

Interpreting External-control Survival In Sequential-care Lung Cancer: Time-zero Alignment And Index Date Imputation

C. Hu¹, Q. Le Coent¹, A.K. Sinha², F. Tian², S. Wang², R. Curley², M-C. Wang¹, G. Rosner¹, P. Mishra-Kalyani²

¹*Johns Hopkins University, Baltimore, MD, USA* ²*Food and Drug Administration, Silver Spring, MD, USA*

Correspondence: Chen Hu, huc@jhu.edu

INTRODUCTION

WHY SEQUENTIAL-CARE QUESTIONS ARE COMMON IN LUNG CANCER

Maintenance follows systemic therapy; consolidation follows chemoradiotherapy. At each later decision, patients must be alive, observed, and eligible. Progression, toxicity, performance decline, or poor recovery can prevent entry. Initiators are therefore a clinically selected population.

COMMON LUNG-CANCER DECISION GATES

Setting	Sequence	Who reaches the later decision?
Metastatic NSCLC	Induction → maintenance	Disease control; tolerability; PS
Stage III NSCLC	CRT → consolidation	No progression; recovery; window
Oligometastatic	Systemic → local treatment	Response/stability; lesion fit; fitness
Emerging ctDNA	Timed result → adaptation	Valid result; current eligibility

CHOOSE THE CLINICAL QUESTION BEFORE THE ANALYSIS

Pathway-entry strategy	Includes all patients at initial-care entry, including nonreachers.
Later conditional decision	Eligible patients at reassessment; time zero is the later decision.

- OBJECTIVES**
1. Define the later-decision estimand and its time zero.
 2. Develop index date imputation (IDI) for unobserved control decision dates.
 3. Evaluate the framework in Monte Carlo simulations and E5508.

METHODS

1 WHY ASSIGNING A DATE IS NOT ENOUGH

Wrong clinical clock

Follow-up starts at the later decision for one cohort but at pathway entry for the other; immortal time bias can arise and operate bias can arise in either direction.

Wrong risk set

Controls may include patients who died, progressed, deteriorated, or were no longer observed before that decision.

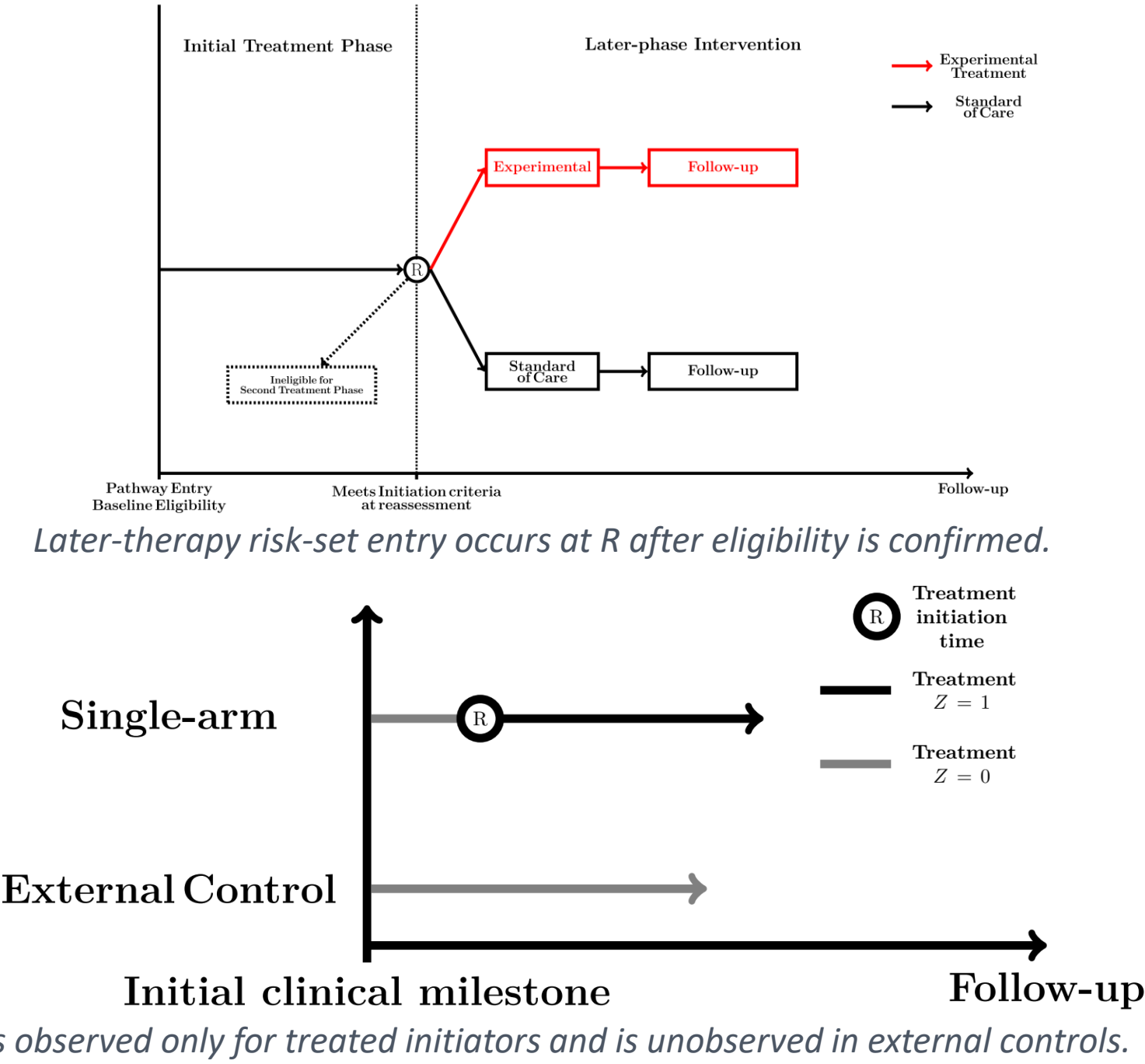
Missing gate information

A pseudo-date cannot recreate unmeasured response, recovery, organ function, performance status, or biomarker eligibility.

Residual noncomparability

Alignment cannot repair poor overlap, unequal outcome ascertainment, or unmeasured confounding.

2 DATA-STRUCTURE ASSUMPTION



3 FULL LATER-DECISION COMPARISON

- 1 Estimate the marginal decision-delay distribution, correcting for initiation times observed only among survivor-selected, eligible initiators.
- 2 Balance measured covariates; assign each control a decision date (R*); retain those alive, observed, and clinically eligible at R*; then restart follow-up at R*.
- 3 Bootstrap the entire process, re-estimating nuisance quantities and re-imputing R* within each replicate.

Required data extend beyond dates: follow-up at R*, every decision-specific eligibility criterion, comparable post-R* outcomes, and adequate overlap. Progression-free status is E5508-specific—not a universal gate.

CLINICAL APPLICATION & DATA REQUIREMENTS

CLINICAL TAKEAWAY

Align the clinical decision—not just the calendar.

E5508: CONTROLLED INDUCTION–MAINTENANCE BENCHMARK

Element	E5508 design and role
Trial setting/design	Phase III advanced nonsquamous NSCLC; 874/1,516 without progression after induction were randomized to bevacizumab, pemetrexed, or both maintenance.
Benchmark question	Among maintenance-eligible patients: overall survival from randomization, pemetrexed versus bevacizumab.
Emulation	Single-arm n=279 pemetrexed initiators. Constructed control n=695 sampled from bevacizumab-maintenance and nonrandomized induction participants to induce measured prognostic imbalance; maintenance date unobserved; naïve OS indexed at registration.
Full reconstruction	Assign maintenance date; require alive, observed, and progression-free; re-index OS; propensity-score weight; repeat all steps in 1,000 bootstraps.

Induction: ≤4 carboplatin/paclitaxel/bevacizumab cycles. Later randomization provides a rare maintenance-gate benchmark; nonrandomization encoded failure of the eligibility process. Methodological test—not new effectiveness evidence.

WHAT FULL RECONSTRUCTION DOES — AND WHAT IT REQUIRES

What it does

- IDI assigns a comparable decision date
- rebuilds the alive/observed/eligible risk set
- restarts follow-up at that date
- propensity-score weighting addresses measured imbalance

Requires

- credible decision-delay information
- every eligibility criterion at R*
- comparable post-decision outcomes
- overlap, effective sample size, and sensitivity analyses

BEYOND E5508: THE GATE MUST MATCH THE DECISION

Later decision	Eligibility information needed at that decision
Maintenance	Disease control/no progression; tolerability; PS
Post-CRT consolidation	No progression; CRT recovery; treatment window
Local consolidation	Response/stability; lesion suitability; fitness
Biomarker adaptation	Valid timed result + current clinical eligibility

Perioperative trials randomized before neoadjuvant therapy estimate the full pathway; IDI applies only to a distinct later conditional question.

FIVE DATA-FITNESS QUESTIONS

- 1 Whole pathway or later conditional decision?
- 2 Is R* a credible analogue of the treated decision?
- 3 Can survival, observation, and every eligibility criterion be assessed at R*?
- 4 Are post-R* outcomes comparable and covariate overlap adequate?
- 5 Do effective sample size, balance, weights, and sensitivity analyses support interpretation?

IDI cannot recover decision-specific eligibility that was never measured at the later gate.

REFERENCES, ACKNOWLEDGEMENTS & DISCLOSURE

1. Le Coent Q, et al. Index Date Imputation for Survival Analysis in Externally Controlled Trials. 2026.
2. Le Coent Q, et al. Beyond Time-Zero Alignment: Risk-Set Reconstruction in Delayed-Initiation External-Control Survival Analyses. 2026.
3. Ramalingam SS, et al. J Clin Oncol. 2019;37:2360–2367.

Supported by the Oncology Center of Excellence, FDA/HHS as part of financial assistance award (U01FD005942). E5508 was conducted by ECOG-ACRIN with NCI support through NCTN/NCORP. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by FDA/HHS, or the U.S. Government.

Artificial Intelligence (AI) Use Disclaimer

In accordance with the standards and guidance set forth by the Accreditation Council for Continuing Medical Education (ACCME), individuals in control of content are required to disclose the use of artificial intelligence (AI) in the development of educational materials. This includes identification of the specific AI tool(s) utilized and a description of how AI contributed to content creation as a slide reference.

IASLC maintains oversight of all educational content and conducts a thorough review process to ensure that materials are evidence-based, fair-balanced, and free from commercial bias, consistent with ACCME Standards for Integrity and Independence.

While AI may be used as a supportive tool in content development, it does not replace clinical expertise or independent medical judgment. Learners are advised to critically evaluate all information presented and apply their own professional judgment when incorporating knowledge into clinical practice.

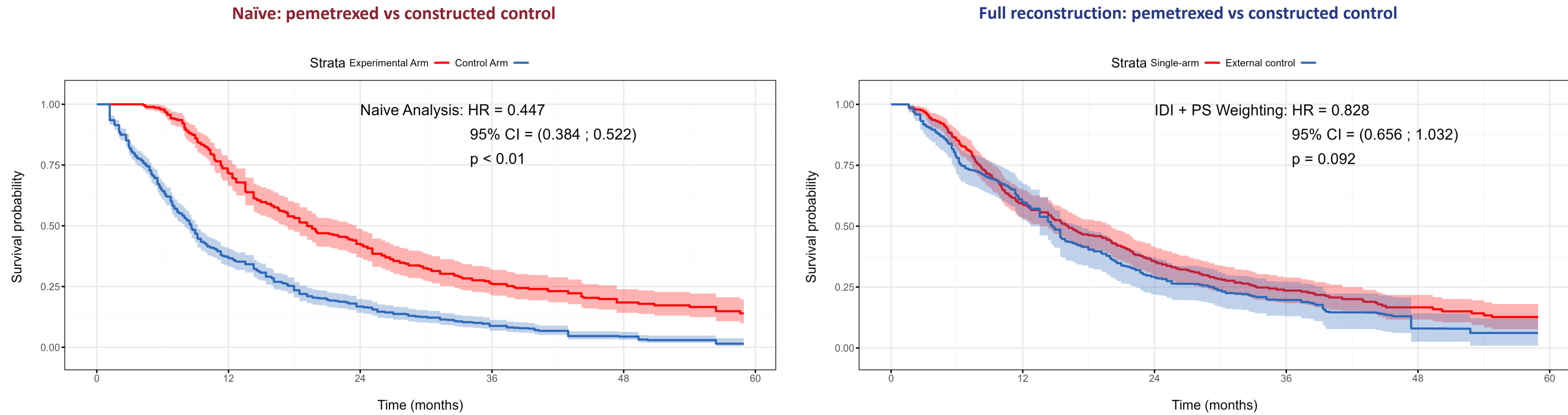
*All individuals responsible for content must disclose within the slides if AI was used in the development of the material and specify which AI tool was utilized.

OpenAI Codex was used for language refinement and poster layout; authors reviewed and approved all scientific content, results, and interpretations.

RESULTS

Constructed E5508 emulation: 36-month OS difference was an apparent +17.0 points with naïve registration indexing, +3.9 after full reconstruction, and +4.2 randomized. Weighting alone, missing eligibility, and poor overlap degraded performance.

RANDOMIZED BENCHMARK | Overall survival: pemetrexed vs bevacizumab | HR 0.860 (95% CI 0.719–1.027) | 36-month difference +4.2 points (–2.6 to +10.9)



Metric	Naïve	Full reconstruction	Randomized
HR (95% CI)	0.447 (0.384–0.522)	0.828 (0.656–1.032)	0.860 (0.719–1.027)
36-month difference*	+0.170 (0.113–0.226)	+0.039 (–0.050–0.127)	+0.042 (–0.026–0.109)

**Treated minus comparator (bevacizumab in the randomized benchmark; constructed external control in emulated analyses).*

MONTE CARLO SIMULATIONS

Method	Abs. log-HR bias	95% CI coverage
Naïve	0.486–0.740	0.000–0.001
Propensity-score weighting only	0.531–0.761	0.000–0.001
IDI + propensity-score weighting	0.006–0.038	0.943–0.952

10,000 datasets/scenario; n=200 treated/800 controls; 3 effects × 2 initiation mechanisms. Simulation assumptions apply.

CHECKLIST: IDI + PROPENSITY-SCORE WEIGHTING

Overlap + eligibility at R*	Abs. bias	95% CI coverage
Sufficient + progression available	0.029–0.058	0.927–0.941
Sufficient + progression unavailable	0.073–0.172	0.723–0.868
Poor + progression available	0.049–0.121	0.794–0.830
Poor + progression unavailable	0.167–0.319	0.588–0.697

10,000 datasets/scenario; n=200/800. Without progression, 31.3%–47.1% of retained controls failed eligibility; effective sample size 49–88 (sufficient overlap) vs 15–22 (poor).

E5508 RESAMPLING REPORT — 1,000 TRIAL-BASED EMULATIONS

Analysis	Abs. log-HR bias	95% CI coverage	Clinical lesson
Naïve baseline-indexed	0.438–0.473	0.001–0.002	Wrong clock and risk set
Propensity-score weighting only	0.238–0.309	0.213–0.370	Balance alone was insufficient
IDI + PS, progression unavailable	0.142–0.219	0.445–0.702	Eligibility remained incomplete
IDI + PS, progression-informed	0.006–0.094	0.796–0.894	Smallest discrepancy

Tipton overlap index 0.78–0.85. Bias and coverage reflect resampling performance, not single-dataset treatment effects.

≈3.1 months mean assigned maintenance-decision delay	68.4–68.6% of excluded controls in progression-informed IDI + PS analyses were excluded for progression	22.0–22.1% of controls retained without progression information would have failed maintenance eligibility	0.032–0.049 post-weighting mean absolute SMD 0.032–0.049 (from 0.280–0.318); ESS 132–135
--	---	---	--

CONCLUSIONS & CLINICAL IMPLICATIONS

In the constructed E5508 emulation, naïve registration indexing produced an apparent +17.0-percentage-point 36-month OS difference, versus +4.2 points in the randomized reference. Full reconstruction—decision time, eligibility, re-indexing, and weighting—yielded 3.9 points (HR 0.828 versus 0.860). Across simulations and trial-based resampling, weighting alone was insufficient; missing eligibility and poor overlap degraded performance. Full reconstruction substantially reduced bias, although trial-based 95% CI coverage remained below nominal.

Constructed methodological benchmark—not a new effectiveness analysis or real-world validation. Progression eligibility for nonrandomized participants was encoded; agreement concerns point estimates, and both reconstructed and randomized CIs include the null. Use beyond E5508 requires a credible decision-time analogue, complete gate-specific eligibility and follow-up, comparable outcomes, and adequate overlap.

BOTTOM LINE

For maintenance, consolidation, or adaptation, verify comparable preceding care and that controls could enter the same clinical gate under the same eligibility rules, with survival follow-up starting there. If not, change the estimand or control source—or present the comparison as descriptive.