Journal for ImmunoTherapy of Cancer

Frontline immunotherapy with response-guided subsequent treatment (FIRST) in cutaneous squamous cell carcinoma: a Bayesian causal analysis of dose intensity, early benefit, and treatment de-escalation

David M Miller ,[1,2,3,4] Ross D Merkin,[1,2,3] Howard L Kaufman ,[1,3,5] Sameer G Gupta,[3,6] Natalie Wolkow,[3,7] Adewunmi Adelaja,[3,8] Ryan J Sullivan ,[1,2,3] Chirayu Patel,[1,3,9] Alexandra Sorrentino,[1] Sonia Cohen,[1,3,10] Christine Cimoch,[1,5] Kevin S Emerick[1,3,5]

To cite: Miller DM, Merkin RD, Kaufman HL, *et al*. Frontline immunotherapy with response-guided subsequent treatment (FIRST) in cutaneous squamous cell carcinoma: a Bayesian causal analysis of dose intensity, early benefit, and treatment de-escalation. *Journal for ImmunoTherapy of Cancer* 2026;**14**:e015029. doi:10.1136/jitc-2026-015029

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jitc-2026-015029).

Presented in part at the Society of Cutaneous Oncology Research Meeting, March 9, 2026.

Accepted 03 July 2026

For numbered affiliations see end of article.

Correspondence to
Dr David M Miller;
dmiller4@mgh.harvard.edu

ABSTRACT

Background Programmed death receptor 1 (PD-1) blockade produces high response rates in resectable and locally advanced cutaneous squamous cell carcinoma (CSCC), but how many doses are needed and whether surgery or radiation after immunotherapy (consolidation) adds benefit in deep responders (patients with substantial clinical responses) remain unclear. Observational inference is challenging because treatment decisions are response-guided, doses accumulate over time, and treatment selection depends on patient characteristics that also affect outcomes.

Methods We conducted a retrospective cohort study of 189 patients with resectable, borderline-resectable, locally advanced, or limited metastatic CSCC treated with immune checkpoint inhibitors as part of management (2019–2025). We summarized responses, treatment discontinuation patterns, and time-to-event outcomes. To evaluate dose-response relationships, we fit Bayesian regression models adjusting for prespecified baseline confounders. Directed acyclic graphs were used to formalize causal assumptions and identify minimally sufficient adjustment sets. For event-free survival (EFS), we used a prespecified landmark analysis among patients receiving ≥2 doses to reduce guarantee-time bias.

Results Objective response occurred in 119/189 (63%), including 52/189 (27.5%) complete responses. Nearly half of patients received ≤2 doses (92/189, 48.7%). Among 50 complete responders managed without surgery, 1 recurrence was observed over a median follow-up of 25.4 months. Dose-response models showed a consistent but modest positive association between additional doses and response: in a linear model, each additional dose was associated with ~1.09 fold higher odds of response (posterior probability >95%). Flexible threshold models suggested early concentration of benefit, strongest at 2 versus 1 dose (93.6% posterior probability of benefit), with persistent uncertainty in effect magnitude (66% probability of ≥6 percentage-point absolute increase). In the EFS landmark modeling cohort (n=177), receipt of ≥3 versus 2 doses showed a 93% posterior probability of reduced event risk, but the 89% credible interval spanned the null (HR 0.61–1.01), indicating substantial uncertainty regarding incremental downstream benefit.

Conclusions In real-world CSCC, durable disease control frequently occurred after limited immunotherapy exposure, and clinical responders often did well without routine surgical consolidation. Although additional doses may modestly improve outcomes, observed gains appear concentrated early with uncertain incremental value beyond two doses. These patterns support conceptualizing treatment as frontline immunotherapy with response-guided subsequent treatment rather than fixed-duration neoadjuvant therapy.

INTRODUCTION

Cutaneous squamous cell carcinoma (CSCC) is the second most common skin cancer worldwide and is usually cured with local therapy.[1] However, a subset of patients present with resectable but high-risk, locally or regionally advanced disease for which surgery—often combined with adjuvant radiotherapy—can result in substantial functional morbidity and remains associated with meaningful recurrence risk. These limitations have driven growing interest in systemic approaches that improve oncological outcomes while enabling treatment de-escalation.

Immune checkpoint inhibitors (ICIs), particularly Programmed death receptor 1 (PD-1) blockade, have demonstrated robust

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Programmed death receptor 1 (PD-1) blockade produces high response rates in resectable and locally advanced cutaneous squamous cell carcinoma, and short neoadjuvant courses can achieve substantial pathological responses. However, the minimal effective dose intensity, the marginal benefit of additional immunotherapy cycles, and the incremental value of surgery or radiation in clinical responders remain undefined, with current practice largely driven by convention rather than comparative evidence.

WHAT THIS STUDY ADDS

⇒ In a large real-world cohort, durable disease control frequently occurred after limited immunotherapy exposure, with suggestive but inconclusive evidence for benefit between one and two doses, and substantial uncertainty regarding optimal minimal dose and incremental gains from prolonged treatment. Deep responders often achieved sustained remission without surgical consolidation, and additional doses were associated with, at most, modest and uncertain improvements in downstream outcomes.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

⇒ These findings support a frontline immunotherapy with response-guided subsequent treatment paradigm, emphasizing response-adapted, less intensive strategies for selected patients and raising the possibility that routine escalation of systemic or local therapy may not be necessary in deep responders. Future research should prioritize identifying patients unlikely to benefit from PD-1 blockade who may require early integration of surgery or radiation, rather than intensifying treatment uniformly across all patients.

activity in advanced CSCC. Early neoadjuvant studies reported high rates of pathological response after short courses of therapy: two doses of cemiplimab produced major or complete pathological responses in 75% of patients,[2] with similar results observed in a multicenter trial where up to four doses achieved 51% pathological complete response.[3] Subsequent trials have extended these observations: the MATISSE trial demonstrated similar efficacy with two-dose nivolumab regimens and supported organ-preservation strategies in selected deep responders (patients with substantial clinical responses) who declined surgery,[4] while the De-Squamate study reported 63% clinical or pathological complete response using response-adapted de-escalation.[5] In the adjuvant setting, the phase III C-POST trial established that cemiplimab reduces recurrence risk after surgery and radiotherapy (HR 0.32),[6] reinforcing immunotherapy's role across the disease continuum.

Despite these advances, fundamental questions remain unresolved. The optimal intensity and duration of neoadjuvant immunotherapy have not been defined, with patients across trials receiving between two and four doses[2–4] and limited evidence regarding the marginal benefit of additional cycles. It remains unclear whether additional immunotherapy cycles meaningfully deepen benefit or merely extend exposure in patients already responding. Similarly, the necessity of consolidative surgery (surgical resection following immunotherapy response) in patients achieving deep clinical responses remains uncertain, with current practice driven largely by convention rather than comparative data. Addressing these questions using observational data is challenging because treatment decisions are response-guided, doses accumulate over time (patients must remain under treatment long enough to receive additional cycles), and treatment selection depends on patient characteristics that also affect outcomes.

To address these gaps, we conducted a retrospective cohort study of 189 patients with resectable, borderline-resectable, locally and regionally advanced CSCC treated with immunotherapy as part of initial management. Natural variation in clinical practice over 5 years—including heterogeneity in treatment duration and selective use of surgical consolidation—provided an opportunity to examine dose-response relationships and the role of surgery in responding patients, mirroring the response-guided strategies increasingly used in routine practice. Using directed acyclic graphs to identify minimally sufficient adjustment sets and Bayesian regression to quantify uncertainty, we sought to clarify the relationship between immunotherapy dose intensity and clinical outcomes—particularly whether benefit accrues primarily with early doses—and to contextualize the role of surgery in the era of highly effective neoadjuvant immunotherapy.

MATERIALS AND METHODS

Study design and oversight

We conducted an Institutional Review Board-approved retrospective cohort study at the Mass General Brigham Cancer Institute. Analytic definitions and modeling plans were developed a priori. Additional methodological detail is provided in the online supplemental methods.

Data sources and cohort assembly

Patients were identified through institutional data repositories and medical record review. The cohort included patients with resectable locally/regionally advanced CSCC treated with immune checkpoint inhibition as part of initial management. Patients with prior systemic antineoplastic therapy for CSCC and those with widely metastatic disease (>5 lesions) at presentation were excluded. The analytic dataset was locked on 23 December 2025.

Treatment intent and clinical variables

Treatment intent at initiation was classified from contemporaneous clinician documentation as neoadjuvant or frontline immunotherapy with response-guided subsequent treatment (FIRST), with actual surgical resection recorded separately from intended strategy. Baseline variables included key demographic, clinical, treatment, and disease characteristics as documented by the treating clinician and supported by available imaging and pathology. Patients without an identifiable primary tumor were designated as unstaged. Immunosuppression

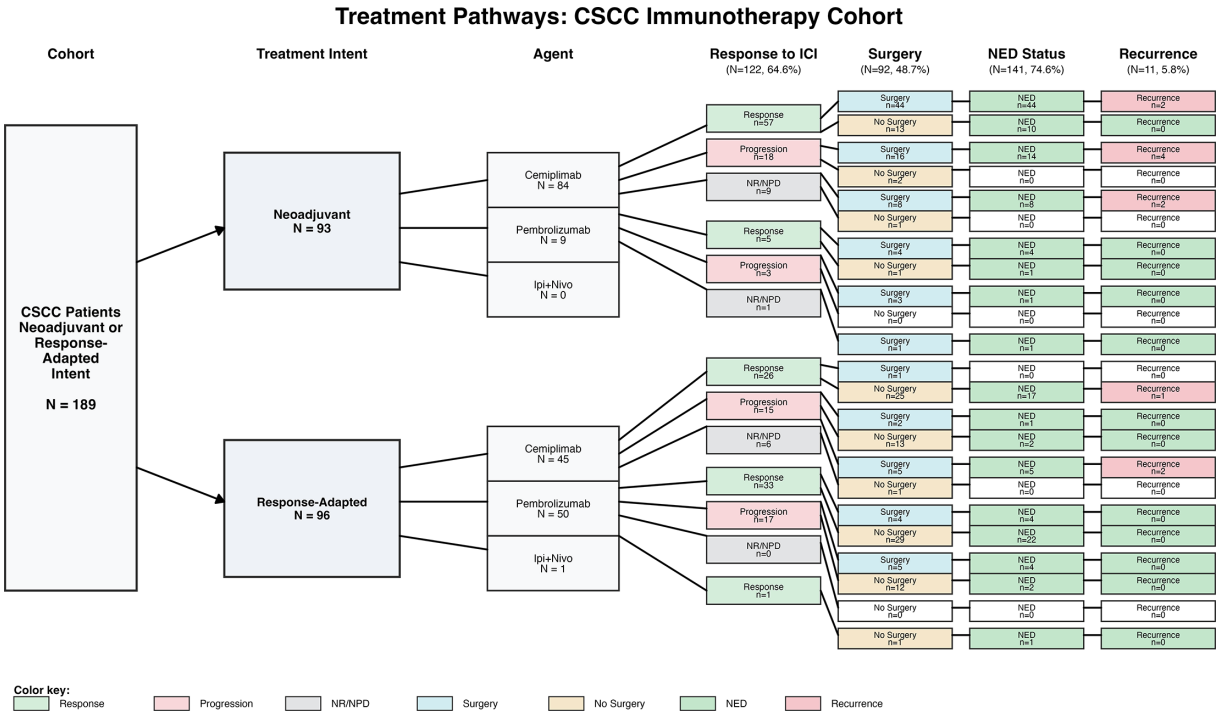

Figure 1 Treatment pathways and outcomes for patients with CSCC treated with immune checkpoint inhibitors. This horizontal flow diagram depicts treatment trajectories for 189 patients with cutaneous squamous cell carcinoma treated with neoadjuvant or response-adapted immunotherapy. Patients are stratified by treatment intent (neoadjuvant vs response-adapted), therapeutic agent (cemiplimab, pembrolizumab, or ipilimumab plus nivolumab), response to immune checkpoint inhibitors (responder, progressed, or non-responder/non-progressed), surgical intervention, achievement of no evidence of disease (NED) status, and subsequent recurrence. Each box displays the number of patients at that stage of the treatment pathway. Overall cohort outcomes are shown beneath column headers as counts with percentages: responders, patients undergoing surgery, patients achieving NED, and patients experiencing recurrence. Colored boxes indicate clinically meaningful states: green denotes favorable outcomes (response or NED), red denotes progression or recurrence, blue indicates surgery, and tan indicates no surgery. For recurrence, green boxes represent patients with no recurrence following NED, whereas red boxes indicate observed recurrence; boxes remain unshaded when no patients are present in that category. The diagram illustrates heterogeneity in treatment pathways and outcomes, highlighting the relationships between immunotherapy response, surgical management, achievement of disease-free status, and recurrence across different treatment strategies. CSCC, cutaneous squamous cell carcinoma; ICI, immune checkpoint inhibitor; Ipi+Nivo, ipilimumab plus nivolumab; NR/NPD, non-responder/non-progressed.

was defined by hematological malignancy, immunosuppressive therapy, congenital immunodeficiency, or HIV infection.

Treatment strategy varied over the study period, with a transition from predominantly neoadjuvant management early in the cohort to a response-guided approach in later years, in which surgery and other local therapies were more selectively applied based on early treatment response. This temporal variation contributed to heterogeneity in treatment pathways. Because surgical decisions were influenced by post-baseline response, surgical pathway was not treated as a baseline proxy for disease extent in the primary causal framework.

Response assessment

Responses were assessed using a clinician-anchored framework reflecting real-world practice rather than formal RECIST or iRECIST criteria. Response dates corresponded to the earliest post-treatment clinician note, radiology report, or date of surgical resection when applicable. Clinical responses were categorized as complete response, partial response, stable disease, or progressive disease using standardized abstraction across clinical, radiographic, and pathological sources. Clinical complete response (cCR) was defined as the absence of clinically detectable disease on physical examination and available imaging, as documented by the treating clinician. In most cases, response assessment was based on cross-sectional imaging (CT and/or MRI), with adjunctive positron emission tomography (PET) imaging performed at clinician discretion rather than according to a standardized protocol. Routine mapping biopsies were not performed, and pathological confirmation of response was generally available only for patients who underwent surgical resection. For patients undergoing surgical resection, pathological response was abstracted as complete, major, partial, or no response. Because distinctions between partial and major response—particularly in the absence of standardized criteria—could not be applied consistently in retrospective clinical documentation, these categories were combined for analysis where appropriate.

Outcomes

Outcomes included response-based endpoints and time-to-event measures: event-free survival (EFS), recurrence-free survival (RFS), disease-specific survival (DSS), and overall survival (OS). Each endpoint was defined using prespecified time origins, event definitions, and censoring rules, with operational definitions provided in the online supplemental methods.

Statistical analysis

Descriptive statistics summarized baseline characteristics, treatment patterns, and outcomes. Unadjusted time-to-event outcomes were displayed using Kaplan-Meier methods for descriptive purposes only. Associations between immunotherapy exposure and outcomes were evaluated using Bayesian regression models with prespecified confounder adjustment. For response outcomes, Bayesian logistic regression models examined alternative dose-response specifications (linear, log-transformed, categorical).

For time-to-event outcomes, median follow-up was estimated using the reverse Kaplan-Meier method to appropriately account for censoring. Because ICI dose exposure accumulates over time, primary survival analyses used a prespecified landmark approach restricted to patients who remained event-free long enough to receive at least two doses, thereby reducing immortal time bias. Bayesian parametric survival models were then used to support stable estimation under limited event counts. Patients receiving fewer than two doses were retained in descriptive analyses to reflect real-world treatment patterns, including both early responders and patients who discontinued therapy due to progression, toxicity, or clinical decision-making.

Directed acyclic graphs were used to formalize causal assumptions and identify minimally sufficient adjustment sets, and sensitivity analyses evaluated robustness to alternative specifications. Full details regarding model formulation, priors, causal structures, diagnostics, and sensitivity analyses are provided in the online supplemental methods.

Uncertainty was characterized using posterior distributions. We report 89% credible intervals (CrI) as a descriptive summary, while inference was based on posterior probabilities rather than interval-based thresholds. The 89% level was chosen to reflect that interval width is a matter of convention and to distinguish credible intervals from frequentist confidence intervals.

Software, reproducibility, and artificial intelligence use

All analyses were conducted in R using a reproducible Quarto-based workflow. Bayesian response models were fit using ulam from the rethinking[7] framework, with prior specification, posterior prediction, and generative simulation implemented in a reproducible Quarto workflow. Bayesian survival models were fit using brms.[8] Computational details and selected analytic code are provided in the online supplemental methods and are available via an interactive project website https://themillerlab.github.io/FIRST/. Relevant methodological components of the analytic pipeline are available in a public GitHub repository https://github.com/TheMillerLab/FIRST.

Artificial intelligence-assisted tools were used to support multiple aspects of the analytic and writing process, including code development, model exploration, result visualization, and manuscript drafting. All analyses, interpretations, and final content decisions were performed and verified by the authors.

RESULTS

Patient characteristics and treatment patterns

Between August 2019 and September 2025, we identified 189 patients with advanced CSCC treated with ICI as part of their initial management plan (online supplemental figures S1 and S2). The cohort reflected real-world heterogeneity in age, immune status, disease stage, and anatomic location (online supplemental tables S1–S3). Most patients received cemiplimab (129 patients) or pembrolizumab (59 patients), while one patient received combination ipilimumab plus nivolumab in an individualized high-risk clinical context (figure 1). Immunotherapy exposure during the frontline treatment period was highly variable (figure 2 and online supplemental figures S3–S5): nearly half of patients (92, 48.7%) received two or fewer doses of systemic therapy, while others received up to 46 doses. Immunotherapy exposure also differed by downstream management strategy (online supplemental figure S3). Among patients who ultimately underwent surgical resection, immunotherapy exposure was concentrated in the first 1–4 doses, consistent with early transition to local therapy for response assessment and/or management of residual disease. In contrast, patients managed without surgery had a broader distribution of immunotherapy exposure, including extended treatment courses and higher cumulative dose counts.

Real-world responses

Despite limited immunotherapy exposure, objective clinical responses were common across all dose strata (online supplemental tables S4–S9). Overall, 119 (63%) patients achieved an objective clinical response, including 52 (27.5%) with complete responses (CRs) and 67 with partial responses (PRs); notably, early disease control was observed in a subset of the 11 patients who received only a single dose. Importantly, complete clinical responses and achievement of no evidence of disease (NED) were observed among patients receiving as few as one or two doses (online supplemental figures S6–S10). Among 50 patients who achieved cCR without surgical resection, only one recurrence was observed over a median follow-up of 25.4 months by reverse Kaplan-Meier estimation, demonstrating durable disease control with immunotherapy alone in selected patients (online supplemental figure S11). Initial responses were rapid, with a median time to first documented response of 42 days (IQR 31–74),

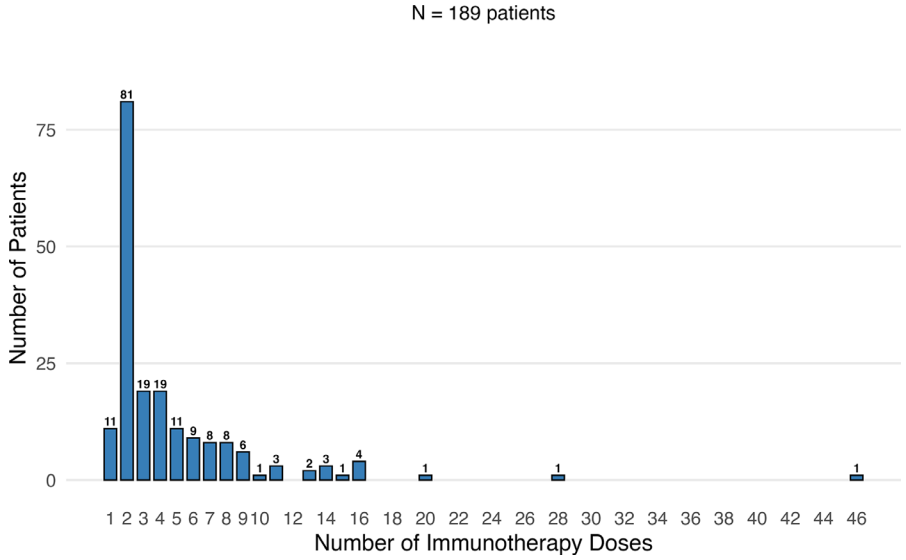

Figure 2 Distribution of immunotherapy doses. Histogram showing the overall distribution of immune checkpoint inhibitor doses received by patients in the cohort. Each bar represents the count of patients who received that specific number of doses.

though formal achievement of NED status occurred later at a median of 182 days (IQR 104–255). The heterogeneity in treatment duration reflected deliberate clinical decision-making based on observed responses.

Treatment discontinuation and reasons for cessation

Consistent with these response patterns, treatment discontinuation decisions were highly variable and reflected the lack of established evidence-based stopping rules, with clinicians adapting treatment duration based on observed responses in real time (online supplemental table S10). Among patients receiving two or fewer doses, the most common reasons for discontinuation were planned transition to surgery as part of a neoadjuvant strategy (53 patients) or achievement of clinical response (23 patients), rather than disease progression (21 patients) or toxicity (16 patients); patients could have multiple reasons for discontinuation. This underscores that limited treatment duration frequently reflected early tumor regression rather than treatment failure, introducing confounding that necessitates causal inference methods.

Survival outcomes

With a median follow-up of 21.4 months, 81 of 189 patients (42.9%) experienced an event (recurrence, progression, or death). Unadjusted Kaplan-Meier analyses demonstrated heterogeneity in outcomes across clinical subgroups and treatment pathways (figure 3, online supplemental figures S11–S13), but these comparisons should be interpreted cautiously given substantial selection effects and time-dependent treatment decisions.

Among patients rendered NED, recurrence events were uncommon overall. In the NED cohort (141 patients), the estimated 1-year, 2-year, and 3-year RFS was 95.1%, 91.2%, and 88.1%, respectively. Stratified analyses showed similar RFS by surgery status (figure 3A), with 10 recurrences among 86 surgically managed patients and 1 recurrence among 55 patients managed without surgery (CR and PR→NED, with or without radiotherapy). In contrast, progression-free survival (PFS) among the 48 patients not rendered NED was markedly lower (figure 3B), with rapid early progression and poor long-term outcomes (3-year PFS: 7.7%).

For the full cohort, OS and DSS were favorable (figure 3C,D), with 40 deaths (28 disease-specific) among 189 patients. One-year, 2-year, and 3-year OS was 86%, 74%, and 70.9%, while DSS was 88.7%, 80.5%, and 79%, respectively.

EFS, which integrates recurrence, progression, and death, provided a pragmatic overall summary endpoint (figure 3E). In the full cohort (189 patients), the estimated 1-year, 2-year, and 3-year EFS was 65.4%, 54%, and 49.4%, respectively, and did not meaningfully differ by surgery status on unadjusted comparison (figure 3E).

Finally, unadjusted EFS stratified by dose count suggested longer EFS with higher cumulative dosing (figure 3F). However, this pattern is particularly susceptible to guarantee-time bias: patients must remain event-free long enough to receive additional doses, and dose accumulation is inherently conditioned on early clinical course and response. Accordingly, dose-stratified Kaplan-Meier curves are presented descriptively as hypothesis-generating and are not interpreted as causal evidence of benefit from prolonged treatment.

Dose-response analysis

To address confounding from non-random treatment assignment and quantify the relationship between immunotherapy dose intensity and outcomes, we used Bayesian regression models with explicit adjustment for baseline confounders. Given that immunotherapy exposure, response, and subsequent surgical decisions were tightly interrelated, we used directed acyclic graphs to formalize causal assumptions and identify minimally sufficient

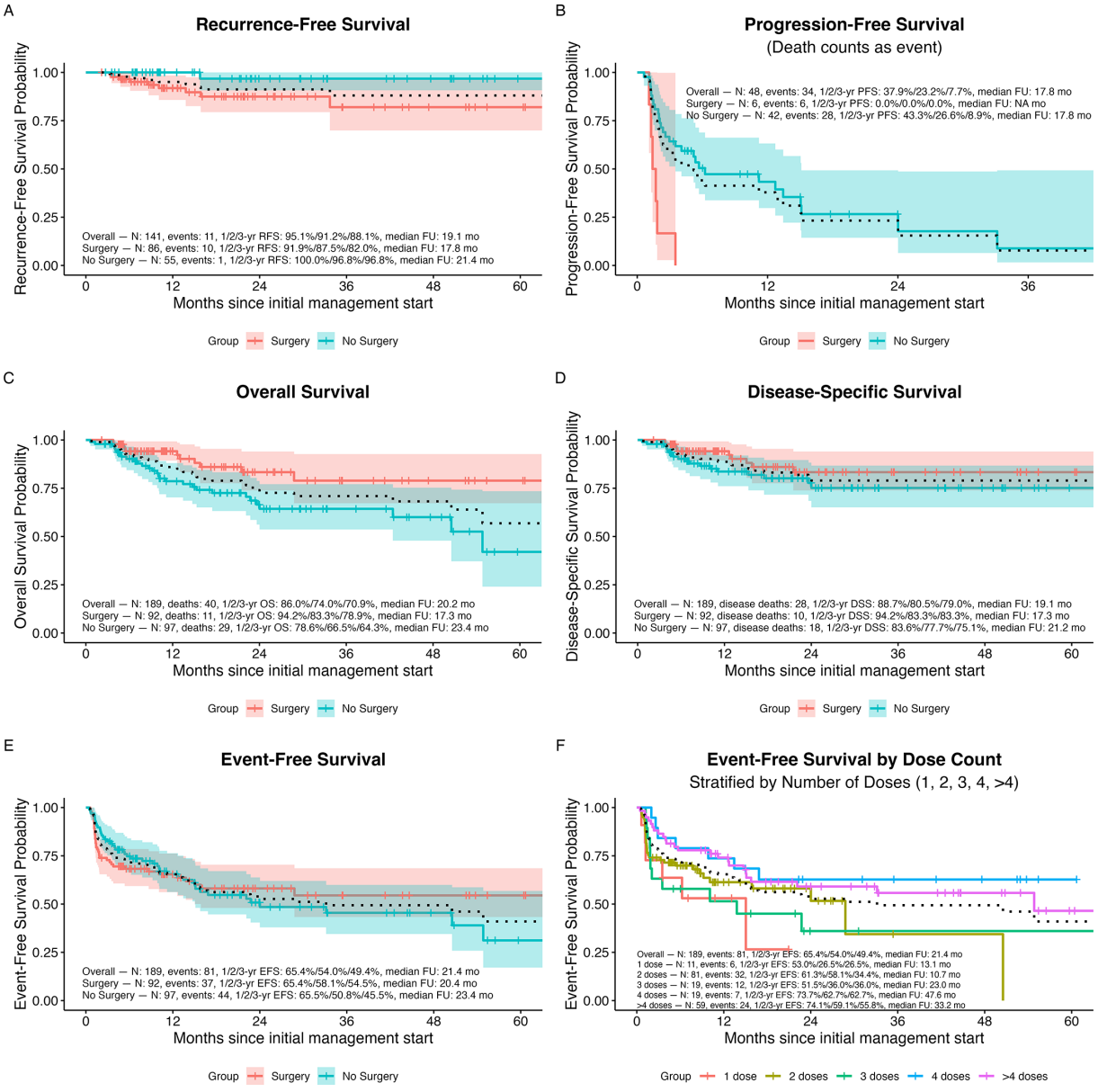

Figure 3 Survival endpoints by surgery status and immunotherapy dose. Paneled Kaplan-Meier curves showing (A) recurrence-free survival (RFS), (B) progression-free survival (PFS), (C) overall survival (OS), (D) disease-specific survival (DSS), (E) event-free survival (EFS) stratified by surgery status and (F) EFS stratified by number of immunotherapy doses received. RFS was evaluated among patients rendered with no evidence of disease (NED) following initial management, which could include systemic therapy, surgery and/or radiation therapy, and was defined as time from initial management start to first recurrence or last follow-up. PFS was evaluated among patients not rendered NED and was defined as time from initial management start to disease progression or death, whichever occurred first, with patients censored at last follow-up if neither event occurred. OS was defined as time from initial management start to death from any cause, with patients censored at last follow-up if alive. DSS was defined as time from initial management start to death attributable to cutaneous squamous cell carcinoma, with deaths from other causes censored at the time of death. EFS was defined as time from initial management start to the earliest of recurrence, progression, or death from any cause. In panel F, EFS is shown stratified by total number of immunotherapy doses received (1, 2, 3, 4, or >4 doses). In each panel, the dotted black line denotes the overall cohort for that endpoint. Shaded regions indicate 95% CIs.

adjustment sets, rather than relying on data-driven variable selection alone (figure 4 and online supplemental figure S13).

Response-based endpoints

Because the true shape of the dose-response relationship is uncertain, we evaluated three complementary model specifications to assess the robustness of the association between cumulative immunotherapy exposure and objective response: a linear model assuming a constant incremental effect per additional dose, a log-dose model allowing diminishing marginal effects as exposure increases, and a categorical threshold model estimating

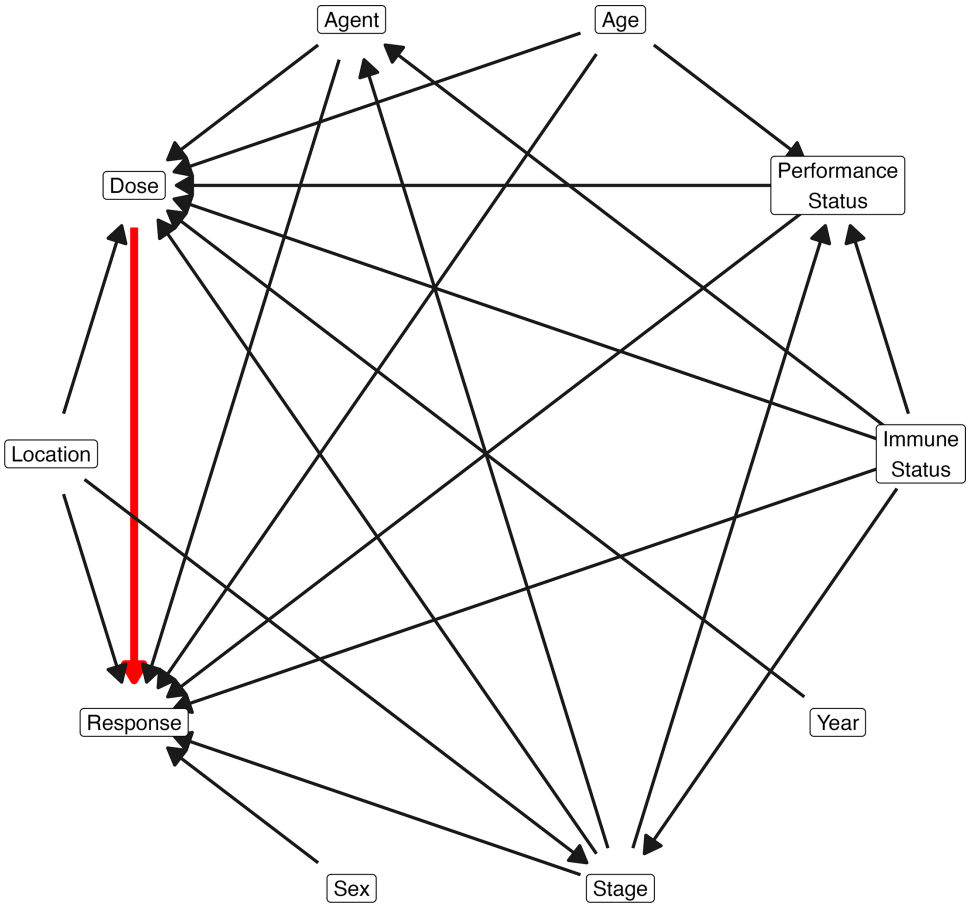

Figure 4 Directed acyclic graph for total causal effect of immunotherapy dose on treatment response. Nodes represent variables and directed edges represent hypothesized causal relationships based on clinical reasoning and domain knowledge. Age, immune status, primary tumor location, disease stage, systemic agent, and performance status were identified as confounders requiring adjustment to estimate the total causal effect of immunotherapy dose on composite treatment response (defined as clinical or pathological complete or partial response). Primary tumor location may affect disease stage at presentation, dose decisions through surgical considerations, and potentially response through anatomic site-specific biology. Performance status might act as both a mediator (influenced by age, immune status, and stage) and a confounder (directly affecting both dose and response decisions). Year of treatment affects dosing patterns through evolving institutional practice but is unlikely to directly affect biological response, and was therefore excluded from the adjustment set. Sex may influence response but likely does not affect dose decisions, and thus does not open backdoor paths requiring adjustment. Post-treatment variables (eg, toxicity, surgery decisions) are not depicted, as adjusting for them would induce collider bias and block causal pathways of interest. The minimal sufficient adjustment set, identified using the backdoor criterion, includes age, immune status, location, stage, agent, and performance status.

the incremental benefit associated with crossing prespecified dose cutpoints. Rather than relying solely on point estimates, we used posterior distributions to quantify the probability that cumulative exposure exceeded clinically interpretable effect thresholds. Together, these models address complementary clinical questions: whether additional doses are generally associated with higher response probability, whether that relationship shows diminishing returns at higher exposures, and whether specific dose thresholds are likely to confer incremental benefit.

Posterior predictions from the linear specification indicated that the probability of objective response increased gradually with increasing cumulative exposure (figure 5A). This pattern was captured by a linear dose model assuming a constant incremental effect per additional treatment cycle. Under a weakly informative prior, the posterior probability that the cumulative dose effect was >0 was 99%. The median posterior effect corresponded to an OR of approximately 1.09 per additional dose, suggesting that each additional treatment cycle was associated with only a modest increase in response probability. Framed in more clinically interpretable terms, the posterior probability that each additional dose increased the odds of response by at least 5% was 82%, whereas the probability of an increase of at least 10% was only 44% (figure 5A and online supplemental figures S15 and S16). At the model-implied baseline response probability of 68% at one dose, these OR thresholds correspond to absolute increases of approximately 1.1 and 2 percentage points in response probability. Taken together, these findings suggest a consistent directional signal favoring additional doses, although the most likely gains per dose are modest.

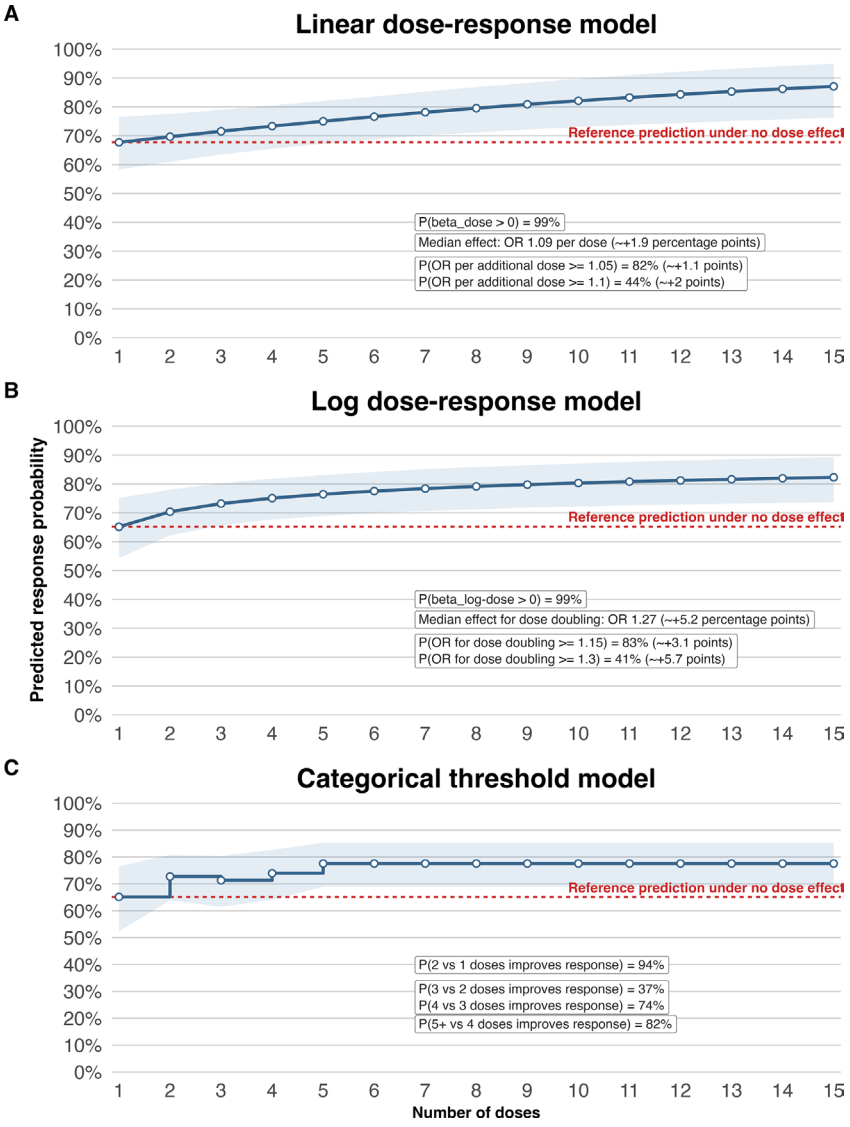

Figure 5 Posterior predicted probability of objective response across cumulative immunotherapy doses under alternative dose-response model specifications. (A) The linear dose-response model, which assumes a constant incremental effect of each additional dose on the log-odds of response. (B) The log-dose model, which allows diminishing marginal benefit with increasing cumulative exposure. (C) The categorical threshold model, which estimates separate incremental effects across successive dose thresholds. Solid blue lines represent the model-estimated predicted response probability at each dose level for a standardized reference covariate profile, and shaded ribbons denote pointwise 89% posterior credible intervals. The dashed red horizontal line indicates the reference prediction under a null dose effect, corresponding to the model-implied response probability if additional doses had no impact beyond the first dose. Importantly, the blue curves should be interpreted as model-based summaries of the estimated dose-response pattern, not as evidence that each additional dose definitively produces the displayed increase in response probability. The strength of evidence for incremental benefit is summarized in the annotation boxes. For the linear and log-dose models, these annotations report the posterior probability that the dose effect is positive and that it exceeds prespecified OR thresholds, with ORs translated into approximate absolute percentage-point changes using the predicted response probability at one dose as the baseline. Thus, the linear model supports a positive dose-response direction, but the estimated incremental benefit of each additional dose is modest, and larger per-dose effects remain uncertain. For the categorical model, annotations summarize posterior probabilities that successive dose thresholds improve response probability. Together, these panels suggest a broadly consistent positive dose-response pattern across alternative functional forms while emphasizing uncertainty in the magnitude and clinical importance of additional doses.

Notably, the model-implied baseline response probability at one dose (68%) exceeded the crude response rate observed among patients who received only a single dose (5/11, 45.5%). This model-implied estimate reflects the predicted response probability for the reference patient in the regression model—at the reference covariate values—rather than the empirical response rate among all patients who happened to receive only one dose. Patients receiving a single dose may have differed systematically from this reference profile, including having less favorable baseline characteristics or early treatment discontinuation because of

progression, toxicity, or other clinical factors. Accordingly, this difference likely reflects baseline risk heterogeneity and confounding by indication rather than a problem with model fit.

To evaluate whether the observed dose-response relationship might reflect variation in treatment timing rather than cumulative exposure itself, we conducted a sensitivity analysis adjusting the primary model for pre-response dose intensity, defined as the average spacing of doses administered prior to response adjudication. Clinically, treatment delays or interruptions may occur because of toxicity, intercurrent illness, or other factors, potentially altering the apparent relationship between cumulative dose and response.

Adjustment for dose intensity did not materially alter the estimated dose-response relationship. Predicted response curves were broadly similar after dose-intensity adjustment when compared with the primary weak-prior linear model (figure 5A, online supplemental figure S15B), and the posterior distributions of the cumulative dose coefficient showed substantial overlap across model specifications (online supplemental figure S16). The dose-intensity coefficient leaned positive but remained uncertain, suggesting limited evidence that treatment timing independently explained the observed dose-response relationship. Together, these findings suggest that the observed association between cumulative dose exposure and response was not primarily explained by variation in treatment timing.

Although the linear model provides a useful baseline description of the observed dose-response pattern, a strictly linear relationship—where each additional dose confers the same incremental benefit regardless of prior exposure—is biologically implausible. Under such a model, predicted response probabilities would continue to increase indefinitely as additional cycles are administered. Clinically, however, immune responses are more likely to emerge early and then plateau, with later treatment cycles maintaining an established response rather than substantially deepening it.

Consistent with this biological expectation, the largest increases in predicted response probability occurred early in treatment, with progressively smaller gains at higher cumulative exposures (figure 5B and online supplemental figure S17). This pattern was captured by a log-dose model in which the effect of cumulative exposure diminishes as the number of doses increases. Under this specification, the posterior probability that the log-dose coefficient was >0 was 99%.

To facilitate clinical interpretation, we examined the effect of doubling the number of doses (eg, from 1 to 2 doses or from 2 to 4 doses). Under the weakly informative prior log-dose model, the posterior median effect for dose doubling corresponded to an OR of approximately 1.27. The posterior probability that doubling the dose increased the odds of response by at least 15% was 83%, while the probability of an increase of at least 30% was 41% (figure 5B and online supplemental figure S18).

Translating these OR thresholds into absolute probability changes using the model-implied baseline response probability at one dose (64%), a 15% increase in odds corresponds to an absolute increase of approximately 3 percentage points in response probability, while a 30% increase corresponds to an increase of approximately 6 percentage points. As illustrated by the model-based prediction curves (figure 5B), the largest increases in predicted response probability occur early in treatment, with progressively smaller gains as cumulative exposure increases, consistent with diminishing marginal returns.

We next fit a categorical threshold model to estimate the incremental benefit associated with crossing successive dose thresholds, thereby relaxing assumptions of linearity and directly addressing the clinically intuitive question of whether moving from one dose level to the next is likely to improve response probability. The strongest evidence for benefit was observed at the earliest threshold: the posterior probability that moving from 1 to 2 doses improved response was 94% (figure 5C, online supplemental figures S19 and S20). In contrast, support for later thresholds was more variable, with posterior probabilities of benefit of 37% for 3 versus 2 doses, 74% for 4 versus 3 doses, and 82% for 5+ versus 4 doses. These results suggest that if cumulative dose improves response, much of that benefit is likely concentrated early—particularly at the transition from one to two doses—whereas the incremental value of later doses remains smaller and more uncertain.

Sensitivity analyses restricted to doses administered before response assessment produced similar posterior dose-effect estimates across linear, log-dose, and categorical ORR models (online supplemental figure S21), suggesting that the observed exposure-response pattern was not driven by inclusion of post-response doses.

This front-loaded pattern suggests that the key downstream question is whether continuing treatment beyond the early response window translates into more durable disease control. To address this, we examined EFS.

Event-free survival

We used a prespecified landmark analysis restricted to patients who remained event-free long enough to receive at least two doses of immunotherapy (n=177). Within this cohort, we compared patients who stopped after two doses (81 patients) with those who received three or more doses (96 patients) (figure 6). This design reduces guarantee-time bias and yielded a relatively balanced comparison, supporting stable estimation.

Figure 6 presents this analysis in two complementary ways. The posterior predicted survival curves show the model-implied EFS trajectories under each dosing strategy, averaged across the observed patient population, while the posterior HR distribution displays the full uncertainty around the corresponding dose effect. The predicted curves modestly favored continued treatment beyond two doses, but separation between groups was limited and the 89% credible intervals overlapped

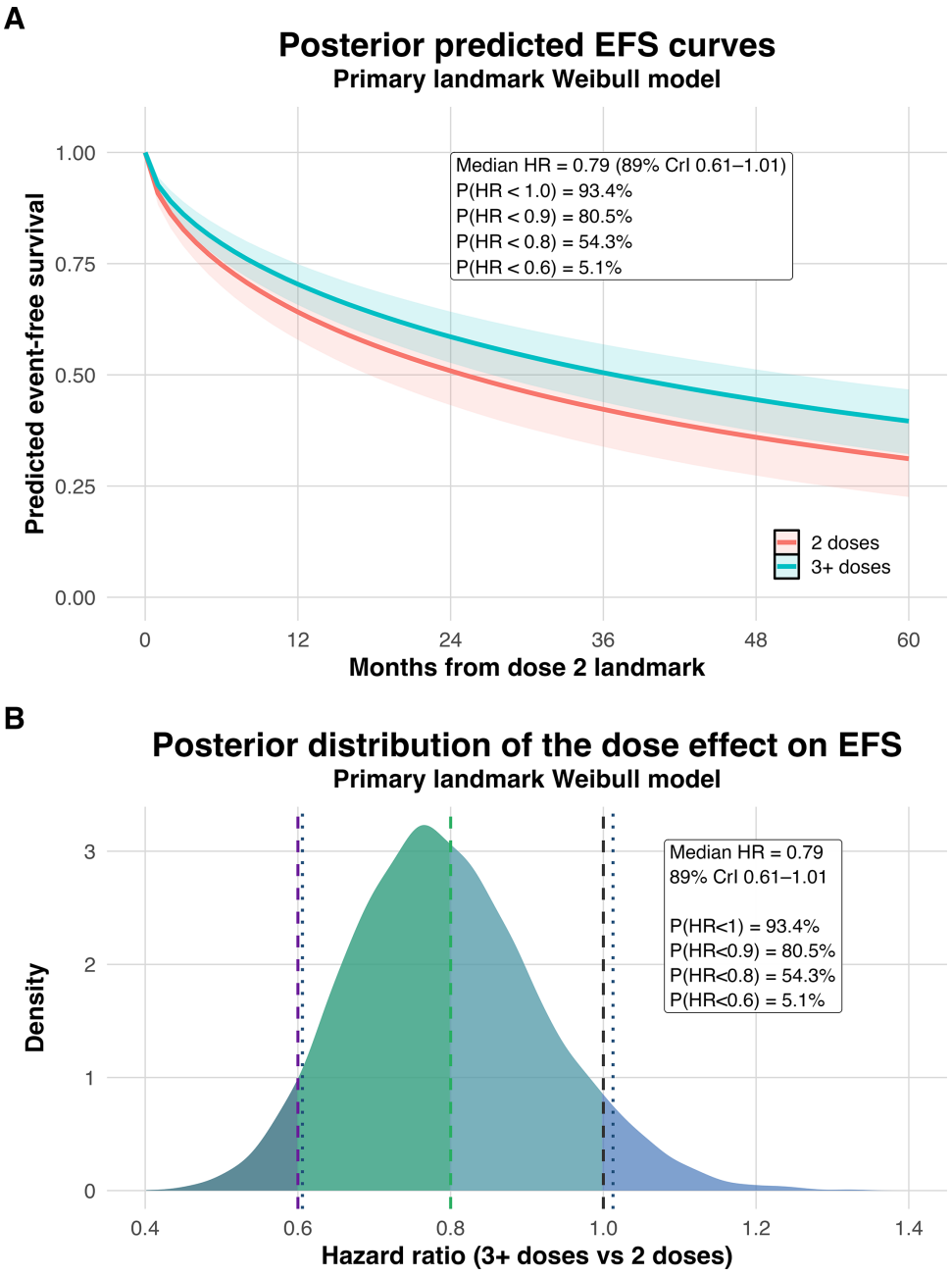

Figure 6 Primary event-free survival analysis integrating posterior predicted survival curves and posterior HR uncertainty. (A) Posterior predicted event-free survival curves from the primary landmark Weibull model. Survival curves represent posterior predictions averaged across the observed patient population. For each posterior draw, predicted survival was computed after assigning all patients to either 2 doses or 3+ doses of immunotherapy while leaving all other covariates unchanged. Shaded regions denote 89% credible intervals. The inset annotation summarizes the corresponding posterior distribution of the HR—linking the separation between curves to the magnitude and uncertainty of the treatment effect. (B) Posterior distribution of the HR comparing 3+ doses versus 2 doses under the same model. Dashed vertical lines denote clinically interpretable thresholds at HR=1.0, 0.8, and 0.6, and dotted vertical lines indicate the 89% credible interval. Shaded density regions represent posterior probability mass below these thresholds. Together, these panels show how differences in predicted survival over time relate directly to the probability that the treatment effect exceeds clinically meaningful thresholds. EFS, event-free survival.

substantially across follow-up. Consistent with this visual pattern, the posterior median HR for 3+ versus 2 doses was 0.79 with an 89% credible interval of 0.61–1.01.

The posterior distribution nonetheless leaned in a protective direction. Under the primary landmark model, the posterior probability that receiving three or more doses was associated with a lower event hazard than stopping after two doses was 93.4%. However, the estimated magnitude of benefit remained modest and uncertain. The posterior probability that the HR was below 0.9 was 80.5%, below 0.8 was 54.3%, and below 0.6 was 5.1%. Taken together, these results suggest that if additional doses improve downstream disease control, the expected effect is more likely to be small than dramatic.

Sensitivity analyses were directionally consistent with the primary model (online supplemental figures S22 and S23). Across the skeptical-prior landmark model, the dose-intensity-adjusted landmark model, and the surgery-adjusted landmark model, posterior HR distributions remained concentrated in a similar range and continued to support, at most, a modest protective association rather than a large treatment effect. The surgery-adjusted analysis should be interpreted cautiously because surgery may lie on the causal pathway between early response and long-term outcome; adjustment may therefore partially block mediation rather than simply account for confounding. A no-landmark model from treatment initiation yielded somewhat stronger apparent benefit, but this pattern is consistent with residual immortal time bias and is therefore better interpreted as a prognostic association than as causal evidence that longer treatment improves EFS.

Taken together, these EFS analyses reinforce the same general message seen in the response models: whatever benefit may accrue from additional immunotherapy exposure appears most likely to be concentrated early, while the incremental value of continuing beyond two doses remains uncertain and, if real, probably modest in magnitude.

Model validation

Because the primary dose-response analyses yielded modest effect sizes with substantial posterior uncertainty, we performed simulation-based calibration to determine whether this uncertainty arose from limited information in the data or from insufficient model sensitivity. Synthetic datasets matched to the observed cohort were generated under prespecified dose-response structures spanning null, linear, non-linear, and threshold effects (figure 7).

Under a null linear scenario (true β=0), the posterior distribution showed a modest positive lean, with an 83.5% posterior probability that β>0. However, the 89% credible interval crossed the null and there was limited posterior probability of a larger effect, consistent with finite-sample variability rather than systematic spurious signal detection

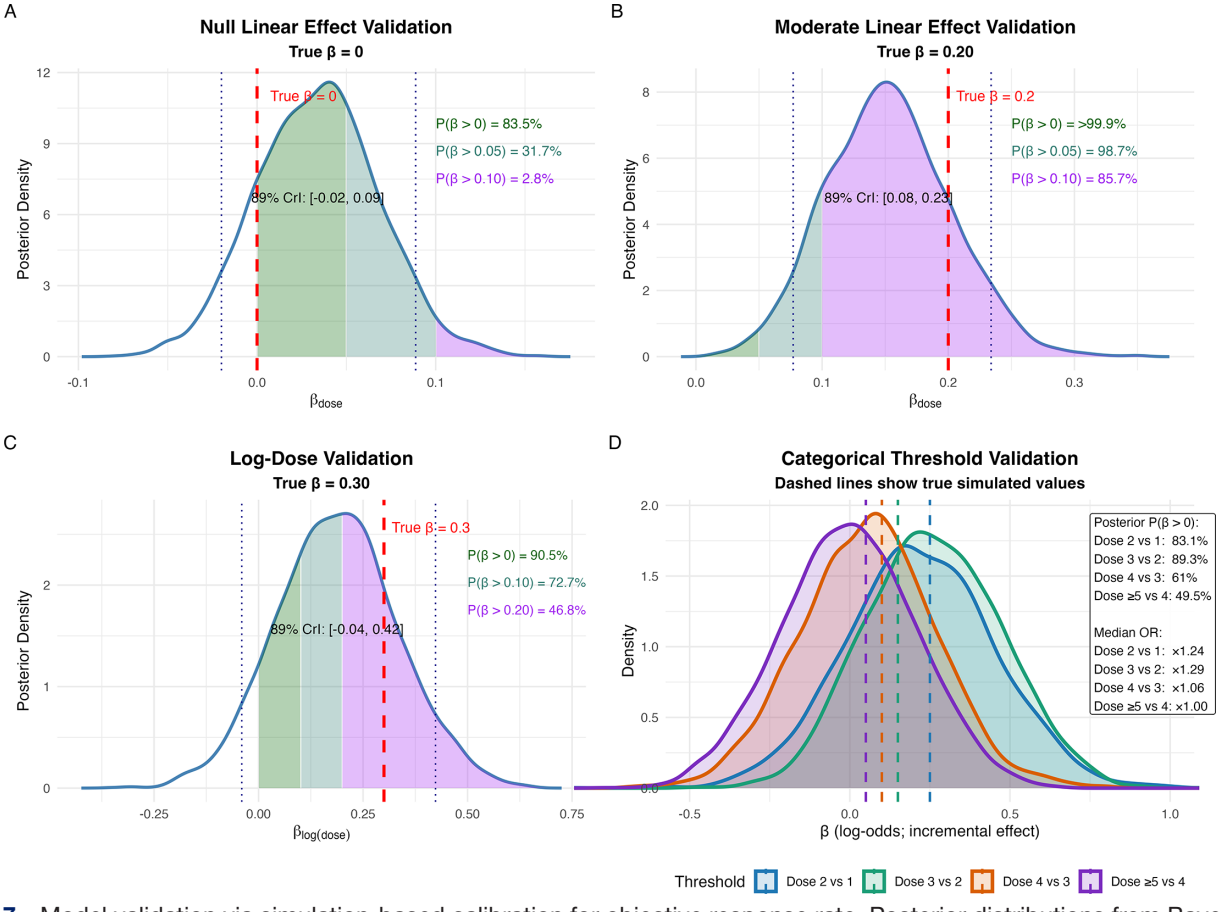

Figure 7 Model validation via simulation-based calibration for objective response rate. Posterior distributions from Bayesian response models fitted to simulated datasets with known dose-effect structures. (A) Null linear scenario: true β=0; the model remains centered near the null while appropriately preserving uncertainty. (B) Moderate linear effect: true β=0.20; the model recovers a clinically meaningful positive dose-response signal. (C) Log-dose scenario: true β=0.30 on the log-dose scale; the model recovers a diminishing-returns pattern consistent with stronger early gains at lower cumulative exposure. (D) Categorical threshold model: true incremental effects β_2=0.25, β_3=0.15, β_4=0.10, and β_{5+}=0.05 (dashed lines); the model recovers heterogeneous threshold effects without imposing a single functional form. Together, these panels show that the modeling framework can distinguish null, linear, non-linear, and thresholded dose-response structures while quantifying posterior uncertainty.

(figure 7A). Under a moderate linear effect scenario (true β=0.20), the model recovered a positive association, with posterior probability >99% that β>0 and 85.7% probability that β exceeded 0.10, consistent with detection of a modest clinically meaningful effect (figure 7B).

In a non-linear log-dose scenario with diminishing returns (true β=0.30), the fitted model recovered a positive log-dose effect, although with meaningful uncertainty and a posterior median below the true value (median β=0.19; 89% CrI, −0.04 to 0.42). The posterior probability that β>0 was 90.5%, while the probabilities that β exceeded 0.10 and 0.20 were 72.7% and 46.8%, respectively (figure 7C).

Under a categorical threshold scenario (true β_2=0.25, β_3=0.15, β_4=0.10, β_{5+}=0.05), the model recovered the expected front-loaded pattern most clearly at the earlier dose thresholds. The posterior probabilities of a positive incremental effect were 83.1% for 2 versus 1 dose and 89.3% for 3 versus 2 doses, compared with 61.0% for 4 versus 3 doses and 49.5% for ≥5 versus 4 doses. Median incremental ORs showed a similar pattern, with larger estimated effects for the earlier thresholds than for later thresholds (OR 1.24, 1.29, 1.06, and 1.00, respectively; figure 7D).

Together, these simulations show that the modeling framework behaved as expected across null, moderate, non-linear, and thresholded dose-response scenarios. The models did not generate strong signals under the null, recovered positive effects when moderate effects were simulated, and reflected greater uncertainty when effects were smaller or distributed across thresholds. Thus, uncertainty in the clinical analyses is most consistent with limited information in the available data rather than an inherent inability of the modeling approach to detect dose-response effects.

Role of local consolidation therapies

Surgical resection was performed in 92 patients, most commonly in the context of partial response (48 patients, 52%), followed by progressive disease (29, 32%) and stable disease (13, 14%). Notably, only two patients with complete clinical responses underwent surgical resection, reflecting a temporal shift in clinical practice over the study period from an initial approach in which most patients proceeded to surgery following immunotherapy to a later strategy of selectively omitting surgery in patients with favorable early responses and reserving operative intervention for those with suboptimal response or persistent disease.

Among the 48 surgical patients with partial clinical responses, 23 (48%) achieved complete pathological response, indicating complete tumor eradication despite persistent clinical abnormalities. An additional 24 (50%) demonstrated pathological evidence of treatment response. Thus, the vast majority of clinical partial responders who underwent resection had substantial pathological tumor regression, highlighting a marked discordance between clinical and pathological assessment.

Although 63% of patients achieved objective clinical responses to immunotherapy, 74.6% were ultimately rendered NED. This difference reflects the contribution of response-guided local therapy to disease clearance among patients with incomplete responses to immunotherapy alone. Radiation was used primarily in postoperative (n=23) or salvage settings (n=14), with six patients achieving NED following radiation therapy (RT) (online supplemental table S11). Together, these findings contributed to an evolving institutional approach in which surgery was increasingly used selectively, both as a consolidative intervention and as a means of confirming depth of response.

Recurrence rates and survival outcomes did not clearly favor surgical consolidation in unadjusted comparisons; however, these comparisons should be interpreted cautiously because decisions regarding surgery and radiation reflected clinical judgment regarding resectability, disease extent, treatment response, and patient fitness. Importantly, both surgery and radiation occurred after initial immunotherapy response, making them post-treatment variables on the causal pathway between immunotherapy exposure and downstream outcomes. Surgery may also serve as a mechanism for response assessment and treatment de-escalation, facilitating earlier discontinuation of systemic therapy, whereas omission of surgery may be associated with longer treatment duration in patients with favorable clinical response. As a result, standard regression adjustment for such post-treatment variables can induce collider bias and obscure causal interpretation.[9 10] Proper estimation of direct and indirect effects would require formal mediation analysis, which was not feasible given the limited number of events. These considerations are discussed further in the Discussion section.

DISCUSSION

In this retrospective cohort of 189 patients with high-risk resectable or locally advanced CSCC treated with immunotherapy as part of initial management, we observed high real-world response rates (63% objective response) and favorable disease control among patients rendered NED. Durable responses were frequently achieved after limited immunotherapy exposure—nearly half of patients received two or fewer doses—and nearly all complete responders managed without surgery remained disease-free (49 of 50 patients). This pattern—frontline immunotherapy followed by response-guided subsequent treatment—characterized the majority of treatment decisions in this cohort and aligns with emerging response-adapted strategies in prospective trials. Using Bayesian causal inference to address confounding from treatment decisions based on response, we found that additional immunotherapy exposure was associated with, at most, modest improvements in outcomes, with substantial uncertainty surrounding the incremental value of prolonged treatment. Taken together, these findings

suggest that minimal effective dosing may be sufficient for many patients, while the added benefit of extended immunotherapy or routine surgical consolidation in deep responders remains unclear.

Our dose-response analyses nonetheless revealed a consistent directional signal favoring additional immunotherapy doses, though the magnitude of benefit was modest and attenuated with continued treatment. In particular, models without strict linearity assumptions suggested that any benefit appears concentrated early in treatment, though substantial uncertainty remains regarding the optimal minimal dose for individual patients, with diminishing incremental gains beyond initial exposure. Immune checkpoint blockade acts by reinvigorating antigen-experienced T cells and enhancing systemic antitumor immunity, and even a single dose of PD-1 inhibition has been shown to induce rapid pharmacodynamic T-cell activation and pathological tumor regression in neoadjuvant settings.[11] Mechanistic studies further suggest that early antigen exposure in the presence of checkpoint blockade promotes T-cell priming and expansion, leveraging the abundant tumor antigen milieu prior to resection.[12] These biological principles align with our observation that much of the apparent clinical benefit accrues early, while additional doses may sustain rather than deepen antitumor responses. Prospective neoadjuvant trials have similarly demonstrated that short courses—typically 2–4 doses—can achieve high pathological response rates, although they were not designed to estimate the marginal benefit of additional cycles beyond the initial response window.[2–4] These findings are further supported by emerging real-world data suggesting that early discontinuation of immune checkpoint inhibition after initial response may be associated with favorable outcomes in selected patients,[13] reinforcing the potential safety of response-adapted treatment strategies. Our findings extend this literature by suggesting that most therapeutic benefit may occur early, while later cycles offer diminishing and uncertain returns.

That said, several limitations warrant consideration. First, the observational design introduces confounding from non-random treatment assignment that cannot be fully eliminated through statistical adjustment. Although we used directed acyclic graphs to formalize causal assumptions, identify minimally sufficient adjustment sets, and ensure we did not adjust for post-treatment variables that would block causal pathways, unmeasured factors—including nuanced clinical judgment regarding disease trajectory—likely influenced both treatment decisions and outcomes. Inclusion of patients with limited treatment exposure may introduce heterogeneity related to early clinical course. However, naïve exclusion of such patients would risk conditioning on post-baseline treatment exposure; therefore, for causal analyses, we used a prespecified landmark approach to restrict the cohort to patients who remained event-free long enough to receive at least two doses, thereby mitigating this bias.

Second, practice patterns evolved substantially over the 6-year study period. In the early years of cohort accrual, most patients were managed with a uniformly neoadjuvant approach and proceeded to surgery following initial immunotherapy regardless of response. Over time, as clinicians observed frequent early responses and favorable outcomes in selected patients, management shifted toward a response-guided paradigm in which surgery and other local therapies were increasingly reserved for patients with suboptimal response, progression, or clinically meaningful residual disease. This evolution reflects real-world clinical learning but introduces important complexity for causal interpretation, as changes in treatment intensity and selection occurred concurrently with increasing experience and refinement in patient selection.

Third, the landmark analysis used for EFS mitigates—but does not eliminate—immortal time bias or time-dependent selection. Post-baseline decisions, including surgical resection, radiation therapy, and continuation or discontinuation of immunotherapy, were made in response to evolving clinical status and therefore limit causal interpretation of downstream outcomes. Because surgery and radiation occurred after initial immunotherapy response, they may lie on the causal pathway between immunotherapy exposure and subsequent outcomes. Conditioning on such variables—through stratification or covariate adjustment—can block part of the treatment pathway, induce collider bias, and distort causal effect estimates.[9 10] Formal estimation of total, direct, and mediated effects would require counterfactual mediation methods,[14] but limited event counts precluded stable estimation using such approaches.

We therefore performed a surgery-adjusted landmark model as an exploratory sensitivity analysis to assess whether accounting for observed surgery use materially changed the estimated association between receiving ≥3 versus 2 doses and EFS. The dose-effect estimate changed minimally after adjustment for surgery (median HR 0.8 with surgery adjustment vs 0.79 in the primary landmark model), suggesting that differential surgery use did not substantially explain the observed dose-EFS association. However, because surgery occurred downstream of response-adapted clinical decision-making, this analysis should be interpreted as a robustness check of the adjusted association rather than as an estimate of the total effect of dose operating through subsequent local therapy decisions.

Fourth, the sample size and number of events constrained analytic scope. We therefore focused on response-based endpoints and EFS as outcomes most amenable to inference, but lacked sufficient events to achieve precise estimates for RFS, DSS, and OS separately. Follow-up was reasonably mature, with a median of 21.4 months for the overall cohort by reverse Kaplan-Meier estimation and 25.4 months among the 50 complete responders managed without surgery. Even so, additional follow-up will be important to more fully define the long-term recurrence risk in this subgroup.

Fifth, we used clinician-anchored response assessment rather than formal RECIST or iRECIST criteria. While this introduces potential misclassification—for example, two patients deemed NED by treating clinicians were later found to have persistent lymph node disease that subsequently progressed (online supplemental figure S6)—this approach more closely reflects real-world clinical practice, where standardized response criteria are rarely applied outside of clinical trials. This trade-off between standardization and ecological validity is inherent to pragmatic observational studies: our findings may be more generalizable to routine practice precisely because they capture the ambiguity and clinical judgment that characterize real-world decision-making. However, such misclassification may influence response-adapted treatment decisions and bias estimates of the relationship between early response and subsequent management.

Sixth, this single-institution experience may limit generalizability. Practice patterns, patient populations, and institutional expertise at a tertiary referral center may differ from community settings, and our findings require validation in broader populations and practice contexts.

A separate methodological consideration concerns our use of Bayesian inference. This approach is well suited to quantifying uncertainty, especially in complex observational settings where treatment decisions are influenced by patient characteristics and early clinical course. Bayesian methods provide direct probabilistic interpretation of treatment effects—allowing statements such as 'the posterior probability of benefit exceeds X%'—which align more naturally with clinical decision-making than frequentist p-values. Multiple prior specifications demonstrated robustness of our conclusions to analytic assumptions, while simulation-based calibration confirmed adequate model sensitivity. However, fundamental uncertainty remains: we cannot definitively establish whether the dose-response relationship is linear, logarithmic, or follows some other pattern from observational data alone. Our comparative approach—fitting multiple model structures—mitigates this limitation. The modest effect sizes and wide credible intervals reflect genuine uncertainty rather than methodological inadequacy, underscoring the challenge of causal inference in pragmatic observational studies with limited sample sizes.

A particularly consequential clinical observation was the striking discordance between clinical and pathological response assessment noted above—a phenomenon that has been reported previously in neoadjuvant immunotherapy studies. In this pattern, residual abnormalities on clinical examination or imaging often corresponded to minimal or absent viable tumor at resection, reflecting treatment-related inflammation or fibrosis rather than persistent disease. Similar discordance has been described in prospective neoadjuvant trials, including the seminal studies by Gross and colleagues, where substantial pathological responses were frequently observed despite incomplete clinical regression.[3 15] In our cohort, repeated real-time observations of clinicopathological discordance influenced evolving treatment decisions. Clinicians became increasingly comfortable with observation and treatment de-escalation in patients demonstrating early and convincing clinical responses, recognizing that apparent residual disease on examination or imaging might not represent viable tumor. This institutional learning occurred contemporaneously with emerging prospective evidence supporting response-adapted strategies. Notably, in the MATISSE trial, 10 patients with deep responses elected to forgo surgery, with nine achieving durable remission without additional therapy, while the De-Squamate study similarly demonstrated high rates of durable disease control using response-adapted de-escalation approaches.[4 5] Together, these data reinforce that observed discordance between clinical and pathological response can be leveraged safely to individualize local therapy in selected patients. Importantly, these observations should not be interpreted as evidence against surgery per se, but rather as underscoring the absence of comparative data defining its incremental benefit among patients with substantial clinical responses.

These findings have several clinical implications. First, although durable responses frequently occurred after limited immunotherapy exposure—with suggestive but inconclusive evidence that some patients derive additional benefit at the transition from one to two doses, and substantial variability in optimal dose intensity across patients—immune checkpoint inhibition is not a benign intervention. Immune-related adverse events, including rare but life-threatening toxicities, remain an important consideration, particularly in an older and comorbid population. Response alone therefore does not imply that continued or intensified immunotherapy is necessarily the optimal strategy for every patient.

Second, because the observational nature of these data precludes definitive treatment recommendations, prospective trials directly comparing abbreviated versus extended dosing regimens would be highly informative. However, such trials face substantial feasibility challenges and may never be conducted at scale. In this context, careful and transparent analysis of real-world data—despite inherent limitations—remains essential for informing daily clinical decision-making in the absence of randomized evidence.

Third, our experience suggests that partial clinical responses do not invariably mandate immediate escalation—whether to additional immunotherapy doses or local consolidation. In carefully selected patients managed with close surveillance, continued tumor regression was frequently observed over time without continuing systemic therapy, and we did not observe loss of surgical opportunity or compromise of subsequent treatment plans by allowing responses to mature. This supports a deliberate, response-adapted approach—sometimes described in clinical practice as 'slow medicine'—in which clinicians exercise restraint in escalating to additional treatment modalities, allowing both the immune response and underlying tumor biology to declare

themselves before committing patients to additional immunotherapy cycles or irreversible local interventions. This approach recognizes that both systemic and local therapies carry meaningful toxicity and morbidity, and that premature escalation may expose patients to unnecessary harm when observation alone might suffice.

Taken together with the accumulating evidence from prospective trials, these findings support a shift away from fixed-duration neoadjuvant immunotherapy with predetermined surgical intent toward a response-adapted strategy we term FIRST. In this approach, immunotherapy is initiated upfront, clinical responses are assessed early and iteratively, and subsequent management—including continued systemic therapy, surgery, radiation, or observation—is individualized based on response depth, durability, and patient-specific considerations. Importantly, this approach is not novel in practice: response-adapted strategies have already been operationalized in prospective studies such as De-Squamate and MATISSE, though without a unifying conceptual framework.

We propose FIRST as more than semantic reframing: it clarifies treatment duration as an adaptive clinical decision rather than a predetermined protocol requirement and foregrounds the key unmet need of identifying non-responders who require early local therapy integration. Accordingly, future research should prioritize identifying patients with incomplete, heterogeneous, or non-durable responses—those most likely to benefit from surgery or radiation—enabling more precise targeting of multimodal therapy to those who need it most.

Author affiliations
[1]Mass General Brigham Cancer Institute, Boston, Massachusetts, USA
[2]Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA
[3]Harvard Medical School, Boston, Massachusetts, USA
[4]Dermatology, Massachusetts General Hospital, Boston, Massachusetts, USA
[5]Otolaryngology, Mass Eye and Ear, Boston, Massachusetts, USA
[6]Mass Eye and Ear, Boston, Massachusetts, USA
[7]Department of Ophthalmology, Mass Eye and Ear, Boston, Massachusetts, USA
[8]Dermatology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[9]Department of Radiation Oncology, Mass General Brigham Inc, Boston, Massachusetts, USA
[10]Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA

Acknowledgements The authors thank the members of the Society of Cutaneous Oncology who participated in the March 9, 2026, Research Meeting for their thoughtful review and constructive feedback. We specifically thank Dr. Ann Silk for suggesting an analysis accounting for treatment delays and other deviations from the planned dosing schedule, which prompted the development of a regimen-adjusted dose-intensity variable for use in sensitivity analyses.

Contributors DMM conceived and designed the study; performed patient-level data abstraction; conducted data curation, statistical modeling, and formal analyses; developed the analytic codebase; and drafted the initial manuscript. All authors contributed to patient care, provided critical input on study design and clinical interpretation, and participated in review and revision of the manuscript for important intellectual content. All authors approved the final version of the manuscript for submission. DMM is the guarantor of the study and accepts full responsibility for the integrity of the work as a whole, including the accuracy of the data, the validity of the analyses, and the decision to submit for publication. Artificial intelligence-assisted tools were used to support manuscript preparation. ChatGPT 4o and o1 (OpenAI, San Francisco, CA) assisted with literature review, manuscript drafting, and editing. SolveitAI (fast.ai) was used for code development and refinement. All study design, statistical analyses, data interpretation, and final content decisions were made by the authors, who reviewed and revised all AI-generated content for accuracy.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests DMM has received honoraria for serving as a consultant or participation on advisory boards for Almirall, Bristol Myers Squibb, Merck, EMD Serono, Regeneron, Sanofi Genzyme, Pfizer, Castle Biosciences, Checkpoint Therapeutics, and Sun Pharmaceuticals. He serves on the Unloxcyt steering committee for Sun Pharmaceuticals. He has stock options from Avstera Therapeutics. He has received institutional research funding from Regeneron, Kartos Therapeutics, NeoImmune Tech, Inc, Project Data Sphere, ECOG-ACRIN, and the American Skin Association. HLK is an employee of Ankyra Therapeutics. He has served in advisory roles for Castle Biosciences, Marengo Therapeutics, Tatum Biosciences, and Virogin Biotech; serves on the Board of Crichton Biosciences; holds stock in Replimune, Inc.; and has received honoraria from the Society for Immunotherapy of Cancer. RJS has served as a consultant or participated on advisory boards for Merck, Pfizer, Replimune, Marengo, PharmaMar, and Novartis. He has received institutional research funding from Merck. CP has served as a consultant for Novartis (ended April 2025). He has received an investigator-initiated research grant from Varian (ended December 2024). KSE has served as a consultant or participated in advisory boards for Regeneron Pharmaceuticals. He has received institutional research funding from Haystack by Quest Diagnostics.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Massachusetts General Hospital (IRB Protocol #: 2019P002459). Informed consent was waived by the Institutional Review Board for this retrospective study using de-identified data from electronic medical records. Waiver of informed consent was granted because the research involved no more than minimal risk, could not practicably be conducted without the waiver, and the waiver did not adversely affect the rights and welfare of subjects.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. De-identified individual participant data underlying the results reported in this article may be made available upon reasonable request to the corresponding author, subject to institutional review, data use agreements, and applicable privacy and regulatory requirements. The analytic code and reproducible workflows used for this study are publicly available at: https://themillerlab.github.io/FIRST/https://github.com/TheMillerLab/FIRST.

ORCID iDs
David M Miller https://orcid.org/0000-0001-9736-9119
Howard L Kaufman https://orcid.org/0000-0003-1131-004X
Ryan J Sullivan https://orcid.org/0000-0001-5344-6645

REFERENCES

1 Rogers HW, Weinstock MA, Feldman SR, *et al*. Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the U.S. Population, 2012. *JAMA Dermatol* 2015;151:1081–6.
2 Ferrarotto R, Amit M, Nagarajan P, *et al*. Pilot Phase II Trial of Neoadjuvant Immunotherapy in Locoregionally Advanced, Resectable Cutaneous Squamous Cell Carcinoma of the Head and Neck. *Clin Cancer Res* 2021;27:4557–65.

3 Gross ND, Miller DM, Khushalani NI, *et al*. Neoadjuvant Cemiplimab for Stage II to IV Cutaneous Squamous-Cell Carcinoma. *N Engl J Med* 2022;387:1557–68.
4 Breukers SE, Traets JJH, van Dijk SW, *et al*. Neoadjuvant ipilimumab and nivolumab in resectable cutaneous squamous cell carcinoma: a randomized phase 2 trial. *Nat Med* 2025;31:4055–64.
5 Ladwa R, Lee JH, McGrath M, *et al*. Response-Adapted Surgical and Radiotherapy De-Escalation in Resectable Cutaneous Squamous Cell Cancer Using Pembrolizumab: The De-Squamate Study. *J Clin Oncol* 2025;43:2888–96.
6 Rischin D, Porceddu S, Day F, *et al*. Adjuvant Cemiplimab or Placebo in High-Risk Cutaneous Squamous-Cell Carcinoma. *N Engl J Med* 2025;393:774–85.
7 McElreath R. Rethinking: Statistical rethinking book package. 2024.
8 Bürkner P-C. brms: an R package for Bayesian multilevel models using stan. *J Stat Softw* 2017;80:1–28.
9 Hernán MA, Robins JM. *Causal inference: what if*. Chapman & Hall/CRC, 2020.
10 Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* 2008;8:70.
11 Huang AC, Orlowski RJ, Xu X, *et al*. A single dose of neoadjuvant PD-1 blockade predicts clinical outcomes in resectable melanoma. *Nat Med* 2019;25:454–61.
12 Topalian SL, Taube JM, Pardoll DM. Neoadjuvant checkpoint blockade for cancer immunotherapy. *Science* 2020;367:eaax0182.
13 Ladwa R, McGrath M. Outcomes from early discontinuation of immune checkpoint inhibitor therapy following response in patients with advanced cutaneous squamous cell carcinoma. *EJC Skin Cancer* 2025;3:100759.
14 VanderWeele TJ. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2015.
15 Gross ND, Miller DM, Khushalani NI, *et al*. Neoadjuvant cemiplimab and surgery for stage II-IV cutaneous squamous-cell carcinoma: follow-up and survival outcomes of a single-arm, multicentre, phase 2 study. *Lancet Oncol* 2023;24:1196–205.