MRU data, associated with medical encounters, will be collected on an ongoing basis whenever an encounter occurs, including information regarding utilization of healthcare services (including the timing and type of services). Protocol-mandated procedures, tests, and encounters are excluded.

Descriptive summary of MRU data will be provided, including the number and duration of medical encounters, duration of hospitalization, and frequency of outpatient medical encounters by treatment cohort.

5.7.6. Subgroups Analysis

Subgroup analysis will be performed for the selected potential prognostic variables (as listed in the table below) to assess the consistency and robustness of the treatment benefit. The subgroup variables and the cut-off values are subject to change, if warranted, to better represent the data.

Subgroup analysis will be presented graphically in a forest plot.

Table 4: Subgroup Variables and the Definitions

Subgroup	Definition
Region	America: USA Asia: South Korea, Israel Western Europe: Belgium, France, Germany, Italy, Netherlands, Poland, Spain, UK
Age Group	⩽ 70 ＞70
Gender	Female Male
Race	American Indian or Alaska Native Asian Black or African American Native Hawaiian or other Pacific Islander White
Nicotine Use History	Current Former Never
Tumor Stage at Screening	cT2 cT3-4a
TURBT at Screening	Visibly complete Incomplete and ⩽3 cm
Baseline ECOG status	0 1
PD-L1 status	High Low/Negative

5.8. Interim Analyses

Two interim analyses are planned. The first interim analysis (IA1) will be conducted after approximately 40 participants in total (i.e., approximately 25 in Cohort 1 and approximately 15 in Cohort 2) complete the RC and have the disease assessment for pCR at RC. The second interim analysis (IA2) will be after approximately 80 participants in total (i.e., approximately 50 in Cohort 1 and approximately 30 in Cohort 2) complete the RC and have the disease assessment for pCR at RC. The analyses will utilize the Bayesian approach described in Section 5.3.3.

At IA1, the cetrelimab cohort may be stopped for futility if the posterior probability of pCR rate <25% is greater than 80%, i.e, the number of participants with pCR is <3 out of 15. The criteria ensure at least 60% probability of early termination for futility if the true pCR rate for the cetrelimab cohort is less than 15%.

At IA2, the cetrelimab cohort may be stopped for futility if the posterior probability of pCR rate <25% is greater than 80%, ie, the number of participants with pCR is ⩽5 out of 30. Similarly, an early success may be declared for the TAR-200 and cetrelimab cohort if the posterior probability

of pCR rate >45% is greater than 80%, i.e., the number of participants with pCR is ≥26 out of 50. The criteria ensure at least 71% probability of early termination for futility if the true pCR rate for the cetrelimab cohort is less than 15%, or 71% probability of early success for the combination if the true pCR rate for the combination cohort is greater than 55%.

5.8.1. Independent Data Monitoring Committee (IDMC)

An IDMC of at least one medical expert in the relevant therapeutic area and at least one statistician will be established to monitor data on an ongoing basis to ensure the safety of the subjects enrolled in this study. The safety review will focus on deaths, treatment discontinuations, serious adverse events, Grade ≥3 events, and emerging safety data. Based on the results from these scheduled safety review meetings, the IDMC chair may request additional safety monitoring measures. Throughout the conduct of the study, all individual patient deaths, treatment discontinuations, and serious adverse events will be reviewed by the Sponsor's medical monitor and the safety physician on an ongoing basis to identify specific patient safety issues, and the IDMC will be informed of any new potential signals of concern. The plan for monitoring subject safety and evaluating the efficacy, and the roles and responsibilities of the IDMC, are detailed in the IDMC Charter.

The plan for monitoring subject safety and evaluating the efficacy, and the roles and responsibilities of the IDMC, are detailed in the IDMC Charter. Please refer to the IDMC Charter and IDMC SAP for details as to specific outputs to be reviewed by the IDMC for the futility and efficacy analysis.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1: List of Abbreviations

AE	adverse events
ALT	alanine aminotransferase
ANC	absolute neutrophil count
AST	aspartate aminotransferase
CR	complete response
CrI	credible interval
CT	computed tomography
dFdU	2′,2′-difluorodeoxyuridine
ECG	electrocardiogram
eCRF	electronic case report form(s)
EORTC	European Organization for Research and Treatment of Cancer
ECOG	Eastern Cooperative Oncology Group
FACT-Bl	Functional Assessment of Cancer Therapy – Bladder
FDA	Food and Drug Administration
HRQoL	health-related quality of life
ICF	informed consent form
ICH	International Council for Harmonisation
IDMC	Independent Data Monitoring Committee
irAE	immune-related adverse event
IV	intravenous
MedDRA	Medical Dictionary for Regulatory Activities
MIBC	Muscle Invasive Bladder Cancer
MRI	magnetic resonance imaging
MRU	medical resource utilization
OS	overall survival
pCR	pathologic complete response
PD-1	programmed-cell death protein 1
PD-L1	programmed-cell death ligand 1
PD-L2	programmed-cell death ligand 2
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PK	pharmacokinetic(s)
pOR	pathologic overall response
pPR	pathologic partial response
PRO (ePRO)	patient-reported outcome(s) (paper or electronic as appropriate for this study)
Q2W	once every 2 weeks
Q3W	once every 3 weeks
Q4W	once every 4 weeks
RC	radical cystectomy
RFS	recurrence-free Survival
SAE	serious adverse event
TURBT	transurethral resection of bladder tumor
UTI	urinary tract infection
WBC	white blood cell

6.2. Appendix 2: Changes to Protocol-Planned Analyses

There has been a correction of TSP definition. The new definition removed

'Development of symptomatic deterioration on the basis of global deterioration of health status', since the global deterioration of health status based on ECOG evaluation may results from other causes than disease progression.

6.3. Appendix 3: Demographics and Baseline Characteristics

The number of participants in each analysis set will be summarized and listed by cohort and overall. In addition, the distribution of participants by region, country/territory, and site ID will be presented unless otherwise noted.

Table 5 and Table 6 present a list of the demographic variables and baseline characteristics that will be summarized by cohort and overall, for the Full and Efficacy Evaluable Analysis Sets. Demographics and baseline characteristics will also be summarized by region using the Full Analysis Set.

Table 5: Demographic Variables

Continuous Variables:	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum], and IQ range).
Weight (kg)	
Height (cm)	
Body Mass Index (BMI) (kg/m^2)	
Categorical Variables	
Age (≤ 70, >70)	Frequency distribution with the number and percentage of participants in each category.
Sex (male, female, undifferentiated)	
Region[a] (America, Asia, Western Europe)	
Race[b] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Multiple)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino)	
BMI ([underweight <18.5 kg/m2, normal 18.5-<25 kg/m2, overweight 25-<30 kg/m2, obese >=30 kg/m2])	

[a]America includes USA; Asia includes South Korea, Israel; Western Europe includes Belgium, France, Germany, Italy, Netherlands, Poland, Spain, UK
[b]If multiple race categories are indicated, the Race is recorded as 'Multiple'

Table 6: Baseline Characteristics

Categorical Variables	**Summary Type**
Nicotine Use History (Current, former, never)	Frequency distribution with the number and percentage of participants in each category.
Baseline ECOG (0, 1)	
Tumor Stage (cT2, cT3-4a), Tumor sites, Type of Histology	
Lymph Node Stage (N0-N3, NX)	
Metastasis Stage (M0, M1, MX)	
PD-L1 status (high vs. low/negative)	

6.4. Appendix 4: Protocol Deviations and Quality Tolerance Limits (QTL)

In general, the following list of major protocol deviations may have the potential to impact participants' rights, safety, or well-being, or the integrity and/or the primary endpoint of the clinical study. Participants whose major protocol deviations are determined to meet the above criteria (have the potential to impact participants' rights, safety or well-being, or the integrity and/or the primary endpoint of the clinical study) will be identified by medical and statistical review prior to database lock and will be summarized by the following categories:

- Developed withdrawal criteria but not withdrawn
- Entered but did not satisfy criteria
- Received a disallowed concomitant treatment
- Received wrong treatment or incorrect dose
- Other

A Protocol Deviation Specification (PDS) has been developed to provide more information about the major protocol deviations. Periodic meetings are required to investigate each potential protocol deviation.

Quality Tolerance Limit parameters and thresholds are defined and will be monitored in this study. QTL parameters will be summarized. More details are described in the Integrated Analytical Risk-Based Monitoring (iARBM) Plan.

6.5. Appendix 5: Prior and Concomitant Medications

Prior and Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). Prior medications are defined as any therapy used before the day of the first dose (partial or complete) of the study intervention. Concomitant medications are defined as any therapy used on or after the same day as the first dose of the study drug, including those that started before and continue after the first dose of the study drug.

Descriptive summaries of concomitant medications will be presented by treatment cohort. The proportion of participants who receive each concomitant medication will be summarized as well as the proportion of participants who receive at least 1 concomitant medication.

Prior medications will be summarized by treatment cohort.

6.6. Appendix 6: Medical History

A listing of participant's medical history will be provided.

6.7. Appendix 7: Intervention Compliance

Compliance will be summarized descriptively for each study drug within a study intervention. Compliance with randomized intervention versus actual intervention will be presented in a summary table.

Compliance will be calculated as follows:

For Cohort 1: Compliance (%) = number of TAR-200 insertions completed/the number of TAR-200 insertions expected

For Cohorts 1 and 2: Compliance (%) = number of IV cetrelimab infusions completed/the number of IV cetrelimab infusions expected

6.8. Appendix 8: Adverse Events of Special Interest

There are no AEs of special interest.

6.9. Appendix 9: Medications of Special Interest

Concomitant medications of special interest include steroids and immunosuppressive medications used to treat immune-related adverse events related to IV cetrelimab.

6.10. Appendix 10: ECOG Performance Status Scale[5]

Grade	ECOG Performance Status Grade
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light housework, office work
2	Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair
5	Dead

6.11. Appendix 11: PRO – Patient Reported Outcome

6.11.1. FACT-BI

CCI

[redacted]

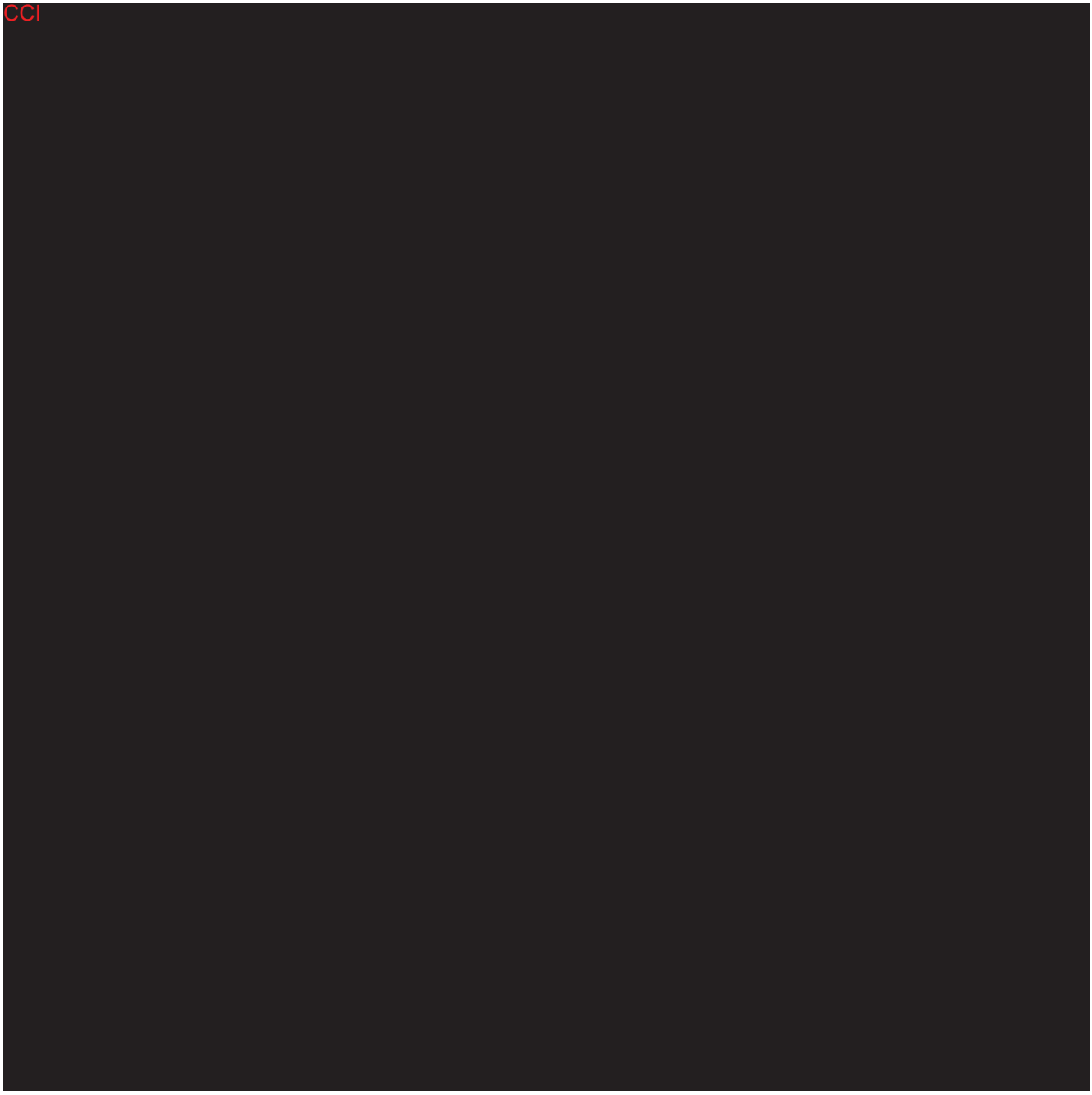

CCI


CCI

6.11.2. FACT-BI Scoring Guidelines

CCI

[redacted]


CCI

6.11.3. PGIS - Patient's Global Impression of Severity of Cancer

6.11.4. PGIC - Patient's Global Impression of Change of Cancer

CCI

6.12. Appendix 12: Cardiovascular Risk Assessment Prior to Study Enrollment and Prior to Radical Cystectomy

The individual cardiovascular risk of each study participant must be assessed during the screening period prior to enrollment to the study. Participants undergoing radical cystectomy are at a high risk for cardiovascular and thromboembolic events. Participants can only be enrolled when the risk assessment has been performed and all subsequent steps for diagnostic procedures and cardiac clearance have been conducted as outlined below, which needs to be documented.

In addition, the cardiovascular risk of each study participant must be reassessed prior to radical cystectomy (RC) as outlined in the Schedule of Activities: i.e., at approximately 1-2 weeks prior to the anticipated RC date. Diagnostic assessment and subsequent steps for diagnostic procedures and cardiac clearance prior to surgery must be conducted and documented. If a participant is assessed as not eligible for RC, surgery must be delayed until cardiac clearance is obtained. If cardiac clearance cannot be achieved, the participant should continue on study, but the Investigator might decide to adjust the treatment plan for a participant and consider alternative treatment options. In such cases, the Investigator should contact the Sponsor to agree on next steps.

To assess the cardiovascular risk of participants prior to enrollment in this study (17000139BLC2002) and the actual cardiac risk prior to RC, the formula outlined below adapted from the revised Cardiac Risk Index for Pre-Operative Risk (RCRI) should be used to calculate participants' point values reflecting the individual cardiovascular risk. The revised RCRI accurately stratifies participants based on their individualized risk prior to surgery.

Assessment of Individual Cardiovascular Risk

Risk Factor	Description	Points
History of ischemic heart disease	• History of myocardial infarction • Electrocardiogram with pathological Q waves • History of positive exercise test • Current chest pain considered due to myocardial ischemia/angina • Use of nitrate therapy • History of coronary artery disease • Prior percutaneous coronary intervention • Prior coronary artery bypass surgery	+1
History of congestive heart failure	• Pulmonary edema, bilateral rales or S3 gallop • Paroxysmal nocturnal dyspnea • Chest x-ray showing pulmonary vascular redistribution	+1
History of cerebrovascular disease	• Prior transient ischemic attack • Past stroke	+1
History of cardiac valvular disease	• Aortic stenosis	+1
Other risk factors[a]	• Peripheral arterial disease • Atrial fibrillation or other cardiac arrythmia	+1
Pre-operative treatment with insulin		+1
Pre-operative creatinine >2.0 mg/dL (176.8 µmol/L)		+1

[a] Only risk factors listed specifically on the RCRI form should be counted in risk assessment. Any other risk factors such as hypertension, hyperlipidemia, obesity, smoking, diabetes, etc, will not be included in the RCRI however, these factors should be taken under consideration for monitoring cardiac safety throughout the trial.

Interpretation of Risk Assessment and Next Steps for Assessment

Revised Cardiac Risk Index Score/Points*	Next steps during the study (17000139BLC2002) screening period	Next steps prior to RC
0	No additional diagnostic work-up	Standard pre-operative surgical and anesthesiology assessment; subsequent diagnostic procedures and clearance for cardiac risk factors prior to surgery depending on results of pre-operative assessment
≥1	Cardiological consultation prior to enrollment and assessment for eligibility for RC	Cardiological consultation prior to surgery; subsequent diagnostic procedures and cardiac clearance prior to surgery depending on results of pre-operative assessment.

* Any changes in risk factors between study entry and date of RC should prompt re-evaluation.

6.13. Appendix 13: Dynamic Borrowing of External Data

6.13.1. Materials and Methods

Define, $i = 1, \ldots, m$, where 1 refers to the current study and 2 to m refers to the historical studies. Let $D = \{D_1, \ldots, D_m\}$ represent the control data of the current and the historical trials. Then, the likelihood of the current and the historical trials are $L(\beta_i|D_i); i = 1, \ldots, m$. We are interested in improving the estimation of the true effect size β_1 using the estimation of the historical effect sizes $\beta_i; i = 2, \ldots, m$.

6.13.1.1. Independent Analysis (IND)

This method assumes non-ignorable heterogeneity between studies, i.e., borrowing information would mislead decisions and not support gaining efficiency while increasing the bias and FPR. Using this method, no information is borrowed across studies with independent non-informative prior distributions on effect sizes of each study, i.e., $\beta_i \sim N(0, \sigma^2);\ i = 1, \ldots, m$.

6.13.1.2. Pooled Analysis (Full).

Unlike the independent analysis, the pooled analysis assumes studies are entirely homogeneous and the true effect sizes are the same. That is, $\beta_1 = \cdots = \beta_m = \beta$, and β has non-informative prior, where $\beta \sim N(0, \sigma^2)$.

6.13.1.3. Meta-analytic Prior (MAP)

The meta-analytic-predictive (Neuenschwander et al., 2010) approach accounts for heterogeneity in effect sizes between the current study and the historical studies when quantifying the posterior distribution of the effect size of the current study. The MAP approach assumes that the model parameters of the trials are exchangeable and each β_i follows the same normal distribution, $\beta_i \sim N\left(\mu, \frac{1}{\tau}\right); i = 1, \ldots, m$ and non-informative hyperparameter priors as $\mu \sim N(0, \sigma^2)$ and $\tau \sim Gamma(r, \lambda)$, where the degree of borrowing τ is estimated based on the joint posterior distribution. The MAP prior is generally derived from a random-effect meta-analysis of historical data via Markov Chain Monte Carlo (MCMC) algorithms as the analysis is commonly not tractable analytically. The difference between the MAP and the conventional meta-analysis approach is that the MAP aims to estimate the current trial parameter, whereas meta-analysis aims to estimate the overall mean parameter.

6.13.1.4. Regularized Bayes (RB)

RB performs dynamic borrowing of information to achieve optimal performance using a regularization term, called the Regularized Bayes (RB) approach. This approach calibrates the trade-off between bias and uncertainty associated with borrowing via an explicitly defined objective function to achieve an optimal balance. The degree of borrowing is determined by minimizing the mean square error (MSE) of the group of interest alone (i.e., current data), which minimally depends on the exchangeability assumption relying on external data. The optimally calibrated degree of borrowing is then incorporated into the Bayesian model.

The Regularized Bayes approach formularizes β_1 for the current study and $\beta_i = \beta_1 + \delta_i, i = 2, \ldots, m$ for the historical studies. The independent MLEs of each study, without borrowing, are

denoted by $\hat{\beta}_1^*, \ldots, \hat{\beta}_m^*$. By adding a regularization term, $\frac{\tau}{2}\sum_{i=2}^{m} w_i(\beta_1 - \beta_i)^2$, to borrow information, the RB approach solves the question given a fixed amount of borrowing τ, written as

$$\hat{\beta}_1(\tau) = \text{argmin} - l(\beta_1) + \frac{\tau}{2}\sum_{i=2}^{m} w_i(\beta_1 - \beta_i)^2 - \sum_{i=2}^{m} l(\beta_i) \qquad (1)$$

where β_1 is the true effect of interest of the current study; and $l(\beta_1)$ and $l(\beta_i)$ are the log-likelihood of the current and i^{th} studies, respectively. τ is a non-negative tuning parameter. A large value of τ enforces more similarity between the current and the historical studies, resulting in more borrowing. $\tau = 0$ indicates no-borrowing and as $\tau \to \infty$, the RB approach tends to borrow the full historical data (see Figure 3). Furthermore, $w_i = \frac{1}{(\hat{\beta}_1^* - \hat{\beta}_i^*)^2}$ accounts for the different borrowing weights based on the estimates without borrowing.

The degree of borrowing, τ, is calibrated by minimizing the MSE of the current study $E(\hat{\beta}_1(\tau) - \beta_1)^2$, which can be approximately solved by a bootstrapping method and with β_1 approximated by the MLE, $\hat{\beta}_1^*$, without borrowing. Bootstrapping is a random sampling approach to approximate the distribution of statistic[7].

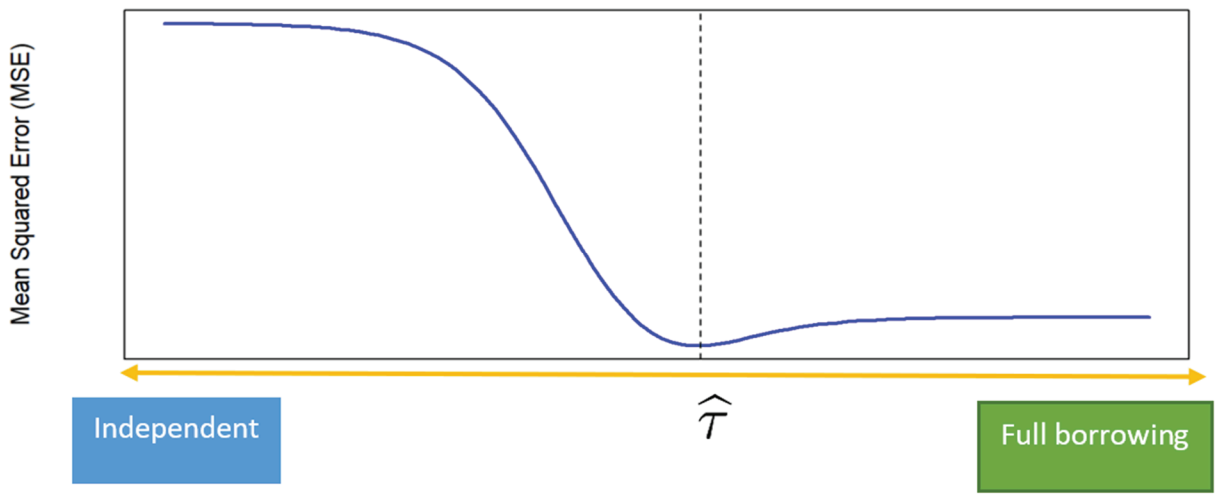


Figure 3: Plot of MSE vs. $\hat{\tau}$. An example of how MSE varies with the estimate of τ while borrowing information. MSE is approximated using the MLE of β_1. The vertical line represents the estimated τ based on the bootstrapping at which the MSE is minimum and the balance between the bias and variance is optimal.

After the degree of borrowing is obtained, the Bayesian model is used to obtain the estimate and credible interval (CrI) with prior distributions $\beta_1 \sim N(0, \sigma^2), \delta_i \sim N(0, 1/(\hat{\tau} w_i))$ and make inference.

Algorithm 1: Algorithm of Regularized Bayes (RB)

1) Denote $\hat{\beta}_1^*$ and $\hat{\beta}_i^*;\ i = 2, \dots, m,$ as the maximum likelihood estimate (MLE) of the current and i^{th} historical studies, respectively. Obtain the MLEs for all studies.
2) Denote $\widehat{w}_i = \frac{1}{(\hat{\beta}_1^* - \hat{\beta}_i^*)^2}$ as the borrowing weight based on the estimates without borrowing of the i^{th} study. Obtain $\widehat{w}_i$ for all historical studies.
3) **For $\boldsymbol{\tau = \tau_1, \dots, \tau_k}$, where $\boldsymbol{\tau \in [0, \infty)}$ is the degree of borrowing, do:**

   i) Denote $\hat{\beta}_{1b}(\tau)$ as the estimate of the true effect using current and historical data for the b^{th} bootstrap. Obtain $\hat{\beta}_{1b}(\tau); b = 1, \dots, B$, via optimization problem (1).
   ii) Denote $MSE_\tau = E(\hat{\beta}_1(\tau) - \hat{\beta}_1^*)^2 \approx \frac{1}{B}\sum_{b=1}^{B}(\hat{\beta}_{1b}(\tau) - \hat{\beta}_1^*)^2$ as the MSE. Calculate MSE_τ using bootstrapping.

   end;
4) Denote $\hat{\tau}$ as the optimal value of τ. Obtain $\hat{\tau}$ that corresponds to the minimum MSE_τ such that $\hat{\tau} = argmin\ E(\hat{\beta}_1(\tau) - \hat{\beta}_1^*)^2$.
5) Fit Bayesian model using the optimized $\hat{\tau}$.

6.13.2. Simulation Study

We conducted simulations assuming a binary response as is the primary objective in this study. The RB method can also be applied to other types of responses. The simulations evaluated the performance of the RB and the other Bayesian Borrowing methods aforementioned: 1) Independent Analysis (IND; no-borrowing), 2) Pooled Analysis (FULL; full-borrowing), and 4) Meta-analytic Prior (MAP), where IND is used as the benchmark. The performance was evaluated in terms of false-positive rate (FPR) and true positive rate (TPR).

Define, $\text{i} = 1, \dots, \text{m}$, where 1 refers to the current study and 2 to m refers to the historical studies. Let Y_i denote the number of participants achieving pathological complete response (pCR) and N_i denote the sample size of the i^{th} study. Under the Bayesian model, let Y_i denote a random variable with a binomial distribution conditioned on the true pCR rate, P_i, such that $\text{Y}_\text{i} = \text{Binomial}(\text{N}_\text{i}, \text{P}_\text{i})$ and $\text{logit}(\text{P}_\text{i}) = \beta_\text{i}; \text{i} = 1, \dots, \text{m}$, where β_i is the true effect size. In addition, $\sigma^2 = 1000$ and $r = \lambda = .001$ was used for the non-informative prior.

For the RB method, the degree of borrowing, τ, is calibrated by minimizing the MSE of the current study such that

$$\hat{\tau} = \text{argmin}\ \text{E}(\hat{\text{p}}_1(\hat{\tau}) - \text{p}_1)^2 \approx \text{argmin}\frac{1}{\text{B}}\sum_{\text{b}=1}^{\text{B}}(\hat{\text{p}}_{1\text{b}}(\hat{\tau}) - \hat{\text{p}}_1^*)^2 \text{; where } \text{logit}(\text{p}_1) = \beta_1.$$

Furthermore, we considered two sample sizes $(\text{n} = 30, 60)$ for the current study and two historical simulated data with sample sizes 88 and 114 to mimic the actual application of the two historical anti-PD-1/PD-L1 antibody clinical trials: ABACUS (Powles et al., 2019) and PURE-01 (Necchi et al., 2020), respectively. The current data were simulated for the true pCR rate $(\text{pCR} = .15,\ .20,\ .25,\ .30,\ .35,\ .40,\ .45,\ .50)$ and historical data were generated for the mean pCR rate $(\text{HpCR} = .20,\ .25,\ .30,\ .35,\ .40)$. Two hypothesized pCR cutoffs $(\text{p}_0 = .20, .30)$ and 80% Credible Interval (CrI) were used to obtain FPR and TPR. Note that, in the frequentist approach, p_0 can be considered as the value of a null hypothesis.

6.13.2.1. Simulation Results

Figure 4 and Figure **5** shows the false positive rate (FPR) and true positive rate (TPR, respectively, of the methods for the sample sizes and historical pCR rates against the true pCR. FPR was controlled at $\alpha = 0.2$.

The FPR is controlled for the no-borrowing (IND) method regardless of whether the true pCR or historical pCR matches, which is expected since IND does not borrow historical information; whereas other borrowing methods are affected when the historical pCR rate is higher than the true pCR rate (Figure 4). RB and MAP have comparable inflated false positive rate (FPR) of at most 10%, but the RB tends to control better when there is a large discrepancy between the current and historical data, which supports that RB can recognize when there is a difference, and less borrowing is needed.

The RB provides competitive TPR than the no-borrowing method (IND), and substantially better TPR over the MAP and Full borrowing methods (Figure **5**). When $HpCR > p0$, more borrowing obtains higher TPR, e.g., MAP and RB; When $HpCR < p0$, less borrowing maintains a good TPR, e.g., IND and RB. In both situations, RB performs as good as the best methods without considering the full borrowing.

In summary, MAP tends to borrow more from the historical data and the performance in terms of TPR depends on the historical data and the null hypothesis rate. It performs better than the IND analysis by borrowing a good historical result but also suffers from issue of borrowing a poor result. On the other hand, RB appears to recognize when there is a difference, and less borrowing is used. RB improves the TPR over both IND and MAP, at about 12% to 25%.

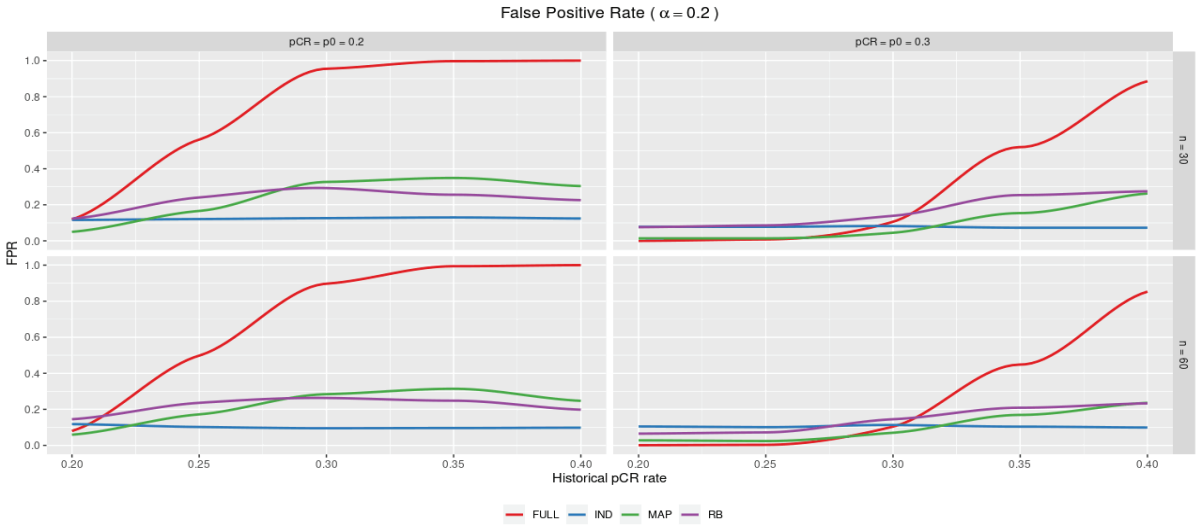


Figure 4: False Positive Rate (FPR) of methods for different historical pCR rates (***HpCR***) and samples sizes (***n***) for two cutoffs ($\boldsymbol{p_0}$).

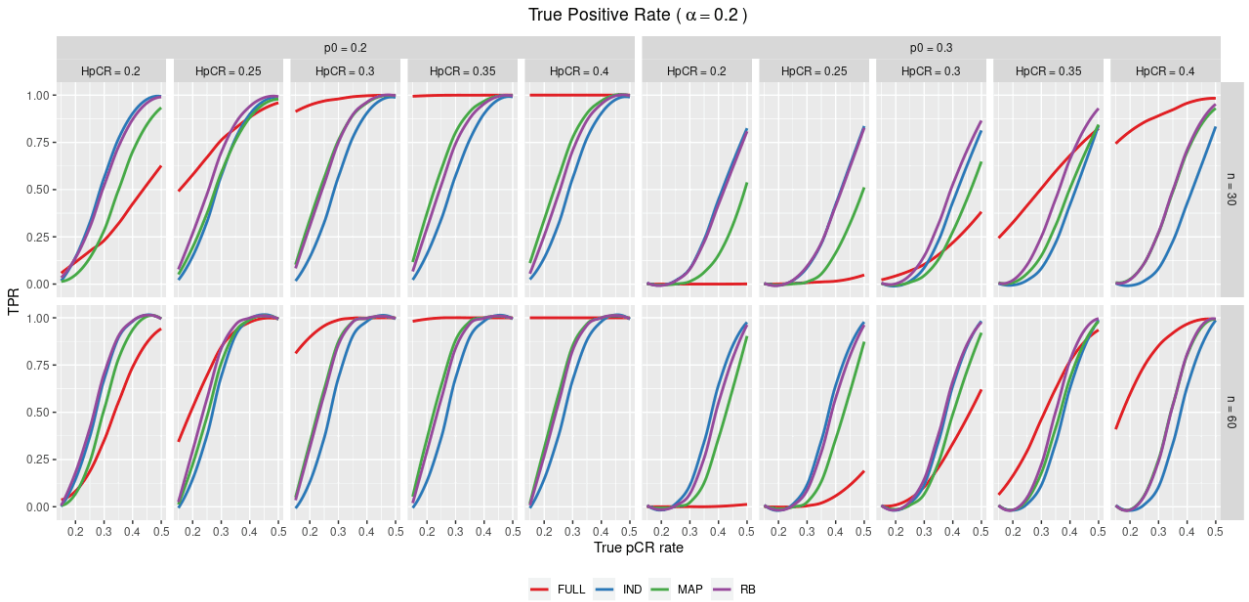


Figure 5: True Positive Rate (TPR) for different historical pCR rates (***HpCR***), sample sizes (***n***), and cutoffs ($\boldsymbol{p_0}$) against different true pCR.

7. REFERENCES

Brant A, Kates M, Chappidi MR, et al. Pathologic response in patients receiving neoadjuvant chemotherapy for muscle-invasive bladder cancer: Is therapeutic effect owing to chemotherapy or TURBT? *Urol Oncol Semin Orig Investig*. 2017;35(1):34.e17-34.e25. doi:10.1016/j.urolonc.2016.08.005

Degboe A, Ivanescu C, Rohay JM, Turner RR, Cella D. Validity and performance of the Functional Assessment of Cancer Therapy-Bladder (FACT-Bl) among advanced urothelial cancer patients. Support Care Cancer. 2019 Nov;27(11):4189-4198.

Efron, Bradley and RJT. *An Introduction to the Bootstrap*.; 1993. doi:https://doi.org/10.1201/97804292465938.

Grossman, H. Barton et al. Neoadjuvant Chemotherapy Plus Cystectomy Compared With Cystectomy Alone for Locally Advanced Bladder Cancer. *N Engl J Med*. Published online 2003:141-146. doi:10.1093/med/9780190655341.003.0025.

CCI [redacted]

Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clin Trials*. 2010;7(1):5-18. doi:10.1177/1740774509356002

Oken MM, Creech RH, Davis TE. Toxicology and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol Cancer Clin Trials*. 1982;5(6):649-655. doi:10.1097/00000421-198212000-00014

CCI [redacted]