Cancer Biol Med 2026. doi: 10.20892/j.issn.2095-3941.2025.0759

PERSPECTIVE

Protocol and statistical designs in classic clinical research—toripalimab series trial analysis

Yingjian He[1]*, Li Bian[2]*, Zefei Jiang[2]

[1]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Breast Center, Peking University Cancer Hospital & Institute, Beijing 100142, China; [2]Senior Department of Oncology, Chinese PLA General Hospital, Beijing 100071, China

Both large-scale prospective randomized controlled trials (RCTs) and smaller investigator-initiated trials are essential for evaluating the efficacy and safety of medical interventions. Robust protocols and statistical designs ensure the reliability of trial outcomes and improve the credibility of research findings. By reviewing the statistical approaches used in the TORCHLIGHT, NCC2167, and NeoTENNIS trials, this article illustrates the principles underlying large-sample confirmatory RCTs, small-sample exploratory adaptive designs, and single-arm two-stage designs. This discussion is aimed at helping researchers apply these design methods more effectively, to increase the likelihood of success in clinical studies.

TORCHLIGHT trial

TORCHLIGHT was the first phase III trial in China to investigate a programmed cell death protein 1 (PD-1) inhibitor combined with chemotherapy. It evaluated the efficacy and safety of toripalimab plus taxane as a first-line treatment for advanced triple-negative breast cancer (TNBC). The study used a multicenter, randomized, double-blind, placebo-controlled design, with nab-paclitaxel as the chemotherapy regimen (**Figure 1**). The results were published in Nature Medicine in 2024[1]. At the trial design stage, several key factors guided the choice of chemotherapy agent for combination with toripalimab. The selected drug was required to be effective, have low toxicity, and be compatible with immunotherapy. Nab-paclitaxel offered substantial advantages over other chemotherapy agents[2]. First, paclitaxel nanoparticles bind natural albumin, thereby increasing drug delivery and bioavailability, and consequently enhancing the therapeutic response, inducing immunogenic tumor cell death, and promoting tumor antigen release. Second, paclitaxel is rapidly transported out of the blood circulation through albumin binding, enters endothelial cells *via* receptor-mediated endocytosis, and accumulates in tumor tissues through the enhanced permeability and retention effect of the tumor vasculature. As a result, damage to normal tissues is mitigated, immune cell survival is preserved, and immune efficacy is supported. Third, because nab-paclitaxel does not cause solvent-related allergic reactions, glucocorticoid premedication is unnecessary. Therefore, immune suppression is avoided, and the efficacy of immune checkpoint inhibitors is maximized. The selection of nab-paclitaxel as the chemotherapy partner for toripalimab was crucial to the success of the TORCHLIGHT trial.

The trial used a 2:1 parallel design between the experimental and control groups, involving unequal randomization. Although allocating more participants to the experimental group is more costly and requires a larger total sample size than a 1:1 design, this approach is widely used in confirmatory studies. Moreover, this design aligns with ethical principles and may accelerate trial progression. The sample size difference between groups should remain moderate; for example, a 2:1 or 3:2 ratio is recommended. The primary endpoint was progression-free survival (PFS), assessed through blinded independent central review (BICR) in the intention-to-treat (ITT) population. The key secondary endpoint was overall survival (OS) in the ITT population. BICR assessment minimized evaluation bias, and investigator-assessed PFS served as a sensitivity analysis to strengthen the reliability of the findings. In 2019, the protocol was amended to include a programmed death-ligand 1 (PD-L1) positive subgroup for both primary and key secondary endpoints. Type I error across the primary and key secondary endpoints was controlled through a fixed-sequence hierarchical strategy involving PFS in the PD-L1 positive subgroup, PFS in the ITT population, OS in the PD-L1 positive subgroup, and OS in the ITT population. Interim analysis indicated larger between-arm differences in both PFS and OS in the PD-L1 positive subgroup than the ITT population, and a negative outcome was observed in the interim analysis.

*These authors contributed equally to this work.
Correspondence to: Zefei Jiang
E-mail: jiangzefei@csco.org.cn
ORCID ID: https://orcid.org/0009-0005-1627-1010
Received November 27, 2025; accepted February 5, 2026.
Available at www.cancerbiomed.org
©2026 The Authors. Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

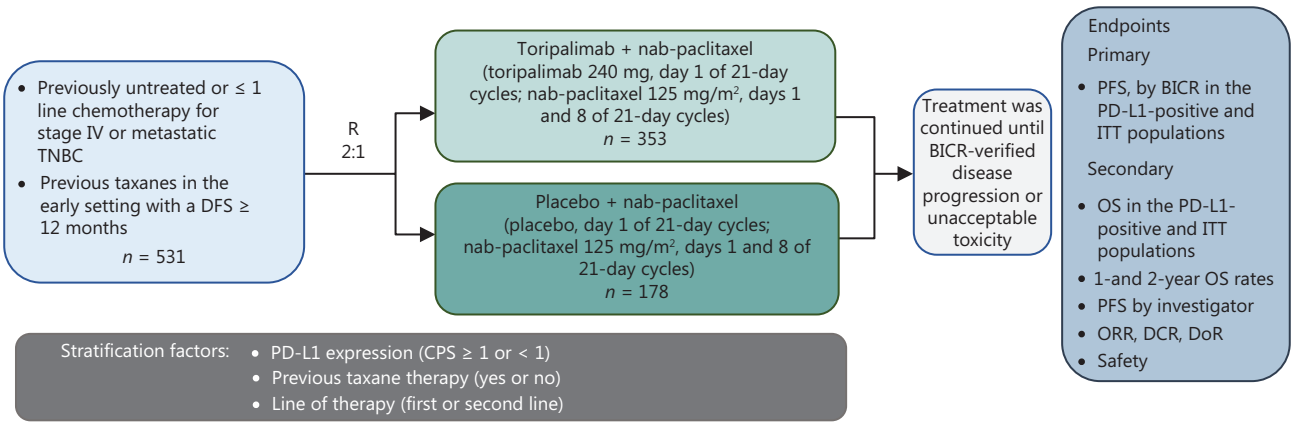


Figure 1 Study design of the TORCHLIGHT trial. BICR, blinded independent central review; CPS, combined positive score; DCR, disease control rate; DFS, disease free survival; DoR, duration of response; ITT, intention-to-treat; ORR, overall response rate; OS, overall survival; PD-L1, programmed death-ligand 1; PFS, progression-free survival; TNBC, triple-negative breast cancer.

In contrast, the IMpassion130 trial[3] controlled type I error for OS first in the ITT population and then in the PD-L1 positive subgroup. Because the OS in the ITT population was negative, the analysis for the PD-L1 positive subgroup could only be descriptive. Therefore, in future immunotherapy research, prioritizing biomarker-positive subgroups in a fixed-sequence hierarchical framework might enable earlier drug approval and allow more patients to benefit sooner.

However, prioritizing biomarker-positive subgroups increases the complexity of controlling type I error. For example, the EMBER-3 study[4] added PFS in the biomarker-positive [estrogen receptor 1 (ESR1)] population *via* protocol amendment, but ESR1 status was not pre-specified as a random stratification factor. In contrast, TORCHLIGHT pre-specified PD-L1 expression as a stratification factor in the design stage, thus ensuring balance between the experimental and control groups within the PD-L1 positive subgroup. This design increased the reliability of subgroup results and is a strategy worthy of consideration in future clinical trials.

NCC2167 trial

Small-sample exploratory studies are better suited than large confirmatory trials to achieving the timely findings required for innovation. Adaptive design is likely to become a key trend in future clinical research. For example, the NCC2167 trial was a phase II study using a Bayesian adaptive design for multi-regimen selection based on metronomic chemotherapy (**Figure 2**). It enrolled only 103 participants (97 of whom were evaluable), and its final results, including detailed design, were rapidly published in Nature Medicine in 2024[5], thus demonstrating the efficiency of the trial model.

The NCC2167 trial compared the efficacy of multiple combination regimens (five arms) of metronomic chemotherapy, anti-tumor angiogenesis therapy, and the PD-1 inhibitor toripalimab in human epidermal growth factor receptor-2 (HER-2) negative advanced breast cancer. The use of metronomic chemotherapy[6] was a novel concept: beyond its cytotoxic effects, it suppresses tumor neovascularization and modulates immune responses, thus enabling synergistic effects in combination with immunotherapy. The Bayesian response-adaptive randomization method[7] used in this trial aligns more closely with ethical principles than classical covariate-adaptive randomization. On the basis of early efficacy data, subsequent participants have a higher probability of being assigned to treatment arms showing superior benefit. This approach not only increases the likelihood of benefit for participants but also decreases the required sample size, lowers trial costs, and increases the probability of identifying the optimal immunotherapy combination. Although Bayesian designs remain debated in academia, the New England Journal of Medicine, in its 2019 Statistical Reporting Guidelines[8], has anticipated their wider use, thus laying a foundation for their broader application.

Another important aspect of the NCC2167 trial was its integration of translational research and preliminary mechanistic exploration. Exploratory immune analyses were performed on peripheral blood mononuclear cells with mass cytometry flow technology. Metronomic (vinorelbine + capecitabine + cyclophosphamide) combined with toripalimab substantially increased the expression of CCR4 (a chemokine receptor involved in DC–T cell interactions) on $CD4^+$ and effector memory $CD8^+$ T cells. The proportion of cluster 19 (efficacy-associated NK cell subsets), as well as clusters 29 ($CX3CR1^+$ monocytes) and 30 ($HLA\text{-}DR^+$ monocytes), markedly increased after treatment. These findings indicated that

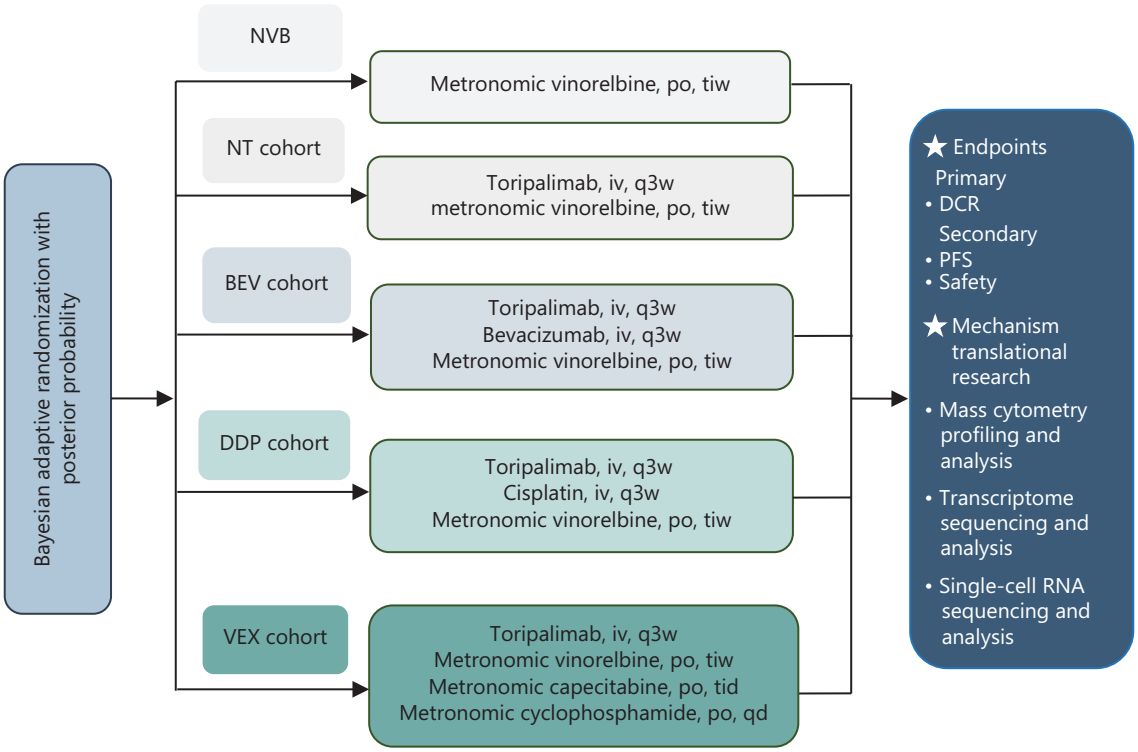


Figure 2 Study design of the NCC2167 trial.

immune reprogramming is closely associated with immunotherapy efficacy. Various combination strategies might exert distinct effects on the immune system and therefore influence clinical outcomes. These results highlight that in-depth mechanistic investigations can enhance the scientific value of small-sample trials and support their publication in top-tier journals.

NeoTENNIS trial

The NeoTENNIS trial used the classic single-arm Simon two-stage design (**Figure 3**), which is the most commonly applied design in exploratory clinical research and can secure enterprise funding more easily than RCT designs. The trial results were published in eClinicalMedicine in 2024[9].

NeoTENNIS investigated a neoadjuvant strategy for early-stage TNBC, in which induction chemotherapy is followed by combination treatment with an immune checkpoint inhibitor. In the first stage, dose-dense epirubicin plus cyclophosphamide was administered as an induction chemotherapy before immunotherapy. In the second stage, toripalimab was combined with albumin-bound paclitaxel. The results indicated a 55.7% total pathological complete response rate, 58.6% breast pathological complete response (pCR) rate, 58.6% rate of a Miller-Payne score of 5, 65.7% rate of residual cancer burden class 0-I, and 92.9% MRI-confirmed objective response rate. After 4 cycles of EC induction chemotherapy, $CD8^+$ lymphocyte infiltration significantly increased, thus indicating that induction chemotherapy enhanced the anti-tumor immune response. NeoTENNIS used a de-escalated neoadjuvant chemotherapy and immunotherapy regimen for TNBC, involving chemotherapy without platinum agents and limiting immunotherapy to 4 cycles combined with nab-paclitaxel. This approach minimized chemotherapy- and immunotherapy-related toxicity while maintaining favorable efficacy and safety profiles. The findings provide a valuable reference and guidance for clinical practice.

From a sample size perspective, in single-arm trials, the Simon two-stage design[10] offers several options. When confidence in the investigational drug is low, or prior data are limited, the optimum design can minimize the number of patients in the first stage and the expected total sample size, thereby limiting the cost of trial and error. When confidence in the drug is high, the minimax design can be used to limit the total sample size and control overall costs. To achieve a balance between trial-and-error costs and total sample size, the admissible design criterion, developed with Bayesian methods, is a suitable alternative. Notably, from a feasibility perspective (e.g., under limited drug supply), a precision-driven approach to sample size calculation is often more aligned with

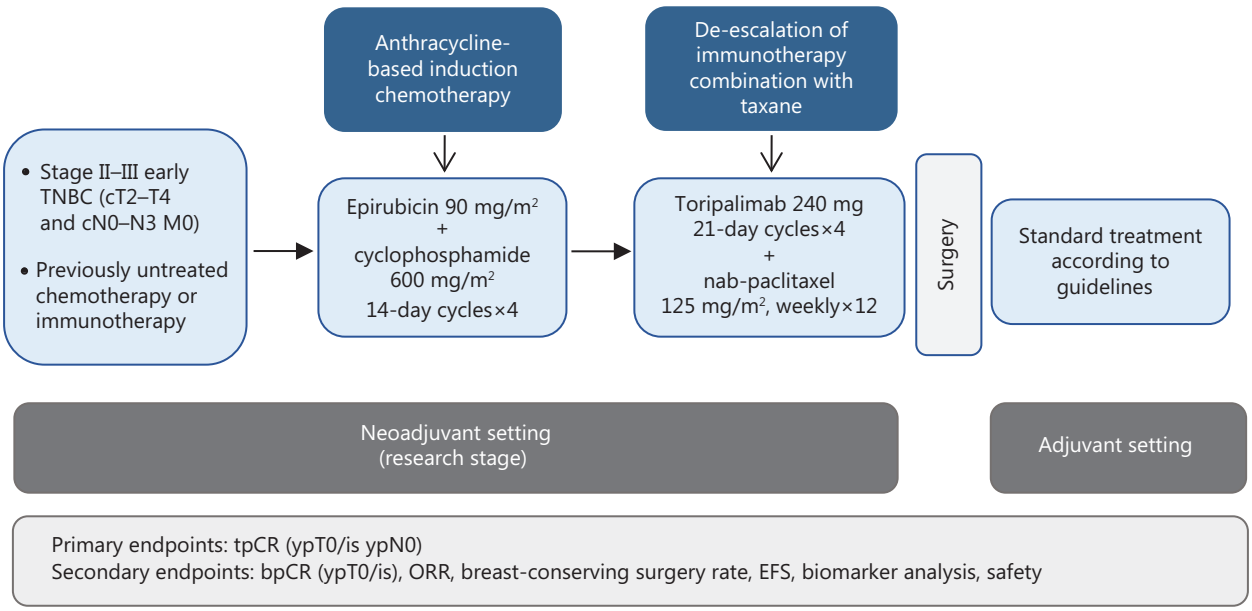


Figure 3 Study design of the NeoTENNIS trial. TNBC, triple-negative breast cancer; tpCR, total pathological complete response; bpCR, breast pathological complete response; ORR, overall response rate; EFS, event-free survival.

real-world clinical requirements. By controlling the precision of the primary endpoint (e.g., the pCR rate in NeoTENNIS), typically the width of the 95% confidence interval, within regulatory and scientific boundaries, investigators can tailor the sample size according to available resources. This strategy has been applied in numerous trials. The main limitation of the single-arm design is the absence of a parallel control group, randomization, and blinding. Even when external controls are used, bias cannot be fully eliminated, thus decreasing the reliability of the study findings and limiting the generalizability of the results.

Summary

Regardless of whether a study is an RCT or an investigator-initiated trial, the optimal trial design and statistical methods should be chosen according to the specific research objectives, to ensure that the conclusions are accurate, and clinical questions are addressed effectively. Evidence from the literature and insights from landmark trials allow clinicians to "take the essence and discard the dross," drawing useful lessons fromclassic clinical trials while avoiding flawed designs, thus achieving the best possible results with limited resources.

In summary, drawing on 3 representative studies, this article offers key insights and recommendations for new drug development: align trial design with the phase of clinical development; learn from the methodological limitations of previous trials to optimize subsequent designs; and balance innovation with methodological rigor.

Implications for future clinical research can be derived by examining findings from earlier studies. Because of space constraints, a comprehensive analysis of the broader R&D landscape was not feasible herein. No strategy is without limitations, and understanding the strengths and weaknesses of different methods is essential. For example, Bayesian response-adaptive randomization is more ethical and better suited to small-sample studies than large studies. However, it can result in imbalanced sample sizes between groups, particularly in poorly performing control arms with small numbers, thus decreasing statistical power and increasing the risk of study suspension. Therefore, the selection of the most appropriate trial designs and statistical methods should always consider the practical circumstances of the study.

Grant support

This work was supported by a grant from the National Science and Technology Major Project (Grant No. 2024ZD0519800).

Conflict of interest statement

No potential conflicts of interest are disclosed.

Author contributions

Conceived and designed the analysis: Zefei Jiang, Yingjian He, Li Bian.

Collected the data: Li Bian, Yingjian He.
Contributed data or analysis tools: Zefei Jiang, Li Bian, Yingjian He.
Performed the analysis: Yingjian He, Li Bian.
Wrote the paper: Yingjian He, Li Bian, Zefei Jiang.

References

1. Jiang Z, Ouyang Q, Sun T, Zhang Q, Teng Y, Cui J, et al. Toripalimab plus nab-paclitaxel in metastatic or recurrent triple-negative breast cancer: a randomized phase 3 trial. Nat Med. 2024; 30: 249-56.
2. Zhang Y, Chen H, Mo H, Zhao N, Sun X, Liu B, et al. Distinct cellular mechanisms underlie chemotherapies and PD-L1 blockade combinations in triple-negative breast cancer. Cancer Cell. 2025; 43: 446-63.e7.
3. Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. N Engl J Med. 2018; 379: 2108-21.
4. Jhaveri KL, Neven P, Casalnuovo ML, Kim SB, Tokunaga E, Aftimos P, et al. Imlunestrant with or without abemaciclib in advanced breast cancer. N Engl J Med. 2025; 392: 1189-202.
5. Mo H, Yu Y, Sun X, Ge H, Yu L, Guan X, et al. Metronomic chemotherapy plus anti-PD-1 in metastatic breast cancer: a Bayesian adaptive randomized phase 2 trial. Nat Med. 2024; 30: 2528-39.
6. Munzone E, Regan MM, Cinieri S, Montagna E, Orlando L, Shi R, et al. Efficacy of metronomic oral vinorelbine, cyclophosphamide, and capecitabine *vs.* weekly intravenous paclitaxel in patients with estrogen receptor-positive, ERBB2-negative metastatic breast cancer: final results from the phase 2 METEORA-II randomized clinical trial. JAMA Oncol. 2023; 9: 1267-72.
7. Albain KS, Yau C, Petricoin EF, Wolf DM, Lang JE, Chien AJ, et al. Neoadjuvant trebananib plus paclitaxel-based chemotherapy for stage II/III breast cancer in the adaptively randomized I-SPY2 trial-efficacy and biomarker discovery. Clin Cancer Res. 2024; 30: 729-40.
8. Harrington D, D'Agostino RB Sr, Gatsonis C, Hogan JW, Hunter DJ, Normand ST, et al. New guidelines for statistical reporting in the *Journal*. N Engl J Med. 2019; 381: 285-6.
9. He M, Hao S, Ma L, Xiu B, Yang B, Wang Z, et al. Neoadjuvant anthracycline followed by toripalimab combined with nab-paclitaxel in patients with early triple-negative breast cancer (NeoTENNIS): a single-arm, phase II study. EClinicalMedicine. 2024; 74: 102700.
10. Wang C, Wang Z, Rosner GL, Huh WK, Roden RBS, Bae S. A batch-effect adjusted Simon's two-stage design for cancer vaccine clinical studies. Biometrics. 2021; 77: 1075-88.

Cite this article as: He Y, Bian L, Jiang Z. Protocol and statistical designs in classic clinical research—toripalimab series trial analysis. Cancer Biol Med. 2026; 23: 338-342. doi: 10.20892/j.issn.2095-3941.2025.0759