ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

SHORT COMMUNICATION

Indirect comparison of reconstructed individual patient data from KEYNOTE-905/EV-303 and PURE-01: complementary analyses for interpreting event-free survival in muscle-invasive bladder cancer

B. A. Maiorano[1*] **& A. Necchi**[1,2]

[1]Department of Medical Oncology, Comprehensive Cancer Center, IRCCS San Raffaele, Milan; [2]Vita-Salute San Raffaele University, Milan, Italy

Available online 22 September 2026

Background: Neoadjuvant and perioperative immunotherapy have transformed the treatment landscape of muscle-invasive bladder cancer, with multiple strategies now available. In the absence of randomized head-to-head evidence, cross-trial comparisons are increasingly used to contextualize outcomes, but they may be affected by differences in study populations, treatment strategies, endpoint definitions, and follow-up maturity. We compared event-free survival (EFS) using reconstructed individual patient data from KEYNOTE-905/EV-303 and PURE-01.

Patients and methods: Individual patient data were reconstructed from published Kaplan—Meier curves. EFS was compared using Kaplan—Meier estimates, Cox regression, and log-rank tests. Sensitivity analyses included proportional hazards assessment, restricted mean survival time, and common follow-up analyses at 12, 24, and 36 months.

Results: The analysis included 325 patients (170 KEYNOTE-905/EV-303; 155 PURE-01) and 93 EFS events. EFS did not differ significantly between perioperative enfortumab vedotin—pembrolizumab and neoadjuvant pembrolizumab (hazard ratio 1.33, 95% confidence interval 0.87-2.03; $P = 0.181$), with evidence of nonproportional hazards ($P = 0.021$). Restricted mean survival time and common follow-up analyses yielded consistent results.

Conclusions: This reconstructed indirect comparison does not permit inference on comparative efficacy or the contribution of individual therapeutic components. Differences in study populations, treatment strategies, and follow-up maturity warrant caution. Complementary survival analyses may help generate hypotheses for studies designed to disentangle individual agents and perioperative treatment phases.

Key words: enfortumab vedotin, event-free survival, muscle-invasive bladder cancer, pembrolizumab, reconstructed individual patient data

INTRODUCTION

Neoadjuvant and perioperative immunotherapy have transformed the treatment landscape of muscle-invasive bladder cancer (MIBC), substantially improving pathologically documented response (pathological response) rates and long-term clinical outcomes for patients eligible for radical cystectomy (RC), starting from the cisplatin-ineligible population.[1] The phase II PURE-01 trial established neoadjuvant pembrolizumab as the first prospective proof of concept that immune checkpoint inhibitor (ICI) monotherapy could induce clinically meaningful pathological responses and durable long-term outcomes.[2] More recently, combining ICIs with different therapeutic agents, such as chemotherapy and antibody-drug conjugates, has further improved survival and response in the perioperative setting, establishing a new standard of care in this setting. The phase III KEYNOTE-905/EV-303 trial demonstrated a significant improvement in event-free survival (EFS) with perioperative enfortumab vedotin (EV) plus pembrolizumab compared with RC alone. Notably, the original study design also included a perioperative pembrolizumab monotherapy arm, which was later discontinued, underscoring the clinical interest in defining the incremental contribution of combination therapy.[3,4] Despite these advances, important clinical questions remain regarding the relative contribution of the different therapeutic components across the perioperative continuum, including neoadjuvant versus adjuvant treatment and ICI monotherapy versus combination approaches.

*Correspondence to: Dr Brigida Anna Maiorano, Department of Medical Oncology, Comprehensive Cancer Center, IRCCS San Raffaele, via Olgettina, 60 - 20132 - Milan, Italy. Tel. +39-0226431

E-mail: maiorano.brigida@hsr.it (B. A. Maiorano).

2059-7029/© 2026 The Authors. Published by Elsevier Ltd on behalf of European Society for Medical Oncology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

As multiple effective perioperative strategies become available, indirect comparisons between contemporary clinical trials are increasingly carried out to contextualize efficacy in the absence of randomized head-to-head evidence. However, differences in study populations, endpoint definitions, treatment strategies, and follow-up maturity may substantially influence the interpretation of reconstructed survival outcomes. To explore these issues, we reconstructed individual patient data from published Kaplan–Meier curves of the experimental arm of KEYNOTE-905/EV-303 and the final PURE-01 cohort. We then carried out an exploratory indirect comparison of EFS, complemented by sensitivity analyses designed to account for differences in follow-up maturity.

PATIENTS AND METHODS

Reconstructed individual patient data (rIPD) were generated from published Kaplan–Meier curves of the phase III KEYNOTE-905/EV-303 trial (EV plus pembrolizumab arm) and the phase II PURE-01 study using the algorithm described by Guyot et al[5] (detailed in the Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2026.108597). Published Kaplan–Meier curves were digitally extracted, and rIPD were reconstructed using the numbers at risk and reported survival probabilities according to the Guyot algorithm.[5]

EFS was selected as the common endpoint because sufficiently comparable definitions were available in both studies, although minor differences in event ascertainment were acknowledged. Survival distributions were estimated using the Kaplan–Meier method and compared using the log-rank test. Hazard ratios (HRs) and corresponding 95% confidence intervals (CIs) were estimated using Cox proportional hazards regression. Given the substantial difference in median follow-up between the two studies, the proportional hazards assumption was assessed using Schoenfeld residuals. Sensitivity analyses included restricted mean survival time (RMST) analyses truncated at 12, 24, and 36 months, together with analyses censoring follow-up at the same time points to explore the potential influence of differential follow-up duration on the primary findings. These time horizons provide clinically interpretable early-, intermediate-, and longer-term assessments of EFS within common follow-up windows across both cohorts, enabling evaluation of the robustness of the comparison while limiting the influence of unequal follow-up maturity. The 24-month horizon also approximately corresponds to the median follow-up of KEYNOTE-905/EV-303. As an additional exploratory sensitivity analysis, rIPD from the randomized control arm of KEYNOTE-905/EV-303 were compared with the PURE-01 cohort using the same survival analysis framework.

All statistical analyses were carried out using R version 4.4.2 (R Foundation for Statistical Computing, Vienna, Austria). All tests were two-sided, and P values <0.05 were considered statistically significant.

RESULTS

The rIPD comprised 170 patients from the experimental arm of KEYNOTE-905/EV-303 and 155 patients from the final PURE-01 cohort. Baseline study characteristics, treatment strategies, and endpoint definitions are summarized in Table 1.

The rIPD analysis included 93 EFS events (47 in KEYNOTE-905/EV-303; 46 in PURE-01). The reconstructed comparison showed no statistically significant difference in EFS

Table 1. Baseline characteristics and key design features of the KEYNOTE-905/EV-303 and PURE-01 trials

Characteristic	KEYNOTE-905/EV-303	PURE-01
Study design	Phase III, randomized, open-label, multicenter	Phase II, single-arm, open-label
Patients analyzed, *N*	170	155
Clinical stage, *n* (%)		
cT2	30 (17.6)	75 (48)
cT3-4	133 (78.2)	80 (52)
cN+	7 (4.1)	0
Median age, years, (IQR)	74 (47-87)	68 (62-74)
Female sex, *n* (%)	33 (19.4)	20 (13)
Histology, *n* (%)		
Pure UC	152 (89.4)	111 (72)
Variant histology	18 (10.6)	44 (28)
Neoadjuvant treatment	EV 1.25 mg/kg (days 1 and 8) + pembrolizumab 200 mg (day 1) q3w × 3 cycles	Pembrolizumab 200 mg q3w × 3 cycles
RC carried out, *n* (%)	149 (87.6)	147 (94.8)
Postoperative treatment	EV–pembrolizumab	None (protocol-mandated)
Primary endpoint	EFS	pCR
EFS definition	Progression precluding surgery, biopsy-proven residual MIBC in nonoperated patients, inability to complete curative-intent surgery because of unresectable or metastatic disease, postsurgical recurrence, or death	Disease progression, inability to undergo RC, initiation of additional neoadjuvant chemotherapy, or death
Median FU, months (IQR)	25.6 (11.8-53.7)	61.8 (53.6-68.2)
Published EFS	2-year EFS: 74.7% (95% CI 66.9% to 80.8%)	5-year EFS: 68% (95% CI 61% to 76%)
pCR (ypT0N0), *n* (%)	97 (57.1)	57 (37)

CI, confidence interval; EFS, event-free survival; EV, enfortumab vedotin; FU, follow-up; IQR, interquartile range; pCR, pathologically documented complete response; MIBC, muscle-invasive bladder cancer; q3w, every 3 weeks; RC, radical cystectomy; UC, urothelial carcinoma.

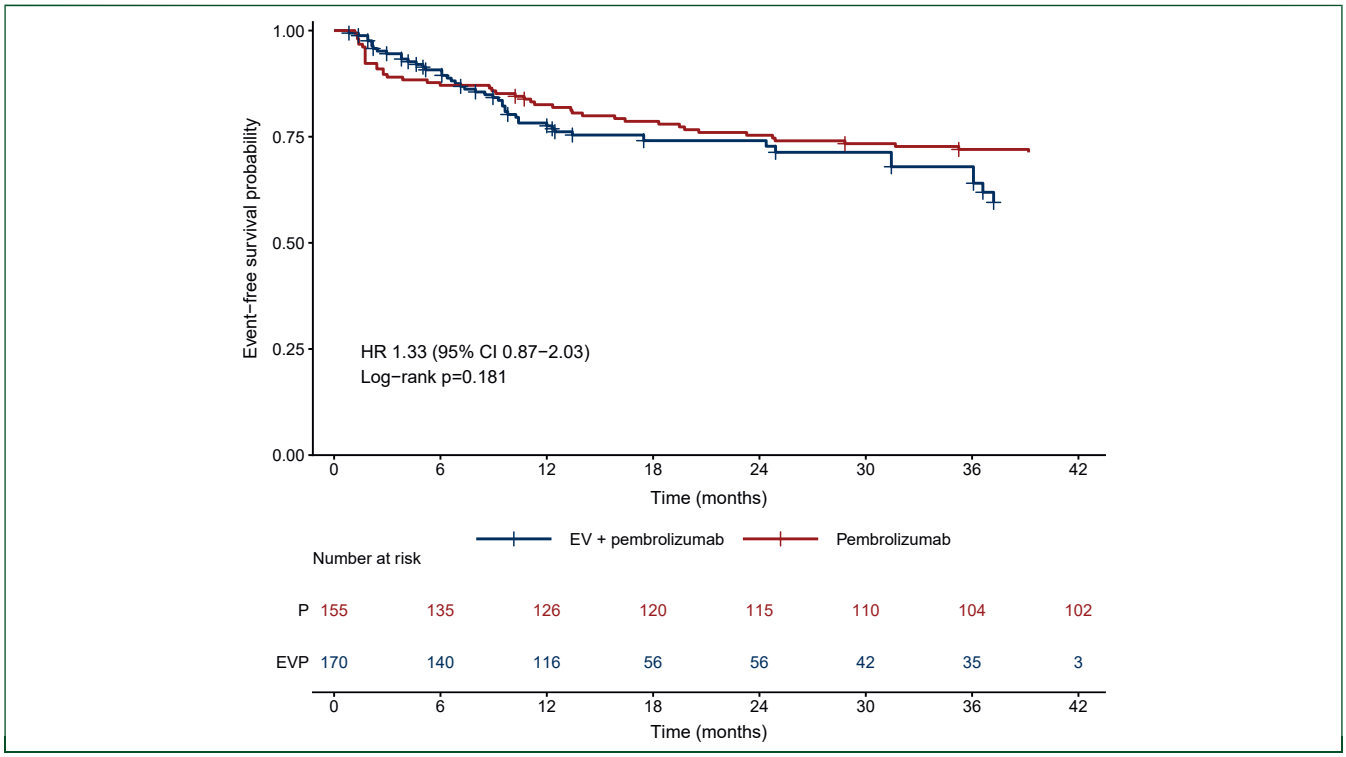


Figure 1. Reconstructed event-free survival curves for the experimental arm of KEYNOTE-905/EV-303 and the final PURE-01 cohort. Kaplan—Meier estimates based on reconstructed individual patient data derived from published survival curves, comparing perioperative EV plus pembrolizumab with neoadjuvant pembrolizumab. CI, confidence interval; EVP, enfortumab vedotin—pembrolizumab; HR, hazard ratio; P, pembrolizumab.

between treatment strategies (HR 1.33, 95% CI 0.87-2.03, $P = 0.181$; Figure 1).

Assessment of the proportional hazards assumption indicated evidence of nonproportional hazards (Schoenfeld test, $P = 0.021$; Supplementary Figure S1, available at https://doi.org/10.1016/j.esmoop.2026.108597). RMST analyses demonstrated no statistically significant differences between treatment strategies at 12, 24, or 36 months. At 12 months, RMST was 10.71 with EV plus pembrolizumab and 10.67 with pembrolizumab (mean difference 0.06 months, 95% CI −0.62 to 0.73, $P = 0.868$), with similarly nonsignificant differences at 24 and 36 months (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2026.108597). Similarly, sensitivity analyses using common follow-up durations yielded consistent results at 12 months (HR 1.26, 95% CI 0.76-2.08, $P = 0.372$), as well as at 24 and 36 months (Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2026.108597). In an additional exploratory analysis, EFS was significantly longer in the final PURE-01 cohort than in the reconstructed KEYNOTE-905/EV-303 randomized control arm (HR for control compared with PURE-01 3.14, 95% CI 2.17-4.55, $P < 0.001$; Supplementary Figure S2, available at https://doi.org/10.1016/j.esmoop.2026.108597).

DISCUSSION

In this reconstructed comparison of individual patient data from the experimental arm of KEYNOTE-905/EV-303 and the final PURE-01 cohort, the primary EFS analysis showed no statistically significant difference between the two approaches. However, the marked discrepancy in follow-up duration between the two studies and the observed violation of the proportional hazards assumption suggested that conventional HR estimates should be interpreted with caution. Complementary sensitivity analyses, including RMST and common follow-up analyses, provided additional context for interpreting the reconstructed survival data. These observations underscore the importance of considering follow-up maturity and time-varying treatment effects when interpreting indirect comparisons between contemporary perioperative trials. Indeed, the apparent divergence observed in the primary reconstructed comparison should be interpreted in the context of important differences between the two studies. Most notably, median follow-up was substantially longer in PURE-01 than in KEYNOTE-905/EV-303 (61.8 compared with 25.6 months), reflecting different stages of outcome maturity. Consistent with this, the proportional hazards assumption was not met in the primary analysis, suggesting that a single HR may not fully capture the relationship between the reconstructed survival curves over time. For this reason, we complemented the primary analysis with RMST and common follow-up analyses, which are less dependent on the proportional hazards assumption and facilitate interpretation in the presence of unequal follow-up. These analyses yielded results that were broadly consistent with the primary comparison. The crossing of the reconstructed EFS curves at approximately 6 months further illustrates the time-varying nature of the observed association. Several factors may have contributed to this pattern. The higher

pathological complete response (pCR) rate reported with perioperative EV plus pembrolizumab (57.1% compared with 37% with pembrolizumab) may have contributed to the numerically favorable early EFS pattern. Although pCR is prognostically informative at the individual-patient level, recent trial-level surrogacy analyses have shown that treatment-induced differences in pCR do not reliably predict corresponding differences in long-term survival outcomes.[6] Differences in treatment sequencing and duration, EFS ascertainment, and the underlying study populations may also have influenced the timing of events. Accordingly, the observed crossing should not be attributed to any single clinical mechanism and further supports caution in summarizing these curves using a constant HR. In addition to follow-up maturity, the studies differed in several clinically relevant characteristics. KEYNOTE-905/EV-303 enrolled a higher proportion of patients with cT3-4 disease, included clinically node-positive tumors, and had a lower prevalence of variant histology compared with PURE-01, whereas PURE-01 included a greater proportion of patients with cT2 disease and excluded cN+ disease. This marked imbalance in baseline clinical stage represents an important source of prognostic heterogeneity and may have influenced the observed EFS patterns. Because rIPD from KEYNOTE-905/EV-303 do not retain individual-level baseline covariates, valid stage-adjusted modeling was not possible without making unverifiable assumptions regarding the relationship between clinical stage and individual event times. Furthermore, although EFS was selected as the common endpoint, minor differences in endpoint definitions remained between studies (Table 1). Collectively, these sources of clinical and methodological heterogeneity should be considered when interpreting the present indirect comparison.

Our findings should not be interpreted as evidence for or against the superiority of either treatment strategy. Rather, they illustrate the inherent challenges of indirectly comparing perioperative trials conducted in different clinical settings and at different stages of follow-up maturity.[7,8] As novel perioperative regimens continue to emerge, cross-trial comparisons are increasingly carried out despite the absence of randomized head-to-head evidence. In this context, reliance on a single summary measure such as the HR may provide an incomplete assessment when proportional hazards assumptions are not met or follow-up differs substantially between studies. Our findings suggest that complementary approaches, including RMST and common follow-up analyses, may facilitate interpretation when reconstructed survival data are compared across studies with unequal follow-up.

Beyond methodological considerations, these findings also highlight an unresolved clinical question in the evolving perioperative treatment landscape: the relative contribution of individual therapeutic agents and treatment phases within increasingly complex regimens. Contemporary phase III studies establish the efficacy of overall perioperative strategies but are not designed to disentangle the contribution of individual agents or to determine how much of the observed benefit derives from neoadjuvant therapy versus postoperative treatment.[9] Consequently, the relative contribution of treatment duration and individual therapeutic components, as well as the incremental benefit of combining ICIs with chemotherapy or antibody-drug conjugates, remains uncertain. This question extends beyond efficacy alone. Combination perioperative strategies inevitably increase treatment complexity, toxicity, and resource utilization, making it increasingly important to identify which patients truly require treatment intensification and those who may achieve durable outcomes with less intensive approaches. Surgical feasibility should also be considered when evaluating treatment intensity. Although most patients in both studies ultimately underwent RC (Table 1), treatment exposure and toxicity differed substantially between the two strategies. Differences in treatment duration, study populations, and adverse event reporting therefore preclude meaningful direct comparisons of safety. Indeed, these considerations have direct implications for patient selection, and both overtreatment and undertreatment remain relevant concerns in contemporary perioperative management. Despite extensive biomarker research, no predictive biomarker has yet been clinically validated to guide perioperative treatment selection in MIBC. Notably, the prospective randomized phase II DUTRENEO study recently showed that an 18-gene tumor inflammation signature, despite prior retrospective validation, failed to enrich for response to neoadjuvant ICIs, further illustrating the challenges of translating candidate biomarkers into clinically actionable treatment selection tools.[10] The development and prospective validation of biomarkers capable of refining perioperative treatment selection—including circulating and urinary tumor DNA—will therefore be essential to optimize treatment sequencing and intensity beyond conventional clinicopathological risk factors.[11,12] Similarly, the expansion of bladder preservation strategies will further increase the need for reliable predictive biomarkers capable of identifying patients who may safely avoid RC without compromising oncological outcomes.[13] An additional unanswered question concerns the independent contribution of perioperative pembrolizumab. Although KEYNOTE-905/EV-303 initially included a perioperative pembrolizumab monotherapy arm, the EV plus pembrolizumab comparison was subsequently prioritized, leaving the incremental contribution of EV beyond checkpoint inhibition incompletely defined.[4] This evidence gap is becoming increasingly relevant as additional multicomponent perioperative strategies emerge. Most recently, positive high-level results from the phase III VOLGA trial were announced, indicating relevant improvements in EFS and OS with perioperative durvalumab plus neoadjuvant EV versus surgery with or without approved adjuvant therapy.[14] However, as these results have not yet been presented or published in full, detailed interpretation is premature. More broadly, these evolving data reinforce the need for future perioperative studies to move beyond evaluating entire treatment strategies and

instead seek to disentangle the contribution of individual therapeutic components and treatment phases, particularly as biomarker-driven treatment escalation and deescalation become feasible.

This study has several strengths. To our knowledge, it represents the first rIPD comparison between perioperative EV plus pembrolizumab and neoadjuvant pembrolizumab in MIBC. Beyond conventional survival analyses, we incorporated the evaluation of the proportional hazards assumption, RMST, and common follow-up analyses, providing a more comprehensive assessment of reconstructed survival data. Compared with the mere visual inspection of published Kaplan—Meier curves, it enables estimation of time-to-event statistics and complementary analyses. Our analytical approach is also supported by the growing methodological literature on rIPD. The Guyot algorithm has become the reference method for reconstructing survival data from published Kaplan—Meier curves and has been widely applied in oncology for indirect comparisons and evidence synthesis. Furthermore, RMST derived from rIPD has been shown to closely reproduce estimates obtained from original trial datasets, supporting its use as a complementary measure when the proportional hazards assumption is violated.[5,7,8] Therefore, the consistency between our RMST analyses and common follow-up sensitivity analyses strengthens the robustness of the present findings despite the inherent limitations of indirect comparisons. However, these findings should also be considered in light of several limitations. Firstly, this was an indirect comparison based on reconstructed rather than original patient-level data. Secondly, differences in eligibility criteria, baseline disease characteristics, treatment strategies, and EFS definitions limit direct comparability between the two studies. In particular, the marked imbalance in baseline clinical stage could not be accounted for because individual-level baseline covariates were unavailable for the reconstructed KEYNOTE-905/EV-303 cohort. Finally, rIPD analyses cannot fully account for unmeasured patient-level confounders or replace randomized head-to-head comparisons. Accordingly, these findings should be interpreted as exploratory and hypothesis-generating.

CONCLUSIONS

Indirect comparisons between contemporary perioperative bladder cancer trials should be interpreted in the context of differences in study design, populations, treatment strategies, endpoint definitions, and follow-up maturity. The present reconstructed comparison does not permit inference on comparative treatment efficacy or on the incremental contribution of individual therapeutic components. Rather, these findings highlight the limitations of comparing increasingly complex perioperative strategies as indivisible treatment packages and reinforce the need for future studies to define the contribution of individual agents and treatment phases. Integrating complementary survival analyses and biomarker-driven treatment selection may ultimately help inform more individualized escalation and deescalation strategies.

FUNDING

None declared.

DISCLOSURE

BAM reports consulting or advisory and speaking and lecture fees with Merck & Co Inc and Astellas Pharma Inc. AN has declared no conflicts of interest.

REFERENCES

1. Cigliola A, Maiorano BA, Dengur D, et al. The management of muscle invasive bladder cancer: state of the art and future perspectives. *Cancers (Basel)*. 2025;17(24):4017.
2. Necchi A, Anichini A, Raggi D, et al. Pembrolizumab as neoadjuvant therapy before radical cystectomy in patients with muscle-invasive urothelial bladder carcinoma (PURE-01): an open-label, single-arm, phase II study. *J Clin Oncol*. 2018;36(34):3353-3360.
3. Tateo V, Basile G, Giannatempo P, et al. Updated 5-year survival results from PURE-01, a phase 2 study of neoadjuvant pembrolizumab followed by radical cystectomy in patients with muscle-invasive bladder cancer. *Eur Urol*. 2026;90(1):7-12.
4. Vulsteke C, Adra N, Danchaivijitr P, et al. Perioperative enfortumab vedotin and pembrolizumab in bladder cancer. *N Engl J Med*. 2026;394(13):1257-1269.
5. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12(1):9.
6. Scilipoti P, Ślusarczyk A, Thibault C, et al. Evaluating the surrogacy of pathological complete response at radical cystectomy in neoadjuvant-based muscle-invasive bladder cancer trials. *Eur Urol Oncol*. In press, https://doi.org/10.1016/j.euo.2026.06.006; 2026.
7. Maiorano BA, Dizman N, Necchi A. Contextualizing the eevelopment of anti-HER2 therapies for patients with advanced urothelial carcinoma. *Eur Urol*. 2026;89(6):479-482.
8. Kamat AM, Hensley PJ, Maiorano BA, et al. Bacillus Calmette-Guérin (BCG) and beyond: is systemic immunotherapy for BCG-naïve non—muscle-invasive bladder cancer progress or overreach? *J Clin Oncol*. 2026;44(16):1478-1482.
9. Powles T, Catto JWF, Galsky MD, et al. Perioperative durvalumab with neoadjuvant chemotherapy in operable bladder cancer. *N Engl J Med*. 2024;391(19):1773-1786.
10. Grande E, Sibai M, Grases D, et al. Spatial architecture contributes to failure of bulk biomarker-guided neoadjuvant immunotherapy selection in bladder cancer: the DUTRENEO study. *Cell Rep Med*. 2026;7(7):102878.
11. Powles T, Assaf ZJ, Davarpanah N, et al. ctDNA guiding adjuvant immunotherapy in urothelial carcinoma. *Nature*. 2021;595(7867):432-437.
12. Powles T, Van Der Heijden MS, Wang Y, et al. Circulating tumor DNA (ctDNA) in patients with muscle-invasive bladder cancer (MIBC) who received perioperative durvalumab (D) in NIAGARA. *J Clin Oncol*. 2025;43(16 suppl):4503.
13. Necchi A, Galsky MD, Dizman N, et al. End points for the next-generation bladder-sparing perioperative trials for patients with muscle-invasive bladder cancer. *J Clin Oncol*. 2025;43(32):3536-3544.
14. AstraZeneca. Perioperative Imfinzi plus neoadjuvant EV showed statistically significant and clinically meaningful improvements in event-free survival and overall survival in muscle-invasive bladder cancer in the Phase III VOLGA trial. Available at https://www.astrazeneca.com/media-centre/press-releases/2026/imfinzi-ev-improves-efs-os-in-bladder-cancer.html. Accessed September 5, 2026.