CRA PORTFOLIO

KEYNOTE-991

Merck | Pembrolizumab + Enzalutamide | mCRPC

Phase III | ~1,200 Patients | Discontinued March 2023

The monitoring architecture was designed for a different kind of drug.
Here is the system I would have built instead.

Martha Wuya Koroma

February 2026

Contents

01 — The Verdict

KEYNOTE-991 did not fail because the science was absurd. It failed because the trial was built for a different kind of drug.

The monitoring architecture system was designed to ask "is this working?" and was calibrated for therapies that show their hand early. Pembrolizumab does not show its hand early. It never has. And the people who designed this trial knew that. Which is what makes the structural decision so interesting to me.

My argument is not that the therapy should have worked. My argument is that the governance system could not see clearly enough, early enough, to distinguish between *the treatment failing* and *the treatment not having finished yet*. Those are not the same thing. And in immunotherapy, confusing them is expensive.

> *You might be wondering why I chose this trial specifically.*

Because the failure is not obvious from the outside. The press release says "did not meet primary endpoints." Standard pharmaceutical termination language. Nothing to see. But when you read the architecture, the endpoint structure, the futility boundary design, the safety monitoring logic, you will find something more interesting than a drug that didn't work. You will find a system that was structurally incapable of asking the right questions at the right time

02 — What Happened

Pembrolizumab (Keytruda) is an immune checkpoint inhibitor. It blocks the PD-1 pathway which is the biological mechanism tumours exploit to hide from the immune system. Remove the brake, the immune system attacks. That logic had worked dramatically in melanoma, some lung cancers, MSI-high tumours. Cancers that were immunologically *present,* inflamed, mutation-heavy, full of T-cell activity that was suppressed but not absent.

Prostate cancer is different. It is immunologically cold. Fewer T-cell infiltrates. Less immune activation. A more suppressive tumour microenvironment. The immune system, in many mCRPC cases, had not been suppressed. It simply had not entered the room.

Merck's hypothesis was that enzalutamide might alter the tumour environment enough to make it more receptive to pembrolizumab. Weaken the tumour signalling, change the inflammatory conditions, possibly warm a cold tumour. There were preclinical signals and early clinical observations suggesting possible synergy. **Keyword:** *possible.*

March 2023. First interim analysis. The Data Monitoring Committee reviewed the data and recommended immediate discontinuation for futility. Both primary endpoints favoured placebo.

Not neutrality — placebo.

Endpoint	Pembro + Enza	Placebo + Enza	Hazard Ratio

rPFS	10.9 months	12.2 months	1.20
OS	Not reached	Not reached	1.16
Time to PSA progression	10.9 months	12.2 months	1.20
Objective response rate	27.6%	23.5%	—

Safety Endpoint	Pembrolizumab Arm	Placebo Arm
Grade ≥3 Adverse Events (any)	61.9%	38.1%
Treatment discontinuation	33.4%	8.2%
Treatment-related deaths	8 patients	2 patients

The trial stopped. No second interim. No extended follow-up. No conditional continuation.

> *You might be thinking: the data looked bad, stopping seems right.*
>
> **Yes. And I am not arguing the DMC was wrong to follow their charter. I am arguing the charter was written for the wrong drug.**

03 Where It Broke

This is where we stop reading the press release and start reading the architecture.

The fracture runs through two places. They look like separate problems. They are the same problem expressed twice.

The problem is time.

Specifically: the trial's governance structure did not understand that immunotherapy lives on a different temporal logic than the drugs the system was built to evaluate.

Crack 2. The Futility Boundary That Could Not See Around the Corner

The DMC stopped the trial at first interim because conditional power had fallen below 20%. Conditional power is the calculated probability that the trial will eventually succeed given what the data looks like now. Below 20%: the mathematics say continue is irrational. Stop.

Mathematically correct. Structurally blind.

Standard conditional power calculations assume constant hazard, that the rate at which the treatment effect emerges is stable across time. For cytotoxic chemotherapy, that assumption is reasonable. The drug works or it doesn't, and it shows you relatively quickly.

Pembrolizumab does not work that way. Immunotherapy survival curves show *late separation* the tail of the curve, where benefit emerges months after the median, is where checkpoint inhibitors often prove themselves. This is well-documented in melanoma, lung cancer, and renal cell carcinoma. It is not a secret. It is the biological signature of how this class of drug behaves.

By stopping at first interim, approximately 50% of events, the DMC assumed the second half of the trial would mirror the first half. In a disease where immunotherapy might require substantially longer follow-up to show OS benefit, that assumption is structurally hostile to the mechanism being tested.

The futility boundary was designed for a drug that shows its hand early. Pembrolizumab does not show its hand early. The system could not see around the temporal corner.

You might be wondering whether I am arguing the trial should have continued.

No. I am arguing the system should have been designed with enough flexibility to *know the difference* between early futility and early ambiguity. Those are not the same. And the charter had no vocabulary for that distinction.

If humans are going to hope anyway, and in large Phase III oncology trials, they will; then the structure itself must account for hope. Not by removing the brakes. By designing brakes that understand what they are stopping.

Crack 4. Safety Monitoring That Forgot What Safety Was For

Grade ≥3 adverse events: 61.9%. Treatment discontinuation: 33.4%.

> ***How are you claiming the treatment failed when a third of the patients couldn't stay on it long enough to fairly test it?***

A third of the pembrolizumab arm exited before they had meaningful exposure. In immunotherapy, where response correlates with duration of exposure, where the mechanism requires time to activate, discontinuation is not a tolerability issue. It is an efficacy killer wearing a tolerability mask.

The monitoring plan saw safety. It did not see safety-as-structural-efficacy.

The system recorded adverse events, serious adverse events, tolerability rates. Standard practice. But it did not ask the operational question underneath: *are safety events removing patients from the efficacy denominator before efficacy can emerge?*

If the answer is yes, and at 33.4% discontinuation, the answer was trending yes, that is not just a patient welfare concern. It is a signal that the trial's ability to detect a real effect is being actively undermined in real time.

There is a second layer the monitoring architecture missed entirely: the rechallenge question. When a patient discontinued due to an immune-related adverse event, the protocol treated that as a closed chapter. Patient exits. Data is censored. Move on.

But Simonaggio et al. (JAMA Oncology, 2019) had already demonstrated, in 93 patients who experienced grade ≥2 irAEs and were rechallenged with the same checkpoint inhibitor, that rechallenge resulted in a second adverse event in 55% of patients, with no increase in severity. Pollack et al. (Annals of Oncology, 2018) found that patients with combination-induced colitis in particular tolerated rechallenge well. Both papers flagged the absence of structured rechallenge protocols as a gap in the field.

That literature existed when KEYNOTE-991 was designed. The protocol did not build from it.

> *You might be thinking: managing rechallenge adds complexity.*

Yes. And complexity managed at protocol stage is always preferable to complexity discovered at futility review.

04. What I Built

Both cracks are expressions of one underlying failure: the trial was designed with cytotoxic chemotherapy assumptions applied to immune checkpoint biology. My intervention rebuilds the monitoring architecture around immunotherapy's actual temporal logic.

Two structures. One thesis.

> **Early uncertainty is acceptable. Permanent uncertainty is not.**

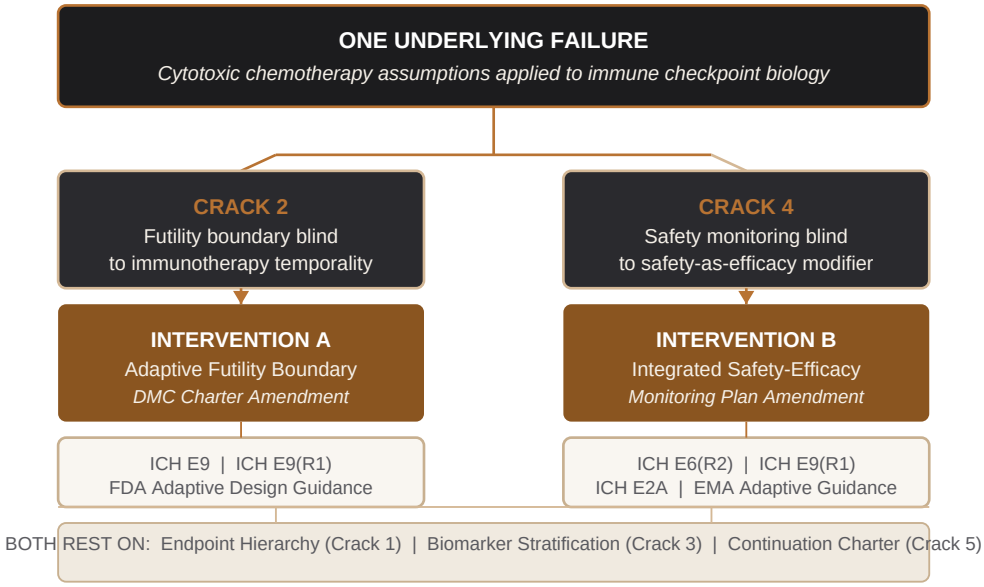

Figure 1. Unified Intervention Architecture — KEYNOTE-991

Intervention A. Adaptive Futility Boundary with Immunotherapy-Aware Conditional Power

Element	Standard Design (KEYNOTE-991)	Adaptive Design
Futility assessment timing	Single interim at ~50% events	Two interims: 40% events and 70% events
Conditional power threshold	<20% = stop	<15% at first interim; <25% at second
Conditional power methodology	Constant hazard assumption	Time-varying hazard model — delayed effect parameter
Late-emerging signal provision	None	Extended follow-up trigger if HR <1.0 with wide CI at second interim

The asymmetric boundaries are intentional. More permissive early, because delayed benefit may be biologically normal in the first half of an immunotherapy trial. More stringent late, because eventually, the treatment must prove itself. The system becomes more patient early and less forgiving over time.

The conditional power methodology draws on Zhang & Pulkstenis (2016), which demonstrated that standard interim monitoring in group sequential designs is biased when proportional hazards are assumed under delayed treatment effects. External immunotherapy data from melanoma, lung, and renal cell carcinoma informs the prior distribution for delayed effect.

DMC Charter Amendment:

> *"For immunotherapy arms, the DMC shall apply a time-varying conditional power model with delayed effect parameter δ estimated from external immunotherapy trial data in solid tumours. Futility boundaries are asymmetric: more permissive at first interim (40% events), more stringent at second (70% events). If conditional power falls below threshold but point estimate HR <1.0, the DMC may recommend conditional continuation with pre-specified biomarker enrichment criteria, rather than automatic termination. Binary stop/go language is insufficient for immunotherapy temporal biology. This charter provides a third option: conditional continuation with defined governance criteria."*
>
> *Regulatory anchor: ICH E6(R2) §5.19 — DMC responsibilities and charter requirements. ICH E9 Statistical Principles — interim analysis and stopping rules. ICH E9(R1) — estimand framework for intercurrent events. FDA Adaptive Designs Guidance (2019) — pre-specification requirements for adaptive futility rules.*

The charter does not remove the brakes. It gives the brakes a vocabulary for ambiguity.

Intervention B. Integrated Safety-Efficacy Monitoring

Element	Standard Design (KEYNOTE-991)	Integrated Design
Safety endpoint	Grade ≥3 AE rate, SAE rate	Dose-intensity-adjusted efficacy endpoint
Discontinuation handling	Censored at discontinuation	IPCW-adjusted survival analysis
irAE management	Standard CTCAE reporting	Grade-stratified rechallenge algorithm
Real-time trigger	None	Dose-intensity dashboard with escalation threshold

On IPCW: Standard analysis censors patients when they discontinue, treating their exit as random. It is not random. The sickest patients, the most toxicity-sensitive patients, exit earliest. The survival curves that remain are systematically optimistic about the patients who stayed. Inverse probability of censoring weights corrects for this by asking what the efficacy picture would look like if discontinuation had been random. When a third of your experimental arm has left before meaningful exposure, your survival curves are not telling the full truth. IPCW asks for the full truth.

irAE Rechallenge Algorithm:

- **Grade 2 irAE** → hold treatment → resolve → rechallenge at full dose with enhanced monitoring
- **Grade 3 irAE** → hold treatment → resolve → rechallenge at reduced dose with prophylactic immunosuppression
- **Grade 4 irAE** → permanent discontinuation

Simonaggio (2019) and Pollack (2018) demonstrated that structured rechallenge is feasible and that severity does not increase on retry when managed carefully.The protocol should have built from that literature. It did not.

Dose-Intensity Dashboard:

A real-time monitoring tool tracking the proportion of patients falling below 80% of planned dose in any given month. If that proportion exceeds 25%, automatic escalation to Protocol Amendment Review. This is something you check, not something you report quarterly.

Monitoring Plan Amendment:

> *"Safety data shall be reviewed not as isolated adverse events but as structural efficacy modifiers. Monthly reports shall include: (1) dose-intensity distribution by arm, (2) efficacy outcomes stratified by exposure duration, (3) IPCW-adjusted survival curves, (4) irAE resolution and rechallenge outcomes. The Medical Monitor and Data Management Lead shall jointly review the dose-intensity dashboard on a weekly basis. Automatic escalation to Protocol Amendment Review is triggered if >25% dose-intensity failure occurs in any calendar month. Discontinuation shall not be the default response to grade 2-3 irAEs in patients otherwise benefiting from treatment. The rechallenge algorithm is pre-specified, not improvised."*

Regulatory anchor: ICH E6(R2) §5.18 — monitoring responsibilities and real-time data review. ICH E6(R2) §5.16 and ICH E2A — expedited safety reporting and SAE documentation. ICH E9(R1) — estimand framework governing IPCW analysis. Protocol amendment procedures: ICH E6(R2) §4.5 and §5.12.

05. Why I Am Confident

> **The guidelines were already there.**

The irAE rechallenge literature existed before this trial enrolled its first patient. The time-varying hazard framework for immunotherapy conditional power existed. The concept of dose-intensity as an efficacy modifier existed in the pharmacokinetic literature. The field had already published the questions this trial failed to ask.

I am not proposing innovation. I am proposing that the people who designed this trial read the room they were already standing in.

My confidence is the quiet certainty that comes from following evidence to its logical operational conclusion.

06. A Final Word

You might be wondering why I did this.

To impress you? Maybe. Ego? Probably.

But there is something people tend to miss, and it matters more than the methodology.

This is what I do on a quiet evening. Not because a job requires it. Because the moment I saw that a third of those patients exited before the treatment could fairly answer for itself, I could not leave the question alone. Because trial governance that speaks only in binary; stop or continue, is a system that has not thought carefully enough about what happens in the middle, where most of the truth actually lives.

> **The instinct to read the architecture, not just the results, is what I am building a career around.**

07. References

Primary Trial Publication

Shore ND et al. Pembrolizumab plus enzalutamide versus placebo plus enzalutamide in metastatic castration-resistant prostate cancer (KEYNOTE-991): a randomised, double-blind, phase 3 trial. *Annals of Oncology.* 2025.
https://doi.org/10.1016/j.annonc.2025.01.010

irAE Rechallenge — Primary

Simonaggio A, Michot JM, Voisin AL, et al. Evaluation of readministration of immune checkpoint inhibitors after immune-related adverse events in patients with cancer. *JAMA Oncology.* 2019;5(9):1310–1317.
https://doi.org/10.1001/jamaoncol.2019.1022

irAE Rechallenge — Combination Therapy

Pollack MH, Betof A, Dearden H, et al. Safety of resuming anti-PD-1 in patients with immune-related adverse events (irAEs) during combined anti-CTLA-4 and anti-PD1 in metastatic melanoma. *Annals of Oncology.* 2018;29(1):250–255.
https://doi.org/10.1093/annonc/mdx642

Adaptive Design Methodology

Zhang L, Pulkstenis E. Group sequential design under non-proportional hazards with delayed treatment effects. *Statistics in Biopharmaceutical Research.* 2016.

Regulatory — ICH E6(R2)

ICH E6(R2). Good Clinical Practice: Integrated Addendum. International Council for Harmonisation, 2016.
https://www.ich.org/page/efficacy-guidelines

Regulatory — ICH E9(R1)

ICH E9(R1). Addendum on Estimands and Sensitivity Analysis in Clinical Trials. International Council for Harmonisation, 2019.
https://www.ich.org/page/efficacy-guidelines

Regulatory — ICH E2A

ICH E2A. Clinical Safety Data Management: Definitions and Standards for Expedited Reporting. International Council for Harmonisation, 1994.
https://www.ich.org/page/safety-guidelines

Regulatory — FDA Adaptive Designs

U.S. Food and Drug Administration. Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry. FDA, 2019.
https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry

Regulatory — EMA Adaptive Design

European Medicines Agency. Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design. EMA/CHMP/EWP/2459/02, 2007.
https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-methodological-issues-confirmatory-clinical-trials-planned-adaptive-design_en.pdf