Keiji AI

EVIDENCE REVIEW · ONCOLOGY REGULATORY SCIENCE

Strong PFS, uncertain survival: what the ivonescimab CRL teaches

A systematic review and meta-analysis of PD-(L)1-based therapy after EGFR-TKI failure, and what a different trial design could have changed.

By Jimeng Sun · 15 November 2026 · Keiji AI evidence review

HR 0.79
HARMONi overall survival; prespecified test missed (p = 0.057)

~48%
Power of 262 deaths to detect that effect

HR 0.86
Pooled survival effect across five randomized trials

15 min
Search to draft review with TrialMind SLR

On 15 November, Summit Therapeutics announced that FDA had issued a complete response letter for ivonescimab plus chemotherapy in EGFR-mutated advanced non-small-cell lung cancer that progressed on EGFR tyrosine kinase inhibitors. By Summit's account, FDA judged that the progression-free survival benefit did not establish a favorable benefit–risk balance without a statistically significant improvement in overall survival. [1]

The decision surprised many observers, because ivonescimab had beaten pembrolizumab head to head and had reported survival benefits in other trials. To see why FDA was not persuaded, I used TrialMind's systematic literature review agent to collect every randomized trial of ivonescimab and every randomized trial of PD-(L)1-based therapy added to chemotherapy

after EGFR-TKI failure, and then pooled them. The search, screening, extraction and first draft took about 15 minutes. The pooled evidence makes FDA's position easier to understand.

Six lessons

1 **A win in one population does not carry over to another.**

The headline win over pembrolizumab came from HARMONi-2: first-line treatment, in China, in patients without EGFR mutations. The US application covered a different group, patients with EGFR mutations whose cancer had progressed on targeted therapy, and rested on a different trial, HARMONi. FDA judged the application on HARMONi alone. [2,4]

2 **In this setting, a PFS gain is not a down payment on survival.**

Across five randomized trials, regimens that add VEGF blockade cut the hazard of progression by about half (pooled HR 0.54), yet their pooled survival effect (HR 0.89) is no better than that of PD-(L)1 inhibitors alone (HR 0.86). FDA's insistence on overall survival has an empirical basis.

3 **The survival analysis was sized for an effect this setting has never shown.**

HARMONi's prespecified OS analysis narrowly missed (HR 0.79, p=0.057). Its 262 deaths gave good power only if the drug cut the risk of death by about 30%; for the effect the trial actually observed, power was close to a coin flip. Later analyses with longer follow-up looked better (HR 0.76, p=0.015), but they were not in the testing plan, so their p-values are only nominal. See "What a different design could have changed" below. [2,6,7,16]

4 **Count patients, not trials.**

HARMONi-A, the Chinese trial with a significant OS result, shares patients with HARMONi. Pooling the two as independent studies gives a convincing-looking HR of 0.76 (p=0.003) built partly on the same people counted twice. [3,8]

5 **A global trial needs global follow-up.**

The Western subgroup was small and immature (HR 0.76, 95% CI 0.52–1.10), and the later global estimates combined updated Western data with an older Asian cutoff. [6]

6 **An earlier approval does not set the bar for the next drug.**

In 2024 FDA approved amivantamab plus chemotherapy for the same patients without a significant survival result, so FDA does not always require one. Each application is weighed on its own PFS effect, toxicity and statistical plan, and ivonescimab's balance fell short where amivantamab's did not. [9,10]

Literature review: the ivonescimab program and the post-TKI landscape

Ivonescimab is a bispecific antibody that blocks PD-1 and VEGF in one molecule. Its randomized program in NSCLC answers four different clinical questions, and keeping them separate is the first step in reading the CRL correctly.

Trial	Setting	Comparison	Key result	Bearing on the US application
HARMONi [2]	Post third-generation EGFR-TKI; global	+ chemotherapy vs placebo + chemotherapy	PFS HR 0.52; OS HR 0.79 (0.62–1.01), test missed	**The pivotal trial**
HARMONi-A [8,22]	Post EGFR-TKI; China	+ chemotherapy vs placebo + chemotherapy	PFS HR 0.46; final OS HR 0.74 (0.58–0.95)	Same question, but overlapping patients
HARMONi-2 [4,5]	First line, PD-L1-positive, no EGFR/ALK; China	Monotherapy vs pembrolizumab	PFS HR 0.51; interim OS HR 0.73	Different population
HARMONi-6 [18,23]	First line, squamous; China	+ chemotherapy vs tislelizumab + chemotherapy	PFS HR 0.60; interim OS HR 0.66	Different population

The program's strongest survival results come from trials in China in other populations. HARMONi, the only trial in the application's population with substantial Western enrollment, did not meet its OS test.

The surrounding literature shows that the post-TKI setting has resisted immunotherapy. KEYNOTE-789 (pembrolizumab) and CheckMate 722 (nivolumab) both missed their primary endpoints. [11,12] Regimens that added VEGF blockade, such as ORIENT-31 (sintilimab plus a bevacizumab biosimilar) and ATTLAS (atezolizumab plus bevacizumab), produced large PFS effects with neutral survival. [13,14,24] Ivonescimab's HARMONi result fits this pattern.

The trials below are the ones pooled in the meta-analysis. Each tested a different drug or combination, so the figure labels each row by drug first and trial second. Three regimens block both the PD-1/PD-L1 checkpoint and VEGF, but by different routes: ivonescimab does it with one bispecific molecule, the others with two separate antibodies.

Trial	Drugs added to chemotherapy	Mechanism	Status in this setting
HARMONi [2]	Ivonescimab (Summit/Akeso)	One bispecific antibody: PD-1 + VEGF	US application rejected (this CRL)
ORIENT-31 [13,24]	Sintilimab + IBI305 (Innovent)	Two antibodies: PD-1 + VEGF (bevacizumab biosimilar)	Approved in China; not in the US
ATTLAS [14]	Atezolizumab + bevacizumab	Two antibodies: PD-L1 + VEGF	Korean academic trial; drugs marketed for other uses
ORIENT-31 [13]	Sintilimab	PD-1 only	Second experimental arm, same trial
KEYNOTE-789 [11]	Pembrolizumab	PD-1 only	Missed PFS and OS endpoints
CheckMate 722 [12]	Nivolumab	PD-1 only	Missed primary PFS endpoint

Grouping by mechanism is a simplification: the drugs differ (PD-1 versus PD-L1, bispecific versus combination), and ATTLAS used carboplatin–paclitaxel rather than platinum–pemetrexed.

Systematic review and meta-analysis

Question. In adults with EGFR-mutated advanced NSCLC after EGFR-TKI progression, does adding PD-(L)1-based therapy, with or without VEGF blockade, to platinum chemotherapy improve PFS and OS compared with chemotherapy alone?

Search and selection. I ran the review in a TrialMind SLR project on 15 November 2026. TrialMind's Auto Mode searched PubMed, Europe PMC, ClinicalTrials.gov, medRxiv and the AACR, ASCO and ESMO meeting libraries (clinical trials from 2015 onward), screened 19 records against the PICO question and included 6 reports. From search to draft report took about 15 minutes. Two targeted PubMed searches checked completeness: one for every ivonescimab trial report (8 records) and one for PD-(L)1 trials after EGFR-TKI failure (19 records). I reconciled the three result sets by hand and grouped reports by trial, so that initial and updated analyses of the same trial count once. Eligible studies were randomized phase 3 trials comparing PD-(L)1-based therapy plus chemotherapy with chemotherapy alone in EGFR-TKI-pretreated, EGFR-mutated NSCLC. Single-arm and non-comparative studies were excluded, as was IMpower151, whose population was only 53% EGFR-mutated. [19]

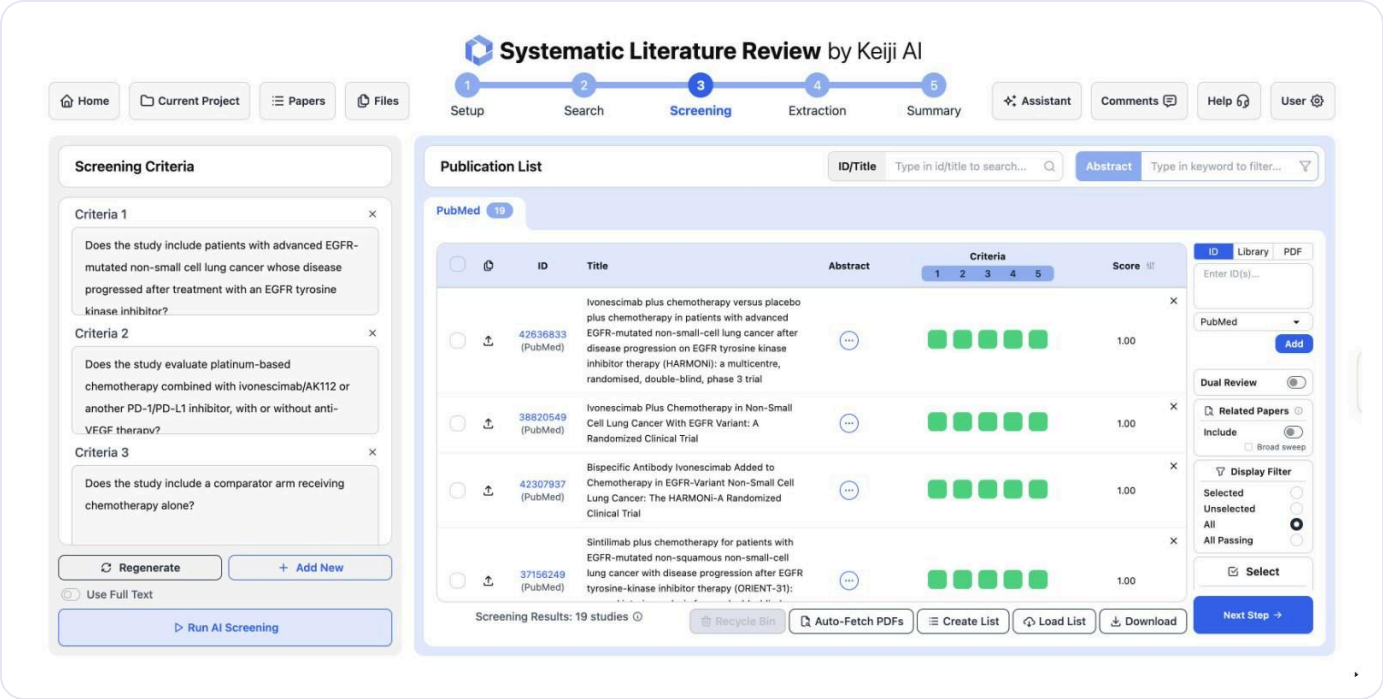


FIGURE 1 Screening in TrialMind SLR: each record is scored against every eligibility criterion, so a reviewer can audit any decision.

Included evidence. Five trials with six comparisons: HARMONi, KEYNOTE-789, CheckMate 722, ORIENT-31 (two experimental arms sharing one control arm) and ATTLAS. HARMONi-A was eligible but was not pooled, because its China-enrolled patients overlap with HARMONi's. It was used only in a sensitivity analysis that replaces HARMONi. Where a trial reported several data cuts, I used the latest peer-reviewed analysis.

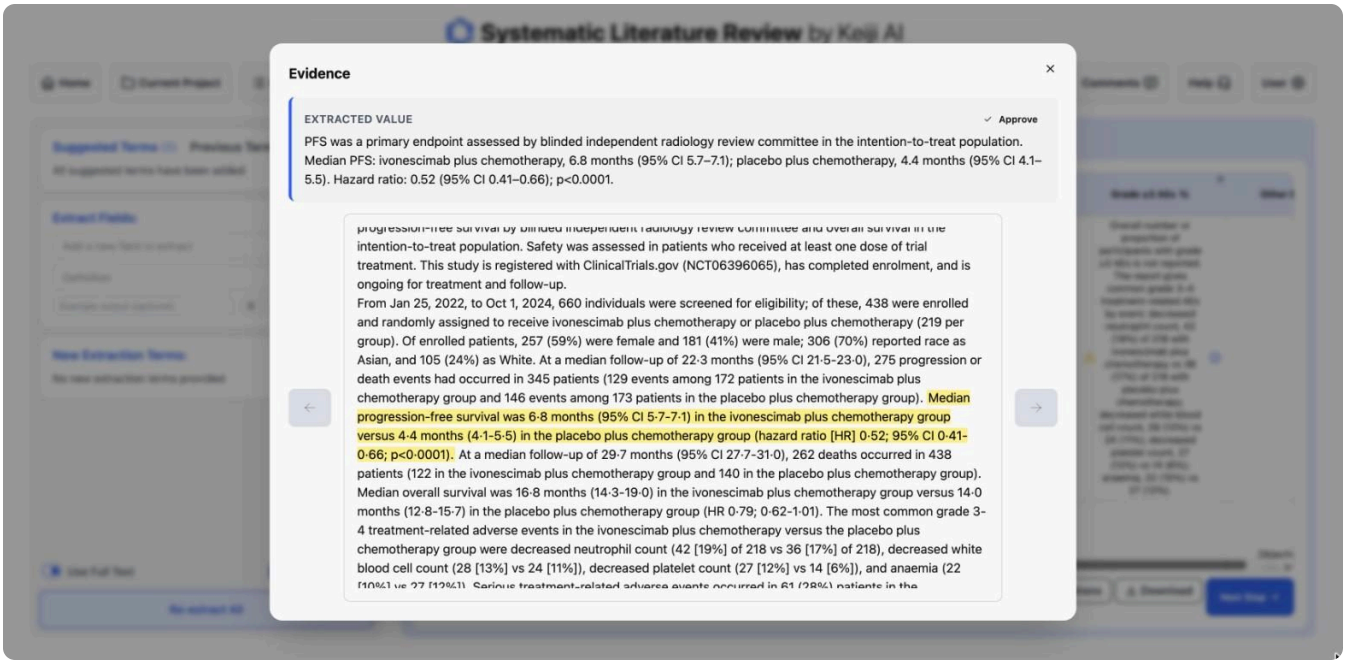


FIGURE 2 Extraction: every value, such as HARMONi's PFS hazard ratio, links to the sentence in the paper it came from.

Analysis. Log hazard ratios were pooled with a DerSimonian–Laird random-effects model. Each trial contributes one comparison to the overall estimate. Subgroups compare regimens with and without VEGF blockade.

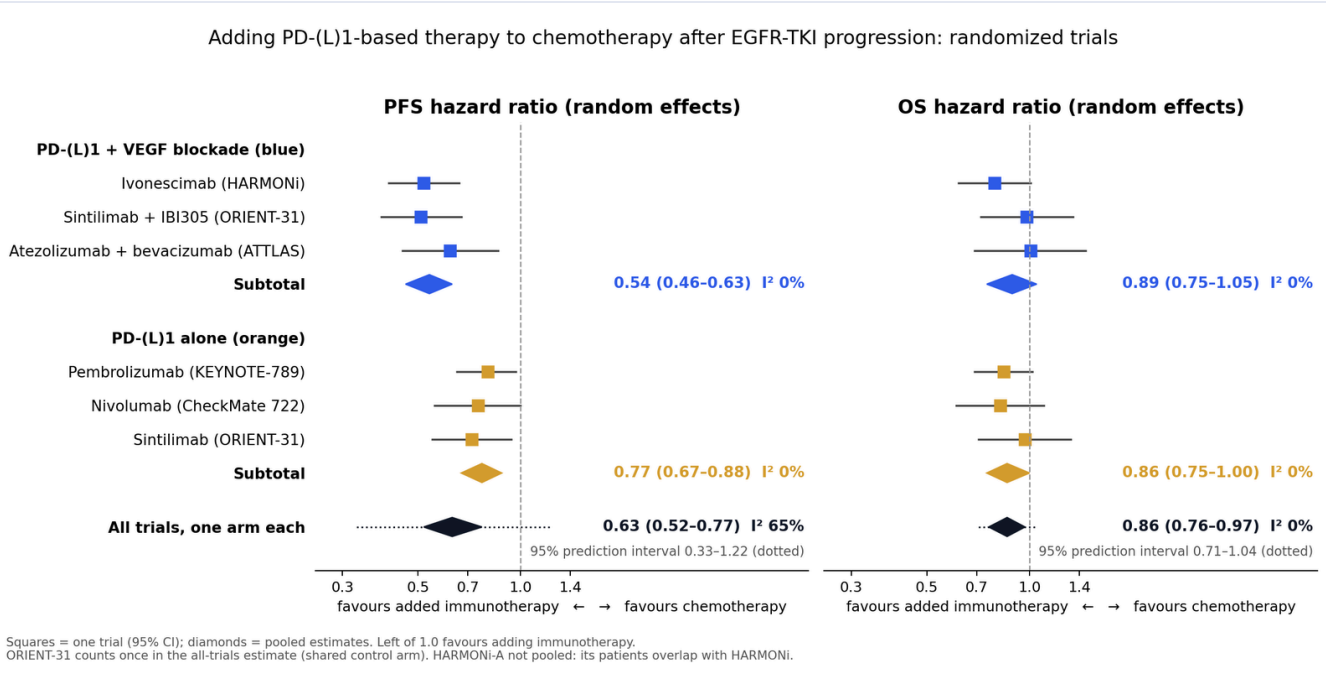


FIGURE 3 Random-effects meta-analysis. Each square is one trial's hazard ratio with its 95% CI; each diamond is a pooled estimate; values left of 1.0 favour adding immunotherapy. Blue rows block PD-(L)1 and VEGF; orange rows block PD-(L)1 only. The dotted line on the bottom row is the 95% prediction interval for a new trial.

Figure 3 shows the main analysis. The table below shows the extra checks that are not in the figure.

Additional analysis	Trials	PFS HR (95% CI)	OS HR (95% CI)
HARMONi-A in place of HARMONi	5	0.62 (0.50–0.77)	0.85 (0.75–0.95)
Excluding ATTLAS (open label; some ALK patients)	4	0.63 (0.50–0.81)	0.84 (0.75–0.96)
Incorrect: HARMONi and HARMONi-A pooled as if independent	2	0.50 (0.41–0.60)	0.76 (0.64–0.91)

The sensitivity checks barely move the estimates, so the conclusions do not depend on any single trial.

Findings. Three results stand out.

1. A bigger PFS effect does not bring a bigger survival effect. Compare the two subtotal diamonds in Figure 3. In the PFS panel (left), the blue regimens with VEGF blockade pool to HR 0.54, clearly stronger than the orange PD-(L)1-only regimens at 0.77. In the OS panel (right), the two diamonds overlap: 0.89 for blue and 0.86 for orange. The I^2 values tell the same story. Pooling all trials for PFS gives I^2 65%, meaning the trials disagree; within each colour group I^2 is 0%, so the VEGF split accounts for the disagreement. For OS, I^2 is 0% overall: the trials agree. Adding VEGF blockade improves progression but not survival.

2. The class has a modest survival benefit, and HARMONi looks typical. In the OS panel, every trial's line crosses 1.0: no single trial proved a survival benefit by itself. Pooled, the black diamond (HR 0.86, 0.76–0.97) stays left of 1.0, which suggests a modest benefit for the class. The dotted prediction interval (0.71–1.04), the range in which a new trial's true effect is likely to fall, still crosses 1.0. HARMONi's square at 0.79 sits close to the pooled diamond: consistent with the class, not clearly better.

3. Counting the same patients twice manufactures certainty. The last row of the table above pools HARMONi with HARMONi-A as if they were separate studies and gets OS HR 0.76 (0.64–0.91, p=0.003), which looks like a confirmed survival benefit. But HARMONi's Chinese patients were also enrolled in HARMONi-A, so part of that precision comes from counting them twice. Deduplicating by title or DOI would not catch this; it takes reading the trial documents. That is why HARMONi-A is left out of Figure 3.

Limitations. This is a rapid review by one reviewer, with machine-assisted screening and extraction and no formal risk-of-bias adjudication. Pooling different drugs, backbones and geographies estimates an average class effect; it cannot rank regimens or substitute for a drug's own confirmatory trial. OS in ORIENT-31 is preliminary and confounded by crossover. ATTLAS was open label and included 13 patients with ALK rearrangement. Five trials give limited power to detect subgroup differences, so the VEGF comparison is descriptive.

What FDA's decision says about evidence

FDA guidance accepts PFS as a basis for approval when the effect size, disease context and benefit–risk support it. [10] The question is why it was not enough here. The meta-analysis offers part of an answer. When a regimen's PFS effect is large but the class record shows that such effects do not translate into survival, the PFS result carries less weight as a surrogate, and toxicity weighs more. HARMONi compared two groups of 218 treated patients each. Both received the same chemotherapy (carboplatin and pemetrexed); one group also received ivonescimab, the other a placebo. Severe (grade 3 or higher) treatment-related adverse events occurred in 109 patients (50.0%) in the ivonescimab group and 92 (42.2%) in the placebo group. Serious treatment-related adverse events, those leading to hospitalization, disability or death, occurred in 61 (28.0%) and 33 (15.1%). Treatment-related deaths were similar: 4 (1.8%) and 5 (2.3%). [2,15]

The amivantamab precedent shows the judgment is case by case. Its approval rested on PFS HR 0.48 with OS not statistically significant at the interim analysis. [9] Different molecules, toxicities and testing plans lead to different conclusions. The full letter and review documents are needed to know which factors decided this application.

For development teams, the practical lessons are concrete:

- **Keep companion trials' populations separate,** and document patient overlap from the start.
- **Use the class record to set assumptions.** A pooled estimate like the one above is the natural planning input for the next trial's survival analysis.

What a different design could have changed

Was HARMONi badly designed? Its PFS analysis succeeded, so the design was not a failure in general. The weak point was the survival analysis, and it is easy to quantify. Using the standard Schoenfeld approximation for a two-sided 5% test, the table shows how much power 262 deaths gave for different true effects, and how many deaths 80% power would have required. If the OS analysis was allocated less than the full 5%, as is likely under a split between endpoints, power was lower still.

True OS hazard ratio	Reduction in risk of death	Power with 262 deaths	Deaths needed for 80% power
0.70	30%	82%	about 250
0.76 (later HARMONi estimate)	24%	60%	about 420
0.79 (primary HARMONi estimate)	21%	48%	about 570
0.86 (class average, this review)	14%	23%	about 1,380

I checked these numbers with TrialMind's Data Science Assistant, asking it in plain language for the power calculation, an alternative design and a simulation to validate both. It wrote and ran the analysis in Python, and its results matched the table: 47.9% power at HR 0.79 and 23.0% at HR 0.86 with 262 deaths. In 5,000 simulated trials per scenario the fixed design achieved 46.5% and 23.8%, close to the formula.

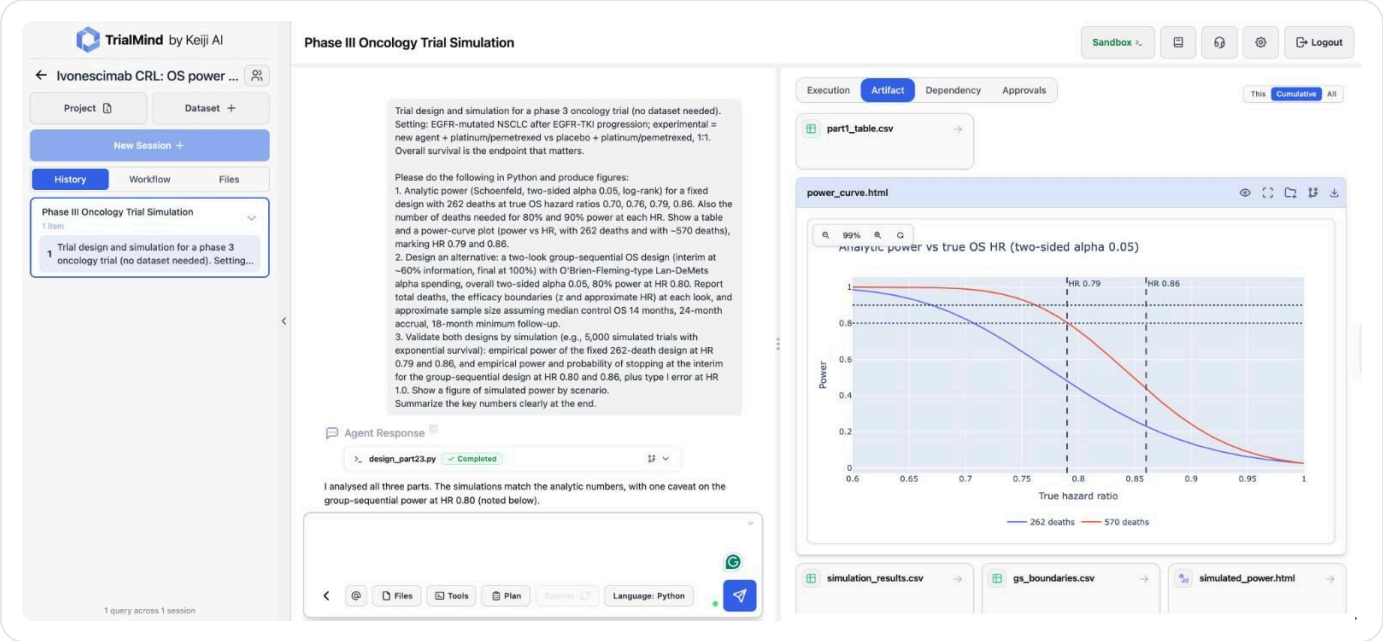


FIGURE 4 Trial design in TrialMind's Data Science Assistant: the request in plain language (left) and the resulting power curves (right). With 262 deaths (blue), power falls below 50% at HR 0.79; about 570 deaths (red) would have given 80%.

The assistant also designed the alternative described below: a two-look group-sequential OS design with an interim analysis at 60% of events and O'Brien–Fleming-type alpha spending, powered at 80% for HR 0.80. It needs about 641 deaths and roughly 890 patients (assuming median control survival of 14 months, 24 months of accrual and 18 months of minimum follow-up), only about 2% more deaths than a fixed design with the same power. In simulation it kept the false-positive rate at the planned level and gave a 36% chance of stopping early for efficacy if the true HR is 0.80.

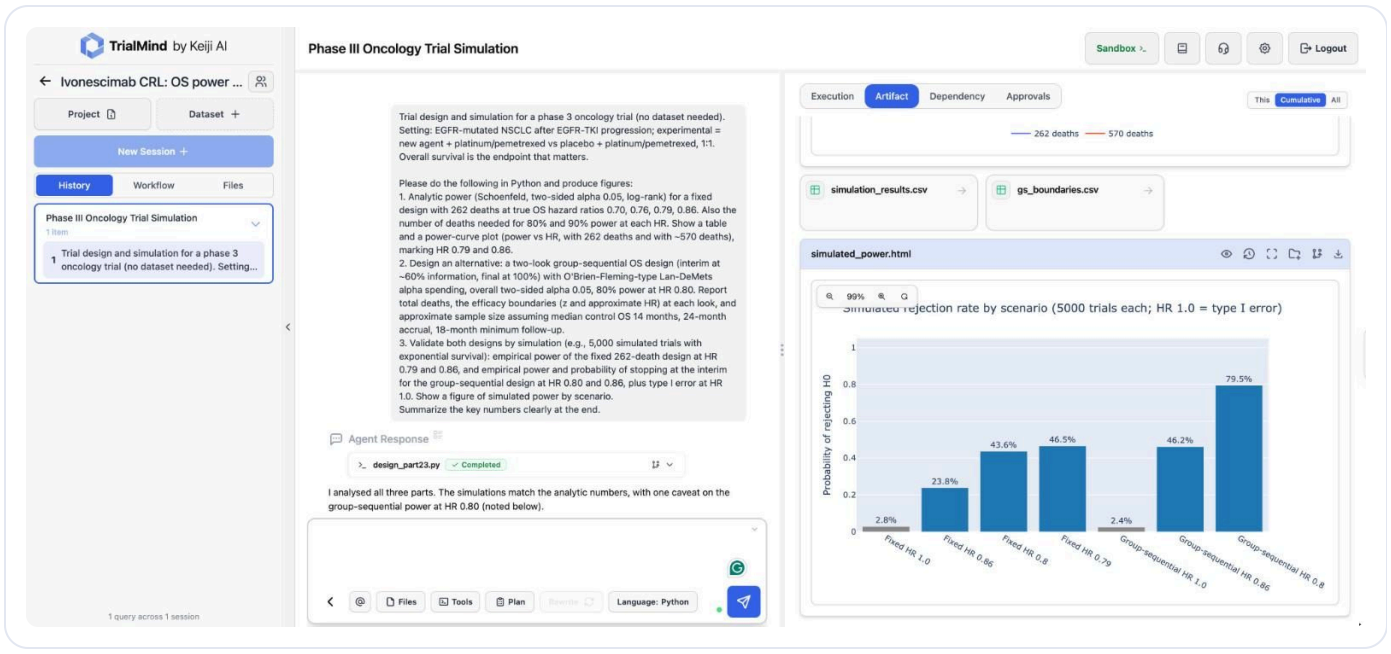


FIGURE 5 Simulated operating characteristics from the same session: probability of a positive OS result for the 262-death fixed design and the 641-death group-sequential design at different true hazard ratios (5,000 trials each; HR 1.0 shows the false-positive rate).

Design work like this, which used to take days of statistician time, took minutes. Running it before enrollment is the point: the simulation shows how likely a trial is to succeed under realistic effects, while the design can still change.

In other words, the survival analysis was large enough only for an effect of about HR 0.70, bigger than any PD-(L)1 regimen has shown after EGFR-TKI failure. PFS effects in this setting are large (HR about 0.5), survival effects much smaller, and a design driven by PFS inherits a survival analysis that is likely to miss.

With hindsight, five choices would have changed the odds:

1. **Size OS for a realistic effect.** Plan on HR 0.80–0.86, the range the class record supports, which means more patients, longer follow-up or both.

2. **Pre-plan the later looks.** A group-sequential design with an interim and a final OS analysis, and alpha spent across them, would have let a longer-follow-up analysis count as confirmatory instead of nominal.

3. **Enroll regions in step.** HARMONi enrolled from January 2022 to October 2024, but not evenly. Enrollment in Asia closed in November 2022; in North America and Europe it continued until October 2024, almost two years later. [15] At the primary analysis, median follow-up in the Western subgroup was only 9.2 months, against roughly 30 months for the trial overall, so Western patients had contributed few deaths. Their OS estimate was correspondingly imprecise. Even after follow-up grew to 23.2 months, it stood at HR 0.76 with a 95% CI of 0.52–1.10, too wide to show that the benefit holds for the US patients FDA must protect. [6]

The later global analyses then combined updated Western data with Asian data frozen at an April 2025 cutoff, which appears to match the final cutoff of HARMONi-A. [6,8] Mixed cutoffs create two problems. Events after April 2025 among Asian patients are left out, so the pooled estimate covers different follow-up periods by region. And choosing which region's data to update after seeing results is itself a data-dependent decision, which is exactly what a prespecified analysis is meant to prevent.

A design that avoids both would open Western sites at the start, set regional enrollment targets so no region trails far behind, require a minimum follow-up in every region before the OS analysis, and apply one data cutoff to all patients. It would also prespecify how regional consistency will be judged, as ICH E17 recommends for multiregional trials, and keep the global trial's patients separate from any companion national trial. [25]

4. **Let the regulatory bar set the sample size.** Summit disclosed that FDA had said significant OS was needed. When OS is the bar, OS, not PFS, should drive the design.
5. **Define how subsequent therapy is handled.** Later treatments dilute survival differences. A prespecified estimand under ICH E9(R1), with sensitivity analyses, makes the OS result interpretable. [17]

There is a real trade-off. A trial powered for HR 0.86 needs roughly 1,400 deaths, which is slow and expensive and may not be worth running. A faster trial can be a rational bet. The lesson is that the bet should be explicit and quantified before enrollment, not discovered at the readout. We have not seen HARMONi's statistical analysis plan, so the effect size it was designed for is inferred from the number of deaths, not quoted.

What this means for AI in clinical research

A good review of this case has to keep HARMONi and HARMONi-2 apart, notice the HARMONi-A overlap, keep the word "nominal" attached to a p-value, and find the amivantamab counterexample to its own first explanation. TrialMind's review agent found the relevant trials, extracted their results and drafted a review in about 15 minutes, work that usually takes a team weeks. Deciding not to pool HARMONi-A, and explaining why, still took a person who read the trial documents. That division of labor is the point: AI makes the evidence base complete and traceable quickly, and expert judgment decides what it means.

Disclosure: The author is co-founder and CEO of Keiji AI, which develops TrialMind.

Evidence notes and sources –

Evidence notes

Regulatory document status. The full CRL was not available. The primary disclosure is Summit's announcement, which must not be described as the FDA letter. FDA's public CRL database had no matching letter when checked on 15 November 2026. [1,20] Established from the sponsor's disclosure: the application was not approved in its current form, and FDA found the evidence insufficient without significant OS. Not established: every deficiency in the letter, FDA's full reasoning, or the path to resubmission.

Data used in the meta-analysis.

Trial	Comparison (N)	PFS HR (95% CI)	OS HR (95% CI)	Source
HARMONi	Ivonescimab + chemo vs placebo + chemo (438; PFS primary 345)	0.52 (0.41–0.66)	0.79 (0.62–1.01)	[2]
HARMONi-A	Ivonescimab + chemo vs placebo + chemo (322)	0.46 (0.34–0.62)	0.74 (0.58–0.95)	[8,22]
KEYNOTE-789	Pembrolizumab + chemo vs placebo + chemo (492)	0.80 (0.65–0.97)	0.84 (0.69–1.02)	[11]
CheckMate 722	Nivolumab + chemo vs chemo (294)	0.75 (0.56–1.00)	0.82 (0.61–1.10)	[12]
ORIENT-31	Sintilimab + IBI305 + chemo vs chemo (476, three arms)	0.51 (0.39–0.67)	0.98 (0.72–1.34)	[13]
ORIENT-31	Sintilimab + chemo vs chemo	0.72 (0.55–0.94)	0.97 (0.71–1.32)	[13]
ATTLAS	Atezolizumab + bevacizumab + chemo vs chemo (228)	0.62 (0.45–0.86)	1.01 (0.69–1.46)	[14]

Checks that changed the interpretation.

- *Patient overlap:* Summit's SEC filing states that China-enrolled HARMONi patients also entered HARMONi-A. [3]

- *Endpoint testing:* HARMONi allocated alpha separately to PFS and OS, with recycling from a successful PFS test to OS. [3]
- *Denominators:* the publication reports 438 randomized; the registry field reads 420; the primary PFS analysis used 345. These are flagged, not harmonized. [2,21]
- *Statistical language:* a hazard ratio is a relative measure, not a survival extension for an individual patient, and non-significance is not evidence of no effect.

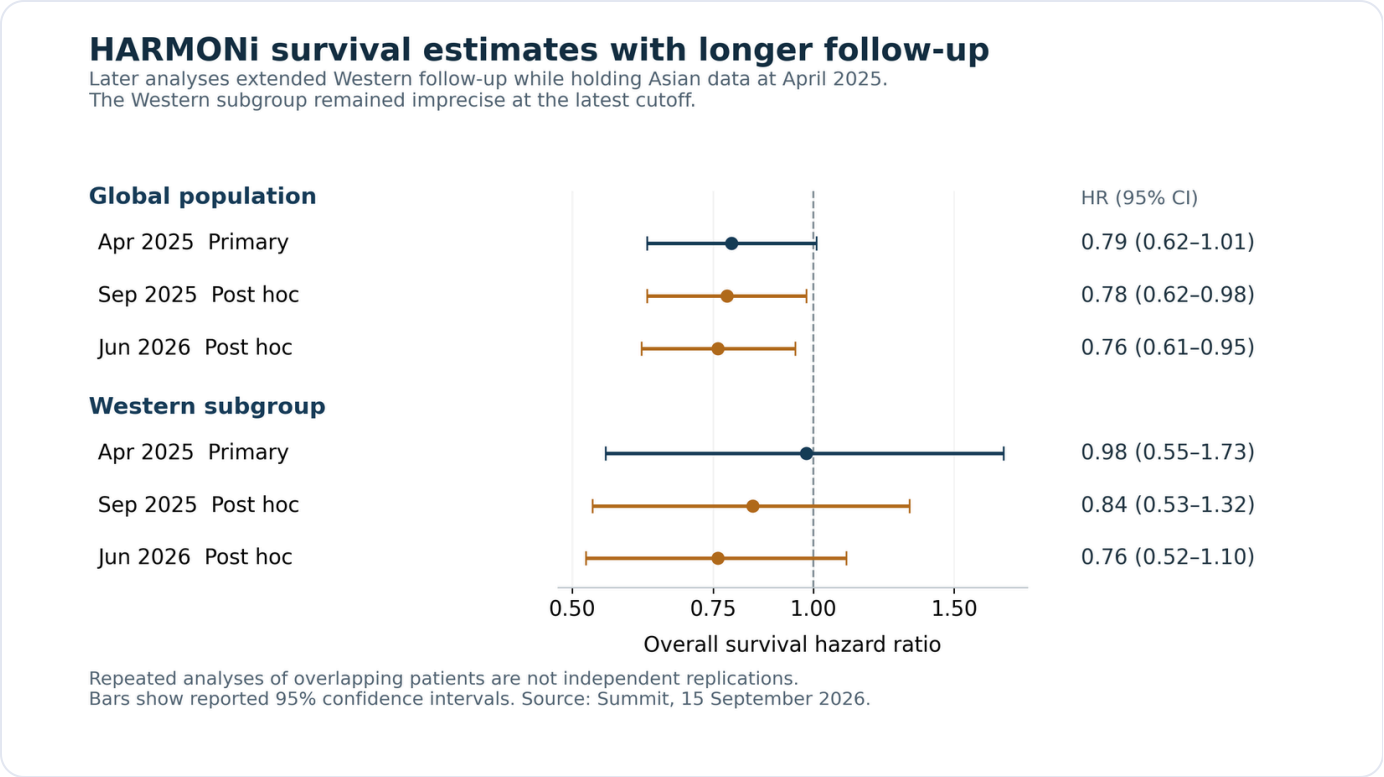


FIGURE 6 Successive HARMONi OS estimates. Later point estimates are encouraging, but their inferential status differs from the primary analysis. [6]

Sources

[1] Summit announcement of the ivonescimab CRL. 15 November 2026. Sponsor disclosure, not the FDA letter.

[2] HARMONi global phase 3 results. Le and colleagues. Lancet Oncology 2026;27:1094–1109. PMID 42636833.

[3] Summit third quarter 2026 Form 10-Q. Patient overlap, endpoint testing and prior FDA feedback.

[4] HARMONi-2 primary publication. Xiong and colleagues. Lancet 2025. PMID 40057343.

[5] HARMONi-2 interim survival disclosure. September 2026.

[6] HARMONi survival follow-up disclosure. 15 September 2026.

[7] ICH E9 statistical principles. FDA guidance.

[8] HARMONi-A final survival publication. JAMA 2026;336:306–314. PMID 42307937.

[9] FDA approval of amivantamab with chemotherapy. 19 September 2024.

[10] FDA guidance on oncology endpoints. December 2018.

[11] KEYNOTE-789 phase 3 results. Yang and colleagues. JCO 2024. PMID 39173098.

[12] CheckMate 722 phase 3 results. Mok and colleagues. JCO 2024. PMID 38252907; OS from full text (PMC11095864).

[13] ORIENT-31 second interim results. Lu and colleagues. Lancet Respiratory Medicine 2023. PMID 37156249.

[14] ATTLAS phase 3 results. Park and colleagues. JCO 2024. PMID 37861993.

[15] HARMONi WCLC 2025 presentation.

[16] FDA guidance on multiple endpoints. October 2022.

[17] ICH E9(R1) estimands and sensitivity analysis. May 2021.

[18] HARMONi-6 interim survival publication. Lu and colleagues. Lancet 2026. PMID 42218899.

[19] IMpower151 phase 3 results. Zhou and colleagues. Nature Medicine 2025. PMID 40379995.

[20] FDA public CRL collection. Checked 15 November 2026.

[21] HARMONi registry record. NCT06396065.

[22] HARMONi-A initial results. Fang and colleagues. JAMA 2024. PMID 38820549.

[23] HARMONi-6 primary publication. Chen and colleagues. Lancet 2025. PMID 41125109.

[24] ORIENT-31 first interim results. Lu and colleagues. Lancet Oncology 2022. PMID 35908558.

[25] ICH E17: planning and design of multi-regional clinical trials. FDA guidance, July 2018.