Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

To Borrow or Not? A Practical Framework for Navigating Non- concurrent Controls Borrowing in Platform Trials with Time Trends

Chen Li

lc.biosta@qq.com

Fourth Military Medical University

Siyu Xiong

Fourth Military Medical University

Binbin Kang

Fourth Military Medical University

Research Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11198716/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

To Borrow or Not? A Practical Framework for Navigating Non-concurrent Controls Borrowing in Platform Trials with Time Trends

Chen Li*, Siyu Xiong and Binbin Kang

Authors'information:

Department of Health Statistics, School of Preventive Medicine, Fourth Military Medical University, Ministry of Education Key Lab of Hazard Assessment and Control in Special Operational Environment, Xi'an, Shaanxi, China.

Corresponding authors: Chen Li, Department of Health Statistics, School of Preventive Medicine, Fourth Military Medical University, Ministry of Education Key Lab of Hazard Assessment and Control in Special Operational Environment, No.169 Changlexilu Road, Xi'an 710032, Shaanxi, China.(Email:lc.biosta@qq.com)

Abstract

Background: Platform trials gained popularity during the last few years as they increase flexibility by allowing new treatment arms entering to an on-going trial with a shared control group. For the new experimental treatment arms, the shared control group includes concurrent control data and non-concurrent controls (*NCC*) that the control patients before the new arm entry. Although borrowing *NCC* data can improve statistical efficiency and reduce sample size requirements, it may also yield estimation bias in case of temporal drift if the appropriate methodology is not used and the assumptions are not met.

Methods: We investigate the operating characteristics of *NCC* borrowing strategies in platform trials for the newly added treatment estimation with both frequentist and Bayesian analysis approaches comprehensive clinically plausible scenarios of sequential treatment entry, varying degrees of enrollment overlap, different lengths of entry intervals, and temporal-trend patterns including linear, stepwise, and inverted-U trends. For trials with continuous or binary outcomes, methods explored include included sample pooling, regression adjustment, step function models, the Bayesian time machine(BTM), dynamic power prior (DPP), meta-analytic predictive (MAP) priors, and the elastic MAP (EMAP) prior. For trials with continuous or binary outcomes, performance was explored based on Type I error rate and power, as well as the bias and mean squared error of the newly added treatment efficacy across various drift parameters and enrollment patterns.

Results: The reliability of NCC borrowing depended primarily on the homogeneity of time trends across treatment and control arms. When trends were homogeneous, *NCC* borrowing was generally reliable, in which the BTM, MAP, and EMAP methods robustly maintained Type I error control to the nominal level while improving power relative to *CC*-only analysis. With long or irregular entry intervals, pooled analysis and DPP showed substantial bias and Type I error inflation. When temporal heterogeneity was limited to a non-target treatment arm, *NCC* borrowing remained generally feasible, and most methods preserved acceptable estimation accuracy. In contrast, when drift occurred only in the control arm or when control and treatment arms had markedly different or oppositely directed trends, all methods exhibited inflated Type I error, with step trends posing the greatest risk. Although the EMAP prior consistently minimized bias but was associated with more conservative power. Strong stepwise heterogeneity produced the most severe failures.

Conclusion: Borrowing of *NCC* data is not intrinsically safe or unsafe.Its reliability in platform trials depends on the temporal-trend structure. A practical safe-borrowing region consists primarily of homogeneous or only mildly heterogeneous trends, especially when the trend is smooth and the entry interval is not excessively long. *NCC* borrowing should be restricted or used only as a sensitivity analysis when strong control-specific, oppositely directed, or abrupt stepwise trends are plausible.

Keywords： Platform Trials; Non-Concurrent Controls; Time Trends; Bayesian Methods

Introduction

Platform trials represent a paradigm shift in clinical trial design, which investigate multiple experimental interventions simultaneously against a shared control arm within a single, overarching master protocol[1]

. This innovative adaptive design possess inherent flexibility. New treatment arms can be seamlessly incorporated into the ongoing trial as new scientific discoveries emerge, while existing arms may be dropped for futility or efficacy based on pre-specified interim decision rules. By sharing the control group across multiple interventions, the dynamic structure optimizes statistical power and confer substantial efficiencies compared to running separate, standalone trials, thereby accelerating the drug development process[2] and encompassing a wide range of therapeutic areas, including but not limited to cancer, Alzheimer's disease, and the COVID-19 pan-demic, and so on[3-7].

Since the new adding treatment arm are permitted to the on-going platform trial, the control data can be further categorized into two groups: control patients randomized concurrently with a new treatment arm, referring as concurrent control (*CC*), and control patients recruited before that new treatment arm entered the platform, that is non-concurrent control (*NCC*) data. As shown in Figure 1, exemplifies with two experimental arms, where the new treatment arm 2 joined the trial at a later time point in a platform trial, the control data before arm 2 enrolled can be considered as its *NCC* data. Actually, Pocock explored the application of historical controls in the design and analysis of randomized controlled trial (RCT) as early as 1976[8], and the academic researches have consistently focused on how to combine the historical controls while avoiding the introduction of bias, and reducing the required total sample size and/or shortening the average total trial duration. As International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) E10 guideline[9] and Jahanshahi's recommendations[10], these historical data can be categorized as non-concurrent external controls. While in that terminology, the NCC in platform trials should be regarded as non-concurrent internal controls, it may still differ from concurrent control[11] and therefore become complicate issue for the new treatment estimation.

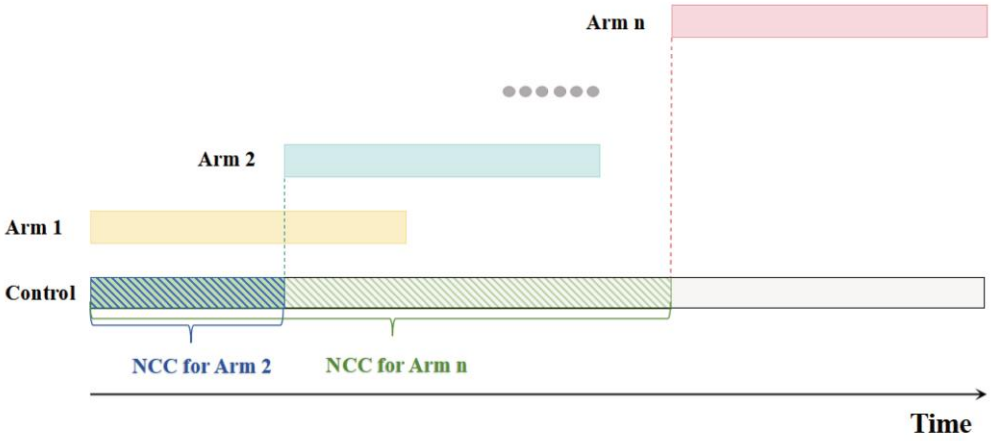

Figure 1 Non-concurrent controls (*NCC*) data in platform trials

It is acknowledged that incorporating all available controls can increase power and yield more precise estimates. For instance, in trials with low event rates (e.g., oncology trials evaluating progression-free survival), *NCC*s can reduce the required sample size by up to 30% compared to *CC*-only analyses[12]. However, treatment response may drift overtime for many reasons, which is defined as the difference between the true unknown concurrent control parameter and the historical control data[13]. For example, sites opening to enrollment at different times, protocol procedure learning curves, a change to standard of care, and centers' different experience in conducting process can all impact trial-wide outcome rates. Since the new treatment arms are added sequentially, while randomization occurs at different time points, temporal drift may yield biased estimates and impact type I error rates for the lack of truly concurrent randomization across time. Therefore, to our knowledge, methodology and guidance for *NCC* application in platform trials remain limited and not been well-studied.

To avoid these risks, Ventz et al.[14] recommends conservative approaches to use only concurrently randomized controls for the analyses of each experimental arm. If drift found, controls enrolled within a pre-specified time interval before the experimental arm is introduced. Hobbs and Yuan [12, 13] have pooling all controls in open platform design only under the stipulation that drift is not a concern. Beyond naively pooling *NCC* or only using *CC*, alternative methods that allow the data to determine the degree of borrowing seem better suitable for guarding against population drift in parameter. For example, the simplest dynamic strategy is the "test-then-pool" method, which

combines *NCC* and *CC* only if the null hypothesis of their equivalence cannot be rejected[13, 15]. Drawing on the strategy of external data borrowing, propensity scores (PS) are employed as a balancing score to adjust for baseline covariate imbalances between historical and concurrent control data, typically via matching, stratification, or weighting. For example, PS-matching[13] or a PS-integrated composite likelihood approach[16] can augment the concurrent control with historical or real-world data. Although it increases statistical efficiency and reduced confounding bias through covariate balance, PS methods are inability to adjust for unmeasured confounding factors. Therefore, it is more commonly employed in scenarios where the external real-world data borrowing, but not been widely observed in platform trials for the leveraging *NCC* as internal control data.

Actually, a prevalent strategy broadly used is to correct for temporal drift between *NCC*s and *CC*s by incorporating time as a covariate in the regression models. In these methods, the time trends always be modeled in as the linear or polynomial trend as continuous covariate variables, grouped into intervals with a separate parameter for each period, or smoothing estimates across pre-defined time increment using a specific model structure. For instance, Lee and Wason[17] proposed frequentist linear regression models with time as a linear or stepwise categorical variable, demonstrating that stepwise adjustment can yield unbiased estimates if time trends are equal across all arms and block randomization is employed. Roig et al.[18] further investigated using a step-wise function relies on the assumption of equal time trends across all arms and on the correct specification of the scale of the time trends in the model. Then, Krotka et al.[19] extended frequentist methods by introducing spline regression to capture complex, non-linear time trends, showing that cubic splines effectively control Type I error for smooth temporal patterns and offer power gains when treatment arm overlap is limited. For the Bayesian approaches, Saville et al.[20] developed the “Bayesian Time Machine” hierarchical model that partitions time into pre-defined time buckets with a normal dynamic linear model to smooth time effects and improved precision. However, the performance of these methods critically depends on the correct specification of the time trend's functional form and scale. The key assumption of equal time trends across all arms is often untestable and may not hold in practice. Violations of this assumption can lead to bias. Furthermore, results can be sensitive to how time intervals are defined (e.g., by arm entry vs. calendar buckets).

From a Bayesian perspective, a common strategy to handling *NCC* issue is discounting its

information and partially borrowing based on its similarity to *CC*s to account for potential heterogeneity between the historical and *CC* data. The most intuitive example is discounting historical data via a scalar weight power parameter that controls its influence on the current trial's posterior estimation[21]. Thus, the choice of the value of this parameter is essential, it is often challenging to specify. Subsequently, many researchers have improved the modified power prior, such as empirical bayes approach based on accumulated observed data[22], incorporating multiple historical control groups[23]. They assess the agreement between historical and concurrent control with a weight based on an equivalence probability weight or tail area probabilities[24]. While the commensurate power prior gives another adaptive modification that use conditional prior distributions of *CC* data to adjust the weight parameter based on a measure of compatibility between historical and concurrent controls. For instance, Normington et al.[25] applied commensurate priors to borrow control data internally across arms in platform trials rather than from external control data. However, their performance critically depends on the appropriate specification of discounting parameters or prior distributions, which is often non-trivial and can lead to operational sensitivity.

To overcome the subjectivity associated with the specification of the discounting parameter, the Meta-Analytic-Predictive (MAP) approach summarize several available historical information as a mixture prior with random-effects framework. This Bayesian hierarchical model accounts for heterogeneity between-trials by assuming exchange ability among the historical and concurrent control parameters to perform a prediction of the control effect from historical control data, or with robust extensions (R-MAP) incorporating weakly informative mixture components to safeguard against extreme discrepancies[26]. Thus, for platform trials, when *NCC* data reflect heterogeneous temporal trends, these can be stratified into distinct time intervals and integrated into a mixture prior with MAP method or R-MAP with a weakly-informative prior. More recent developments include the Bayesian semi-parametric MAP prior (BaSe-MAP)[27], which models random effects non-parametrically via a Dirichlet process mixture for more conservative borrowing, Wang[28] proposed an extension of the R-MAP approach to adjust the time trends in platform trials when the standard of care control arm was permitted to be changed over different recruitment times based on a trials of Ebola virus disease therapeutics. However, since the consistency between *NCC* and *CC* is inherently dynamic with the accruing of the trial participants, the information borrowing of *NCC* should be adaptive adjustment.

Although the aforementioned methods have progressively enhanced the robustness of the *NCC* prior borrowing for varying time trends, they generally implement static or pre-specified borrowing schemes rather than fully adaptive, data-driven dynamic borrowing. As a result, they cannot support dynamic borrowing of *NCC* prior information in response to accumulating trial-internal evidence, limiting their responsiveness to emerging discrepancies between *NCC* and *CC* data.

Based on the dynamic borrowing strategy, Jiang et al.[29] proposed an elastic prior method to quantify the consistency between concurrent and historical data. While the elastic function can take various construction, for instance, as a congruence measure mapped to the interval (0,1) est statistic, it ensures a set of pre-specified criteria such that the prior strongly borrows information when historical and trial data are consistent, but reducing borrowing by inflating the prior variance when the two sources are incongruent. Subsequently, Zhang et al.[30] integrated the elastic prior with the MAP prior framework to develop an Elastic Meta-Analytic-Predictive (EMAP) prior. This approach allows for the synthesis of multiple historical information and enables dynamic borrowing historical information. Although EMAP has been applied in the efficacy evaluation of biosimilars, its extension to platform trials remains unexplored, and its applicability for dynamic information borrowing in such settings warrants further investigation. Besides, the above approaches under different time trends remains unclear. In particular, few studies have compared the relative performance of these above frequentist and Bayesian strategies across a range of time trends, such as linear, stepwise, or non-monotonic patterns.

To address this gap, we conducted a comprehensive simulation study to evaluate the operating characteristics of NCC borrowing methods in platform trials with some commonly used methods—pooling, regression adjustment, Time Machine, MAP and EMAP considering linear, stepwise, and inverted-U-shaped temporal trends, equidistant and non-equidistant treatment-entry patterns, and homogeneous and heterogeneous trends across treatment and control arms. Our primary objective was to characterize the conditions under which *NCC* borrowing improves efficiency without compromising type I error control or estimation accuracy, with particular emphasis on the robustness of EMAP-based borrowing method. We also developed a calibration-based framework to guide whether NCC borrowing should be permitted as a primary analysis, used with safeguards, or restricted to sensitivity analyses to provide practical recommendations for the reliable and efficient *NCC* borrowing in

platform trials.

Method

Trial framework

Consider a randomized, parallel group platform trial including a shared control arm and $K(\mathrm{K} \geq 2)$ experimental treatments entering the trial sequentially. The trial opens with only the first experimental treatment and control and then the new treatment arm added after a number of patients have been enrolled. Denote $k = 0$ for the control arm and $k = 1, \cdots, K$ for the experimental treatment arms, indexed by the trial entry orders. Thus, the total trial stages as time intervals bounded by the time points where the experimental treatment arm added. Once a new treatment arm $k(k \geq 2, \ldots, K)$ entering the trial, it marks the start of a new stage, indexed by $s_k(s = 1, \ldots, S)$, where S is the total number of stages. Let n_{0s}, n_{ks} be the sample sizes of control arm and treatment arm k in stage *s* respectively, so the sample size of stage s_k be $N_s = n_{ks} + n_{0s}$ with the total sample size of the trial $N = \sum_{s=1}^{S}(n_{ks} + n_{0s})$, y_j be the response of the jth patient in his arm k_j at recruitment time s_j. We focus on the efficacy evaluation of the new treatment against the control as soon as the new arm completed. Therefore, all data in the set $D_k = \{(y_j, k_j, s_j), j = 1, \ldots N | s_j \leq s_k^{exit}\}$ would be applied until treatment k leaves the platform at the time point s_k^{exit}. We aim to test the one-sided null hypothesis $H_{0,k}: \theta_k \leq 0 \; vs. H_{1,k}: \theta_k > 0$, where θ_k, a higher-better outcomes, denotes the treatment effect size for treatment arm k against control arm. Several model-based approaches are considered as follow.

Frequentist regression models

To estimate the new treatment effect against the control, regression models always conduct time as a covariate to adjust for potential time trends. The time interval is referred as the trial duration if no change in the number of treatment arms. A new time stage s_k starts when an experimental treatment arm k is added (starts being active) or removed (stops being active) from the platform. To estimate the treatment effect of treatment k against control, Lee and Wason[17] proposed linear regression model including all treatment arms and the control with the time stages before and after the new arm entry.

$$Ma_1:\ y_j = \eta_0 + \sum_{k=0}^{k+1} \theta_k \cdot I(T_j = k) + \gamma \cdot j + \varepsilon_j \quad (1)$$

$$Ma_2:\ y_j = \eta_0 + \sum_{k=0}^{k+1} \theta_k \cdot I(T_j = k) + \tau_s \cdot c_j + \varepsilon_j \quad (2)$$

where η_0 denotes the treatment effect of the control arm in the first stage, T_j be the allocated treatment of subject *j*, θ_k represents the mean difference of the treatment effects for treatment arm *k* compared to the control arm, ε_j be the random errors that are identically and independently normally distributed with mean 0 and variance σ^2. Then, the treatment effect of the new arm was estimated via

$$Mb_1:\ y_j = \eta_0 + \theta_0 \cdot I(T_j = 0) + \theta_{k+1}^{b_1} \cdot I(T_j = k+1) + \gamma \cdot j + \varepsilon_j \quad (3)$$

$$Mb_2:\ y_j = \eta_0 + \theta_0 \cdot I(T_j = 0) + \theta_{k+1}^{b_2} \cdot I(T_j = k+1) + \tau_s \cdot c_j + \varepsilon_j \quad (4)$$

where $\theta_{k+1}^{b_1}$ and $\theta_{k+1}^{b_2}$ represent the mean difference between the new arm $k+1$ and the control arm, c_j is an indicator function for subject *j* entering stage 2 , γ in equations (1) and (3) represents the effect of continuous recruitment, whereas τ_s represents the stage effect (the effect after the new treatment has been added).

Thus, the amount of the control data used to estimate θ_{k+1} in Ma_2 and $\theta_{k+1}^{b_2}$ in Mb_2 are similar to that of using only the concurrent control (stage 2) data, since the new arm is not present in stage 1. While the inference from Ma_1 without adjustment for the linear term in patient ordering is similar to using all control data (including *NCC*) when the new treatment arm $K+1$ is incorporated into the trial. In other words, the stage 1 control data contributes very little to the estimation of the difference in mean responses of the newly added treatment and the control treatment when we adjust for $\tau_s \cdot c_j$ in the analysis. Besides, the borrowing-of-strength (BoS) function was introduced to evaluate whether *NCC* contribute useful information as

$$BoS = \frac{V_0 - V_a}{V_0} \quad (5)$$

where V_0 and V_a denote the effect variances obtained using only *CC* and all control data, respectively. $BoS \approx 0$ indicates that the estimates from using all control data are indistinguishable from those using only *CC* data, suggesting that *NCC* data provides little benefit to efficacy estimation. In contrast, $BoS \approx 1$ means that using all control data does not introduce negative bias into the efficacy estimation,

and the time trend of *NCC* data can be neglected. Therefore, the closer BoS is to 1, the more supportive it is of utilizing *NCC* data. They demonstrated in a simulation study that regression models adjusting for time trends via a step-wise function produce unbiased tests, even when the actual time trend is linear rather than step-wise. However, only if the model is fit based on data of all trial arms, the inclusion of *NCC* improves the power compared to an analysis using *CC* only.

Based on that, Roig further investigated[18] a step function regression model to adjusts for time by a step function based on effect drifts from all treatment arms

$$g\left(E\left(y_j\right)\right) = \eta_0 + \sum_{k \in K} \theta_k \cdot I\left(k_j = k\right) + \sum_{s=2}^{S} \tau_s \cdot I\left(s_j = s\right) \quad (6)$$

where η_0 denotes the treatment effect of the control arm in the first stage, θ_k represents the effect of treatment k relative to the control arm, τ_s denotes a step-wise time effect between stage 1 and s correspondingly with an indicator function $I(\cdot)$. $g(.)$ can be identified as a standard linear model for continuous data or logit link function for binary data and even including an interaction effect for a different effect of time in various of treatment arms. The simulation study based on the continuous and binary endpoints showed that for fixed-sample platform trials with a block randomization design, the stepwise function regression models adjusting for time trends can provide valid treatment effect estimates and asymptotically control the type I error rate only under the assumptions of equivalent time trends across all treatment arms and correct specification of the time trend scale in the model.

Spine regression

To overcome the limitation of the above model, which assumes a constant time trend within each time stage or calendar time interval, Krotka et al.[19] considered estimating the patient response over time using spline regression as

$$g\left(E\left(y_j\right)\right) = \eta_0 + \sum_{k \in K} \theta_k \cdot I\left(k_j = k\right) + f(t_j) \quad (7)$$

The time trend is modeled via a continuous smooth function $f\left(t_j\right)$ of the patient entry time t_j. Specifically, B-spline function with multiple polynomial segments of a specified degree q was conducted to model time trends, in which connected at points known as knots, ensuring that the entire

spline remains continuously differentiable up to the $(q-1)$th derivative. For example. common choices for the polynomial degree include linear $(q=1)$, quadratic $(q=2)$, and cubic $(q=3)$ splines, offering a flexible pattern to balance the flexibility and smoothness. As to the positions of these knots, two strategies analogous can be evaluated based on the periods or calendar-time intervals. The inner knots can be placed at the beginning of each period s = 2, …, S, so that one polynomial of degree q is fitted within each period. Alternatively, they can be positioned at equidistant intervals along the calendar-time axis, thus fitting a polynomial of degree q to each calendar-time interval.

Krotka[19] examine linear, quadratic, and cubic splines corresponding to $\in (1,2,3)$. For a formal definition of the B-spline. Compared to the constant time trend adjusting approaches, it provides greater flexibility, enabling more accurate representation of complex temporal patterns.

Bayesian Time Machine (BTM)

To continue the regression strategy, Saville et al.[20] proposed a Bayesian logistic regression with second-order normal dynamic linear model (NDLM) to smooth the time trends, so that *NCC* data are incorporated as prior information via Bayesian hierarchical structure

$$E(y_j) = \mu_j \tag{8}$$

$$g(\mu_j) = \eta_0 + \theta_k + x_j'\beta + \alpha_{t(j)} \tag{9}$$

Where $g(\mu_j)$ is a link function of mean treatment effect μ_j, η_0 denotes the model intercept, θ_k represents the mean difference of the treatment effects between treatment arm k (corresponding to subject j) and the control, x_j is a vector of covariates related specific to subject j with corresponding parameter β, and $\alpha_{t(j)}$ represents the increment of μ_j in time stage t. Based on the Bayesian hierarchical model (BHM), the prior distributions for the linear predictor parameters follow

$$\eta_0 \sim N(0, \sigma^2) \tag{10}$$

$$\theta_k \sim N(0, \sigma_\theta^2) \tag{11}$$

$$\beta \sim N\left(0, \sum \quad\right) \tag{12}$$

These prior distributions will typically be non-informative prior distributions that depend on the scale of the data and the chosen link function. To capture the time drift with parameter $\alpha_{t(i)}$, we can divide the total trial duration into *T* consecutive intervals, referred as "buckets", counting backward from the most recent interval with $t = 1$ to the beginning of the trial as$(t + 1, t + 2, \dots, T)$ with a reverse chronological order of intervals. Thus, defining $\alpha_1 = 0$ corresponding to the current phase $t = 1$ and α_t rolling forward to subsequent time intervals, the time parameter is modeled with the following Bayesian second-order NDLM.:

$$\alpha_2 \sim N(0, 1/\tau) \tag{13}$$

$$\alpha_t \sim N(2\alpha_{t-1} - \alpha_{t-2}, 1/\tau), t \geq 3 \tag{14}$$

where τ denotes drift parameter the degree of smoothing over the time intervals, for which the following hyperprior distribution as $\tau \sim Gamma(a, b)$.

Dynamic power prior(DPP)

The core idea of the DPP is to treat the degree of information borrowing from *NCC* as a data-driven random variable, rather than a pre-specified fixed value. This method introduces a power parameter $\alpha \in [0,1]$ to regulate the weight of *NCC* data in the posterior estimation of treatment effects, with α itself inferred as an unknown parameter. *NCC* data are used to construct the power prior via the power parameter α, whereas *CC* data enter the likelihood; the posterior is proportional to the product of the *CC* likelihood and the *NCC*-based power prior.

The DPP is constructed based on the normalized power prior framework[31]:

$$Pr(\theta_k, \alpha \mid NCC) \propto \frac{1}{C(\alpha)} L(\theta_k \mid NCC)^{\alpha} \cdot \pi(\theta_k) \cdot \pi(\alpha) \tag{15}$$

where $L(\theta_k \mid NCC)$ is the likelihood function of the NCC data; $\pi(\theta_k)$ and $\pi(\alpha)$ represent the prior distributions for θ_k and α, respectively. Specifically, α follows a uniform distribution $\alpha \sim U(0,1)$. For the treatment effect parameter θ_k, vague (non-informative) priors are specified: $\theta_k \sim N(0, 10^2)$ for continuous outcomes and $\theta_k \sim Beta(1,1)$ for binary outcomes. $C(\alpha)$ is a normalization constant that ensures the posterior distribution adheres to the likelihood principle. With the *CC* data, the joint posterior distribution of θ_k and α can be captured by

$$Pr(\theta_k, \alpha \mid NCC, CC) \propto L(\theta_k \mid CC) \cdot L(\theta_k \mid NCC)^{\alpha} \pi(\theta_k)\pi(\alpha) \quad (16)$$

The posterior estimation of α directly reflects the consistency between *NCC* and *CC* data. When the two datasets are highly consistent (without significant temporal drift), the posterior distribution of α concentrates around 1, indicating full borrowing of *NCC* information; conversely, when there is substantial discrepancy between the two (e.g., heterogeneity caused by temporal drift), the posterior distribution of α tends toward 0, effectively down-weighting the influence of *NCC* to avoid the introduction of bias and Type I error inflation. Thus, the estimation of α is fully data-driven by sampling from the joint posterior distribution using Markov chain Monte Carlo (MCMC) algorithms, ultimately enabling dynamic adjustment of information borrowing from *NCC* data.

MAP prior

The MAP prior approach[26] quantifies between-trial heterogeneity by assuming exchangeability between historical and current control parameters. Based on a Bayesian hierarchical model, it derives a prior that enables the prediction of the concurrent control effect from historical data. Assuming there are k stages *NCC* data $Y_K = \{y_1, \ldots, y_k\}$ with parameters $\theta_K = \{\theta_1, \ldots, \theta_k\}$ and sample size $n_1, \ldots, n_k$, the *CC* data for the new arm $k+1$ are similar with the historical control data Y_K with exchangeable parameters $\theta = g(\psi)$ with the structure of a hierarchical model as

$$y_k|\theta_k \sim F(\theta_k, n_k) \quad (17)$$

$$\theta_k|\psi \sim G(\psi), \psi \sim P \quad (18)$$

where $k \in K = \{1, \ldots, k, k+1\}$, and *F*, *G*, *P* are the sampling, exchangeability (random-effects), and hyper-prior distribution, respectively. Inference for the control parameter ψ in platform trial will be based on direct (y_{k+1}) and indirect evidence (Y_K) with MAP prior, in which $p(\psi_{k+1}|Y_K)$ derived from *NCC* and then combined with the *CC* data y_{k+1} via Bayes' framework $p(\psi_{k+1}|y_{k+1}, Y_K) \propto p(y_{k+1}|\psi_{k+1})p(\psi_{k+1}|Y_K)$.

Moreover, considering a possibility of prior-data conflict of the prior information to render the exchangeability assumption for the control parameters plausible, Schmidli[26] introduced a vague conjugate prior $p_v(\psi_{k+1})$ to robust MAP prior

$$\hat{p}_{RMAP} = (1 - w_R)\hat{p}_{MAP} + w_R p_v(\psi_{k+1}) \quad (19)$$

where w_R the prior probability that the *CC* differs systematically from the *NCC* data.

EMAP prior

Although R-MAP strives to robust the conflict between historical information and current data as much as possible, due to the accumulating nature of platform trial data, the consistency between *NCC* and *CC* data is dynamically changeable. Such full information borrowing is reasonable when *NCC* prior is congruent with concurrent data, but inappropriate when they are incongruent influenced by time trends, which lead to inflated type I errors or reduced power. To address this issue, Zhang[30] proposed EMAP method with an elastic function $g(T)$, which dynamically adjusts the extent of information borrowing based on the consistency between external data and current trial data.

Elastic function $g(T)$ serves as a link function that maps congruence measure T to an information discount factor. Any monotonic function could be used as an elastic function, as long as $g(T) \to 1$ when the value of T corresponds to congruence and $g(T) \to 0$ when the value of T corresponds to incongruence, such as a logistic function

$$g(T) = \frac{1}{1 + \exp\{a + b \times \log(T)\}} \tag{20}$$

where a and $b > 0$ are tuning parameters identified by the grid search approaches calibrated between the congruence and incongruence case.

Such that, $g(T) \in [0,1]$ can serves as a variance inflation factor to integrate with MAP prior to control the information borrowing. Taking binary data for example, the MAP prior follows a beta distribution $p(\psi_{k+1}|Y_K) = Beta(\alpha, \beta)$, EMAP prior can be defined as $p(\psi_{k+1}|Y_K) = Beta(\alpha g(T), \beta g(T))$. Therefore, the elastic function is constructed to proactively control the behavior of information borrowing, such that it induces strong information borrowing when *NCC* and *CC* are congruent, and little information borrowing when *NCC* and *CC* are incongruent.

Simulation

We evaluate the properties of the methods proposed in above sections via simulation studies to discuss the influence of certain time trends, enrollment patterns and design parameters for the binary and continuous endpoints.

Scenario Designs

Suppose there is a platform trial with $K(K \geq 2)$ experimental treatment arms compared to a shared control arm, where these experimental arms enter the trial sequentially with uniform recruitment. In the simulations each patient j being enrolled at each calendar time t_j, so the adding timing of new treatment can be expressed by the cumulative sample size. That is, new arm k $(k = 1, \ldots, K)$ enters the platform after $d_k = d(k-1)$ participants have been recruited to the trial and $d_1 = 0$, where d_k indicates the amount of overlapping sample size between the experimental arms with equidistant entry times of the experimental treatment arms. We assume all experimental arms were equal sized with sample sizes n_{ks} =250, equal allocation among control and treatment in each stage as well as block randomization per period. So the sample size for the control arm depending on the different entry pattern and overlapping between experimental arms.

For the data generation with time trends, we simulate the trial endpoints similar to Roig's[18] function (1) as

$$E(y_j) = \eta_0 + \sum_{k \in K} \theta_k \cdot I(k_j = k) + f(t_j) \tag{21}$$

We assumed effect sizes of $\theta_k = 0.25$ for the treatment-control comparisons and a response of 0 in the control for continuous endpoint and a logit function for binary endpoint with the control response rate of 0.5 and $OR = 0.5$ under the alternative hypothesis. All experimental treatments are set as ineffective and effective under the null and alternative hypotheses, respectively. With these settings, the chosen sample and effect sizes yield 80% power for the treatment-control comparison based on separate concurrent controls only at α = 2.5% (one-sided).

Considering the time trends $f(t_j)$, three patterns are simulated with a parameter λ_k denoting the strength of the time trend as some related reference [18,19,32]:

- **Linear trend**: The mean response in arm k linearly increases with the slope λ_k over time as $f(j) = \lambda_k \cdot \frac{j-1}{N-1}$.
- **Stepwise trend**: Each time a new treatment arm k adding to the trial, the mean response in that arm increases by a jump of size λ_k as $f(j) = \lambda_{k_j} \cdot (c_j - 1)$, where c_j is the number of

experimental treatment arms have already entered the ongoing trial when participant j was enrolled.

· **Inverted U-shaped trend**: The mean response in arm k increases linearly with slope $\lambda_k(\lambda_k > 0)$ until the total sample size reaches N_p, then decreases linearly thereafter. Conversely, it decreases until N_p and increases afterwards for $\lambda_k < 0$. That is, $f(j) = \lambda_k \cdot \frac{j-1}{N-1}$ for $j \leq N_p$ and $f(j) = -\lambda_k \cdot \frac{j-Np}{N-1} + \lambda_k \cdot \frac{N_p-1}{N-1}$, for $j > N_P$, where N_P denotes the sample size where the trend reverses from positive to negative.

To investigate the influence of enrollment patterns along with overlapping between treatment arms, three trial scenarios are considered as follow.

Setting 1: Homogeneous time trends with equidistant entry times between arms. Firstly, we explored the simplest scenario focusing on the operating characteristics of different models under various of time trends. The simulation involved $K = 3$ experimental treatment arms, where each arm k entered the trial after recruiting $d_k = d \cdot (\mathrm{k} - 1)$ participants in the platform. To examine varying overlaps across treatment arms, d was considered three options as $d = 125, 250 \text{ and } 500$. Note that when $d = 500$, a new treatment arm begins enrollment only after the previous arm has finished, resulting in no overlap. All the arms (including control arms) share the same time trend with the strength of the time trend λ_k ranged from -0.15 to 0.15 in increments of 0.075. The evaluation emphasized the efficacy of the new adding arm 3 relative to the control group.

Setting 2: Homogeneous time trends but non-equidistant entry times between treatment arms. In this scenario, we focus on the impact of varying enrollment times and the overlapping among trial arms influence the issue of data borrowing in *NCC*. Assuming there are $K = 4$ treatment arms with the same time trends, arm 2 is introduced after $d_2 = 300$ participants have been recruited, and arm 4 enters after $d_4 = 800$ participants. While for the introduction timing of treatment arm 3, different entry points $d_3 = 300, 425, 550, 675, \text{and } 800$ are considered with increments of 125. Finally, we evaluate the treatment efficacy of arm 4 compared to the control.

Setting 3: Equidistant entry times but heterogeneous time trends between treatment arms.

As for the heterogeneous of the time trends across arms, we considered three conditions with equidistant entry times $d = 250$:

Setting 3.1: The time trends of the original experimental arm 1 differs from all the other arms

($\lambda_0 = \lambda_2 = \lambda_3 \neq \lambda_1$). In practice, the original experimental arm may be introduced during an early operational learning phase, when investigators are still gaining experience with dose adjustment, toxicity management, endpoint assessment, or treatment administration. It may also reflect treatment-specific formulation or manufacturing changes, limited initial site availability, or changes in treatment-specific supportive care. Although the heterogeneous trend arm is not the target treatment being evaluated, this scenario assesses whether an atypical historical experimental arm can indirectly affect the borrowing of NCC borrowing for a later treatment comparison.

Setting 3.2: Only the control arm exits time trend ($\lambda_0 \neq 0, \lambda_1 = \lambda_2 = \lambda_3 = 0$). This setting represents control-specific temporal drift, with stable outcomes in the experimental arms but changing outcomes in the control arm. Such a situation may arise when the standard-of-care regimen is modified during the platform trial, when new background or supportive therapies become available, or when diagnostic procedures, patient selection, adherence, or follow-up practices improve over time. For example, a platform trial conducted over several years may experience a change in clinical guidelines that improves outcomes among patients receiving standard care. Historical control patients enrolled before this change may therefore differ systematically from concurrent controls, creating direct non-exchangeability between *NCC* and *CC* data.

Setting 3.3: The time trends in the control arm differs from the experimental arms ($\lambda_0 \neq \lambda_1 = \lambda_2 = \lambda_3$). This setting represents nonparallel temporal trends, in which both control and experimental outcomes change over time but at different rates or in different directions. This may occur when changes in background therapy interact differently with the investigational treatment and standard care, when earlier diagnosis benefits the experimental treatment more than the control treatment, or when expansion to less experienced sites affect administration of a complex new therapy more strongly than delivery of standard care. Such a pattern reflects a treatment-by-time interaction and challenges methods that assume a common time trend across all treatment arms.

Comparison approaches

To evaluate the impact *NCC* borrowing on the effect inference in platform trial, two reference benchmark approaches were considered. The most conservative approach, referring as *CC*-separate analysis, uses only control data that are fully concurrent with the new treatment arm for the treatment effect estimation. While this method controls type I error, it requires a larger sample size. In contrast,

the most optimistic approach compares the new treatment arm with a pooled sample of all control data (both *CC* and *NCC*). Although this approach substantially reduces the required sample size, it may introduce estimation bias and inflate type I error in the presence of time trends.

The other approaches to be evaluated including frequentist step-function regression model[18], spline regression model[19], BTM approaches[20] for the regression adjust strategies and DPP[31], MAP[26], EMAP[30] approaches as the Bayesian information discounting strategies. The detail parameters for each model and scenarios are summarized in Table 1 as shown in Figure 2 for binary endpoint and in Figure S1 for continuous endpoints, in which detailed selections for these tuning parameters and Bayesian priors are provided in Additional file 1-2. We simulated 5000 replicates of each scenario to compare the Type I error control, power, estimation bias, and mean squared error (MSE) across these approaches according to whether they could improve efficiency over separate analysis without exhibiting the inferential distortion observed under unrestricted pooling.

Table 1 Parameter considerations of the three scenarios in the simulation study

Study Type	K	d_i	λ
Setting 1: Equidistant enrollment and homogeneous time trends	3	$d_i = d \times (i-1)$ $d = 125,\ 250,\ 500$	$\lambda_0 = \lambda_1 = \lambda_2 = \lambda_3 \in [-0.15, 0.15]$(step 0.0375)
Setting 2: Non-equidistant enrollment and homogeneous time trends;	4	$d_1 = 0, d_2 = 300, d_4 = 800, d_3 \in [300,800]$ with increments of 125	All arms: $\lambda_k = \lambda_0 = 0.15$ (for any k)
Setting 3.1 Heterogeneous trends between the original treatment arm 1 and other arms	3	$d_i = 250 \times (i-1)$	$\lambda_1 \in [-0.05, 0.25]$ (step 0.0375); $\lambda_0 = \lambda_2 = \lambda_3 = 0.1$
Setting 3.2 Time trend only in the control arm	3	$d_i = 250 \times (i-1)$	$\lambda_0 \in [-0.05, 0.25]$ (step 0.0375); $\lambda_1 = \lambda_2 = \lambda_3 = 0$
Setting 3.3 Heterogeneous trends between the control arm and other arms	3	$d_i = 250 \times (i-1)$	$\lambda_0 \in [-0.05, 0.25]$ (step 0.0375); $\lambda_1 = \lambda_2 = \lambda_3 = 0.1$

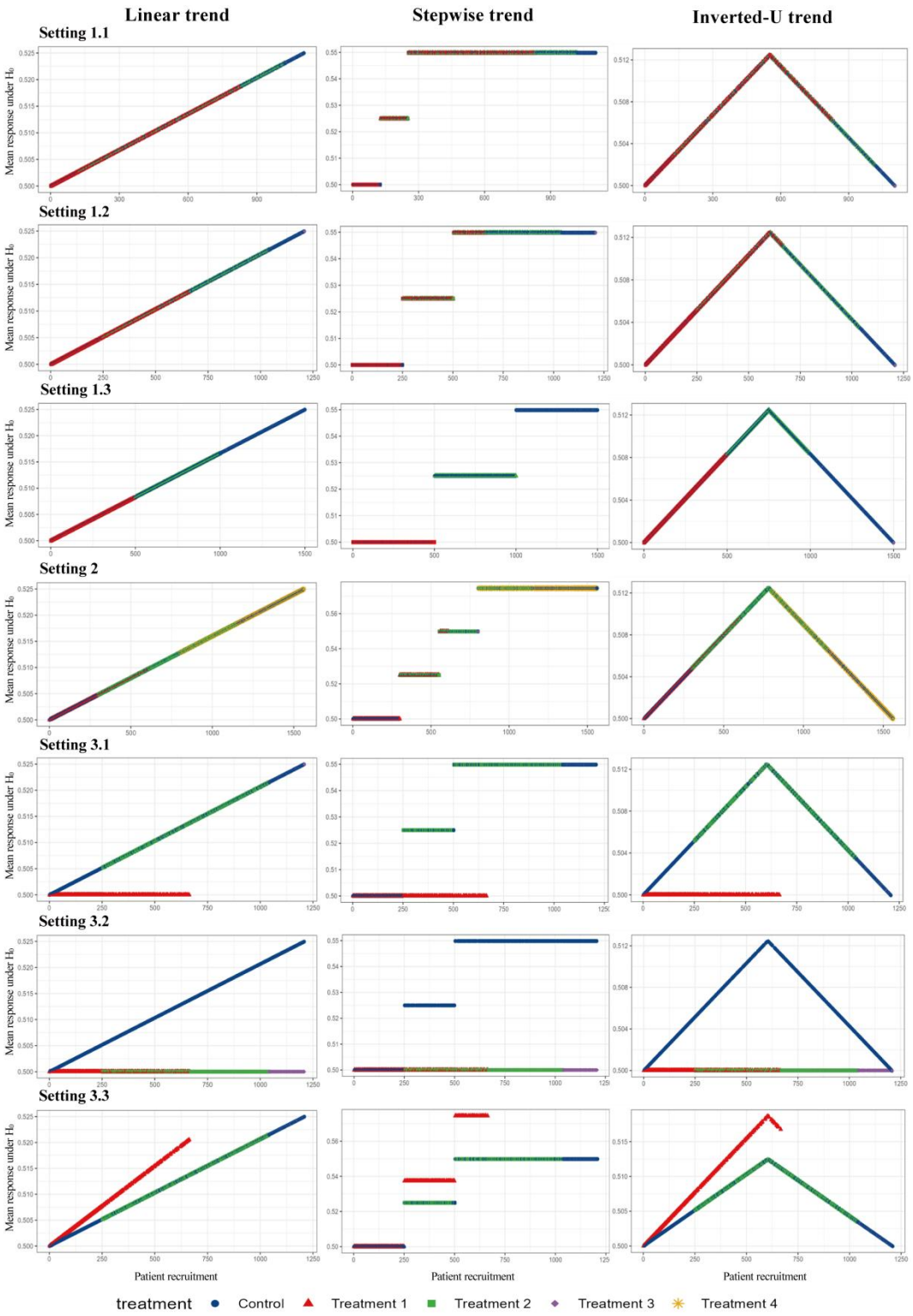

Figure 2　The simulation time trends scenarios for binary endpoint

（Setting 1.1-1.3: Equidistant enrollment and homogeneous time trends for $d = 125,250\ and\ 500$ time, Setting 2: Non-equidistant enrollment and homogeneous time trends; Setting 3.1: Heterogeneous trends between the original treatment arm 1 and other arms; Setting 3.2:Time trend only in the control arm;and Setting 3.3:Heterogeneous trends between the control arm and other arms)

Results

Across all simulation settings, the performance of methods for *NCC*s depended primarily on two factors: whether the time trends were homogeneous across treatment and control arms, and the shape and magnitude of the trends. Simulation results for binary endpoints are presented in Figure 3 and Table 2 (averaged operating-characteristic metrics) and continuous endpoint in Figure S2 and Table S2. Full stratified raw simulation data across all scenarios are provided in Additional file 3.

Setting 1: Equidistant enrollment and homogeneous time trends

We firstly considered the simple scenario in which all treatment and control arms shared the same time trends and new treatment arms entered the platform at equal intervals. The simulations showed that borrowing *NCC* was generally beneficial under homogeneous time trends, but performance differed substantially by the shape of the temporal drift and the analytical method. Across homogeneous linear and inverted-U trends, all time-adjusted methods improved power relative to *CC*-separate analyses and maintained type I error close to the nominal one-sided 2.5% level for both binary and continuous endpoints, regardless of entry spacing. For example, EMAP achieved approximately 0.9060-0.9874 power for binary outcomes and 0.7780-0.8175 for continuous outcomes, compared with approximately 0.7850-0.8155 for separate analysis. Even naive pooling showed little inflation in these settings, reflecting the fact that common temporal effects largely cancelled out in treatment-control comparisons.

Under homogeneous stepwise trends, however, performance differences became more pronounced as the gap between arm entries increased. The main differences emerged under stepwise trends and long treatment-entry intervals. For binary outcomes, when $d = 500$ and $\lambda = 0.15$, Type I error increased to approximately 0.1802 for pooled analysis, 0.1302 for DPP, and 0.1474 for EMAP. In contrast, MAP, regression adjustment, spline regression, and the BTM remained closer to the nominal level. For continuous outcomes, pooled analysis was even more sensitive, with Type I error reaching approximately 0.5440 under the strongest stepwise trend, whereas EMAP remained close to the nominal level.

The increase in apparent power observed for pooled analysis and DPP under severe stepwise drift was accompanied by substantial bias. At $d = 500$ and $\lambda = 0.15$, pooled analysis showed bias

Table 2 The operating characteristics of *NCC* borrowing methods for time trends in platform trials with binary endpoints

Method	Linear				Stepwise				Inverted U-shaped			
	Type I error	Power	Bias	MSE	Type I error	Power	Bias	MSE	Type I error	Power	Bias	MSE
Setting 1：Equidistant enrollment and homogeneous time trends												
CC-separate	0.0271	0.9020	0.0027	0.0339	0.0269	0.8999	0.0025	0.0340	0.0276	0.9022	0.0016	0.0336
Regression	0.0262	0.9151	0.0041	0.0323	0.0260	0.9130	0.0039	0.0325	0.0262	0.9148	0.0030	0.0321
Spline	0.0253	0.9212	0.0058	0.0316	0.0258	0.9195	0.0058	0.0319	0.0257	0.9209	0.0048	0.0314
BTM	0.0275	0.9212	0.0272	0.0354	0.0279	0.9197	0.0271	0.0356	0.0280	0.9213	0.0260	0.0351
MAP	0.0219	0.9225	0.0021	0.0301	0.0236	0.9197	0.0032	0.0307	0.0222	0.9238	0.0013	0.0296
EMAP	0.0259	0.9471	0.0041	0.0280	0.0356	0.9359	0.0040	0.0319	0.0250	0.9477	0.0039	0.0275
DPP	0.0257	0.9505	0.0024	0.0272	0.0319	0.9431	0.0044	0.0304	0.0236	0.9494	0.0004	0.0269
Pooled	0.0267	0.9560	0.0023	0.0269	0.0389	0.9472	0.0051	0.0315	0.0253	0.9549	-0.0010	0.0263
Setting 2: Homogeneous time trends with non-equidistant entry times												
CC-separate	0.0277	0.8938	0.0021	0.0345	0.0250	0.8651	0.0019	0.0377	0.0289	0.9005	0.0018	0.0339
Regression	0.0271	0.9147	0.0036	0.0322	0.0261	0.8910	0.0043	0.0353	0.0259	0.9201	0.0040	0.0313
Spline	0.0261	0.9197	0.0052	0.0320	0.0295	0.9051	0.0163	0.0353	0.0264	0.9228	0.0053	0.0312
BTM	0.0290	0.9217	0.0262	0.0353	0.0291	0.9030	0.0299	0.0387	0.0286	0.9270	0.0264	0.0344
MAP	0.0284	0.9380	0.0191	0.0293	0.0453	0.9140	0.0555	0.0395	0.0200	0.9314	0.0006	0.0283
EMAP	0.0292	0.8941	0.0039	0.0348	0.0270	0.8651	0.0062	0.0381	0.0276	0.9047	0.0040	0.0341
DPP	0.0401	0.9703	0.0356	0.0278	0.1082	0.9793	0.1170	0.0458	0.0246	0.9575	0.0006	0.0256
Pooled	0.0498	0.9816	0.0488	0.0279	0.2301	0.9982	0.1928	0.0655	0.0254	0.9660	0.0001	0.0247
Setting 3.1: Heterogeneous time trend in the original experimental arm 1												
CC-separate	0.0268	0.8978	0.0031	0.0344	0.0269	0.8859	0.0030	0.0355	0.0273	0.9028	0.0022	0.0338
Regression	0.0248	0.9198	0.0055	0.0319	0.0266	0.9117	0.0063	0.0332	0.0257	0.9232	0.0037	0.0315
Spline	0.0253	0.9238	0.0076	0.0319	0.0274	0.9182	0.0125	0.0331	0.0258	0.9264	0.0055	0.0315
BTM	0.0276	0.9263	0.0287	0.0352	0.0292	0.9201	0.0311	0.0366	0.0283	0.9290	0.0268	0.0346
MAP	0.0246	0.9227	0.0092	0.0309	0.0293	0.9235	0.0230	0.0322	0.0222	0.9220	0.0017	0.0298
EMAP	0.0303	0.9024	0.0049	0.0346	0.0282	0.8926	0.0067	0.0350	0.0281	0.9035	0.0033	0.0343

DPP	0.0314	0.9570	0.0196	0.0285	0.0464	0.9614	0.0483	0.0321	0.0254	0.9495	0.0039	0.0276
Pooled	0.0349	0.9696	0.0291	0.0278	0.0664	0.9832	0.0775	0.0339	0.0255	0.9590	0.0048	0.0266
Setting 3.2: Time trend present only in the control arm												
CC-separate	0.0159	0.7880	0.0002	0.0341	0.0192	0.5423	0.0032	0.0344	0.0200	0.8734	0.0050	0.0343
Regression	0.0146	0.8306	0.0090	0.0317	0.0171	0.6007	0.0282	0.0333	0.0187	0.8955	0.0076	0.0316
Spline	0.0146	0.8332	0.0109	0.0316	0.0174	0.6063	0.0316	0.0335	0.0188	0.8987	0.0094	0.0316
BTM	0.0163	0.8408	0.0296	0.0349	0.0189	0.6124	0.0453	0.0367	0.0205	0.9035	0.0298	0.0349
MAP	0.0124	0.8288	0.0080	0.0303	0.0153	0.5861	0.0215	0.0322	0.0162	0.8954	0.0055	0.0303
EMAP	0.0165	0.8928	0.0212	0.0284	0.0175	0.6816	0.0475	0.0352	0.0164	0.9340	0.0101	0.0274
DPP	0.0138	0.8796	0.0159	0.0283	0.0158	0.6548	0.0420	0.0326	0.0186	0.9247	0.0062	0.0278
Pooled	0.0152	0.9111	0.0254	0.0277	0.0154	0.7541	0.0767	0.0387	0.0189	0.9382	0.0071	0.0266
Setting 3.3: Different trend magnitudes in control and treatment arms												
CC-separate	0.0379	0.8739	0.0010	0.0341	0.1100	0.7694	0.0035	0.0352	0.0277	0.9022	0.0048	0.0335
Regression	0.0340	0.8996	0.0032	0.0318	0.0947	0.8083	0.0054	0.0334	0.0262	0.9211	0.0062	0.0310
Spline	0.0345	0.9032	0.0050	0.0318	0.0975	0.8172	0.0115	0.0335	0.0267	0.9232	0.0080	0.0310
BTM	0.0376	0.9072	0.0264	0.0349	0.0998	0.8187	0.0299	0.0365	0.0291	0.9268	0.0294	0.0343
MAP	0.0349	0.9064	0.0091	0.0304	0.1080	0.8145	0.0216	0.0331	0.0234	0.9235	0.0052	0.0296
EMAP	0.0414	0.8782	0.0072	0.0349	0.1139	0.7692	0.0052	0.0361	0.0293	0.9032	0.0042	0.0341
DPP	0.0400	0.9407	0.0176	0.0284	0.1206	0.8743	0.0420	0.0335	0.0256	0.9482	0.0057	0.0272
Pooled	0.0438	0.9606	0.0276	0.0279	0.1271	0.9458	0.0769	0.0398	0.0260	0.9578	0.0063	0.0262

Note: All values are averaged within each trend type (linear, stepwise, and inverted U-shaped) for each simulation setting in Table 1. That is, for each method and each trend type, values are averaged over all parameter combinations specified for that setting. Type I error is evaluated under the null hypothesis; Power, Bias, and MSE are evaluated under the alternative hypothesis. Results are reported for the newly added treatment arm versus the control arm. Detailed parameter settings are provided in Table 1, and stratified results are provided in Additional file 2.

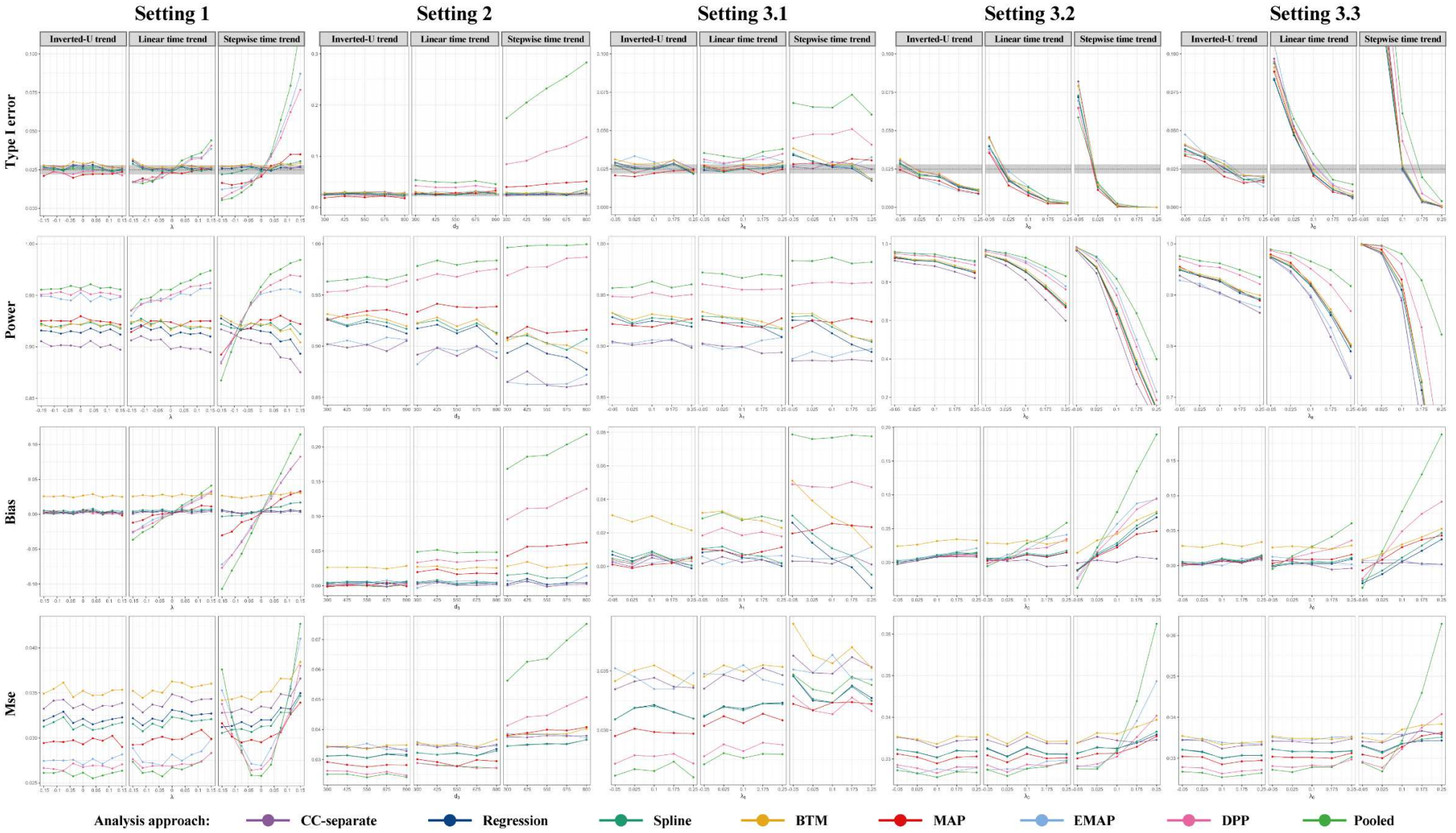

Figure 3 The operating characteristics of methods for *NCC* borrowing across the simulation scenarios with binary endpoints

exceeding 0.1500 for both binary and continuous outcomes. EMAP and the other more conservative methods generally kept bias within approximately 0.0200. For continuous outcomes, pooled analysis reached MSE values of approximately 0.0287, whereas EMAP generally remained at approximately 0.0084.

Thus, under homogeneous trends, *NCC* borrowing was generally feasible. Smooth trends represented a favorable borrowing region, whereas abrupt trends combined with long entry intervals required more conservative or adaptive methods.

Setting 2: Homogeneous time trends with non-equidistant entry times

We next examined a more realistic platform trial configuration in which all arms still shared the same time trend, but treatment entry times were staggered unequally. This allowed us to assess whether irregular overlap between treatment arms affected the reliability of *NCC* borrowing. These simulation results indicate that, when time trends are homogeneous, the pattern of entry times is less important than the trend shape itself. Irregular arm entry had comparatively little effect under smooth trends, but under stepwise trends it magnified the consequences of inappropriate borrowing.

Similar to setting 1, the non-equidistant design itself had only a limited impact under linear and inverted-U drift, with good type I error control, negligible bias, and stable MSE across for all time-adjusted methods, which suggests that irregular enrollment spacing alone does not undermine *NCC* borrowing when the underlying trend is shared across arms. However, the main challenge arose under stepwise drift again. As the delay before entry of arm 3 increased, the incompatibility between *CC* and *NCC* became more consequential for methods that borrowed too aggressively. Simple pooling showed the clearest deterioration, with severe type I error inflation to 0.8080, rapidly increasing bias, and substantial loss of estimation accuracy as reflected by MSE, and the same pattern was observed for DPP of 0.1300 although less dramatically than pooling. MAP and spline regression showed mild degradation in the most extreme cases, particularly for binary outcomes. In contrast, EMAP, regression adjustment, and BTM remained essentially unbiased with stable corresponding MSE within type I error of approximately 0.0185-0.0312(Figure 3, Figure S2).

Therefore, a notable finding in this setting is that irregular arm entry per se was not the primary driver of poor performance. Rather, the key determinant was whether the time trend was abrupt enough

to create meaningful incompatibility between *NCC* and concurrent controls. Under smooth homogeneous trends, all well-adjusted methods performed similarly to those in setting 1; under stepwise drift, methods differed sharply in robustness. This reinforces that borrowing decisions should be driven more by the likely shape of temporal drift than by the arm-entry pattern alone.

Setting 3.1: Heterogeneous time trend in the original experimental arm 1

Setting 3 introduced heterogeneous time trends across arms. In Setting 3.1, only the original experimental arm had a different trend, whereas the control arm and the other experimental arms followed the same trajectory. The simulation results showed that a localized departure from homogeneity in one historical treatment arm did not substantially undermine most methods, especially for smooth time trends. For both binary and continuous endpoints, most methods retained reasonably good control for type I error, bias and MSE under linear and inverted-U trends, with only minor deviations from nominal levels. Even under stepwise drift, the degradation was modest: pooling showed some increase in type I error and bias, while DPP and BTM exhibited mild instability, but EMAP, MAP prior, regression adjustment, and spline regression remained essentially robust(Figure 3, Figure S2). For example, although EMAP was somewhat more conservative with power approximately 0.7940-0.8075 compared to MAP prior and spline regression of 0.8470 and 0.8810, its MSE remained approximately within 0.0076-0.0082, as seen in Additional file 3.

Thus, when heterogeneity is limited to a treatment arm that is not the direct comparator of interest, borrowing *NCC*s can still be effective, provided that a reasonably robust adjustment method is used. This setting suggests that not all forms of trend heterogeneity are equally problematic; the most critical concern is heterogeneity involving the control trajectory.

Setting 3.2: Time trend present only in the control arm

This was the most practically important and most challenging heterogeneous setting that the control arm remains open over long periods while new treatment arms enter sequentially. Here, only the control arm exhibited temporal drift, while all experimental arms had no time trend. Because *NCC* borrowing directly relies on earlier control observations, this scenario violates the core homogeneity assumption underlying most time-adjustment approaches and thus created strong incompatibility between *NCC* and CC data directly.

In this setting, performance diverged sharply across methods. Pooling was consistently the worst performer, with pronounced type I error inflation, large bias, and substantial increases in MSE under all three drift types, especially for stepwise trends. The DPP also showed clear inflation and non-negligible bias. And regression adjustment, spline regression, and BTM all became less reliable as the drift magnitude increased (Figure 3, Figure S2). Although some of these methods retained acceptable behavior when the control trend was weak, their operating characteristics deteriorated all developed noticeable bias once the drift became moderate or strong. Relatively, MAP and EMAP prior remained consistently stable throughout this scenario. Across both endpoint types, they preserved type I error near the nominal level, kept bias close to zero, and maintained comparatively low MSE within 0.0072-0.0083 (Additional file 3). At the same time, they retained substantially higher power than the separate analysis. This trade-off was most evident when the control drift was large: separate analysis remained conservative but under powered, whereas MAP and EMAP prior preserved much more power without sacrificing validity.

This scenario therefore identifies the clearest distinction among methods. When temporal drift is confined to the control arm, most *NCC* borrowing methods methods that rely on direct pooling or on explicit equality of time trends across arms become vulnerable, while MAP-based borrowing remains much more robust.

Setting 3.3: Different trend magnitudes in control and treatment arms

This scenario considered a sustained between-arm mismatch in time trends, with both control and experimental arms drifting over time but at different magnitudes. This represents a broader form of heterogeneity than the above settings and again challenges the exchangeability assumptions behind many *NCC* borrowing approaches.

The overall pattern was similar to Setting 3.2 but somewhat less extreme. Pooling again showed clear inflation of type I error, together with increasing bias and MSE as the discrepancy between control and treatment trends widened. DPP, EMAP, regression adjustment, spline regression, and BTM were more stable than pooling but still showed some sensitivity. Although these methods often still improved power relative to separate analysis, the gain was accompanied by a loss of inferential robustness. Bias and MSE increased with the degree of trend mismatch for nearly all methods except

MAP and separate analysis (Figure 3, Figure S2). The increases were mild under linear but inverted-U patterns (especially when the difference in slope was large) became more pronounced.

EMAP prior remained the more stable approach in this setting. Across the full range of drift magnitudes, it kept type I error close to nominal and produced near-unbiased estimates with the lowest or near-lowest MSE among the borrowing methods. Even under the strongest heterogeneity, its bias is approximately 0.0024 with MSE within approximately 0.0323-0.0340. Importantly, it also maintained a clear power advantage of 0.7410 for inverted-U trends and 0.8762 for linear trends, which again paid a substantial efficiency cost for using concurrent controls only (Additional file 3). Therefore, even when both control and treatment arms evolve over time, EMAP prior provided the most favorable bias–variance trade-off and the most reliable operating characteristics overall.

Summary

Taken together, when control and treatment arms shared the same trend form but differed in trend magnitude, the feasibility of *NCC* borrowing depends primarily on whether temporal drift is homogeneous across arms. Mild heterogeneity affecting a non-target treatment arm may be tolerable, but heterogeneity involving the control arm poses a major threat to validity. Based on the considerations of heterogeneous time trends, enrollment pattern, and trend shapes, we propose a practical decision zones framework (Figure 4) for the *NCC* borrowing in platform trial.

	***NCC* borrowing in platform trials**		
	Acceptable	**Conditional**	**Restricted**
Heterogeneity of time trends	Homogeneous time trends across all arms	Mild heterogeneity, or heterogeneity in a non-target treatment arm, while trends directionally consistent across arms	Markedly different or oppositely directed trends
Enrollment pattern	Adequate overlap times between new treatment and control arms, with short and regular enrollment intervals	Some temporal separation or irregular enrollment, while partial overlap remains	Little overlap times with patients enrolling in different historical periods
Time trends	Smooth, continuous, or mildly changing trends, e.g., linear trend or mild inverted-U trend	Mild stepwise or local abrupt shifts that can be described by models	Strong, non-monotonic, long-term stepwise drift, control-specific drift, or oppositely directed trends
Adjust performance	Type I error close to the nominal level, improved power, and low bias and MSE	Performance varies across adjust methods; Type I error, power, bias, and MSE should be evaluated jointly	No method simultaneously maintains acceptable Type I error, power, bias, and MSE
Borrowing strategy	*NCC* borrowing may be incorporated into the primary analysis	*NCC* borrowing be accompanied by *CC*-seperate and alternative sensitivity analyses	*CC*-seperate remains primary analysis with *NCC* borrwing as a sensitivity analysis

Figure 4 A practical decision zones framework for the *NCC* borrowing in platform trial

As the green zone shown in *NCC* borrowing may be considered acceptable when the CC and *NCC* populations are expected to share homogeneous time trends with adequate overlap in treatment entry. The time trends should be smooth or mildly changing without substantial stepwise shifts or evidence of opposite trends between treatment groups. Before implementation, candidate methods,such as EMAP, MAP, regression adjustment, spline-based models, or BTM, should demonstrate adequate operating characteristics under prespecified plausible scenarios compared with CC separate analysis.

For the yellow zone, *NCC* borrowing is possible but requires safeguards when there is partial heterogeneous time trends between treatment groups, irregular treatment entry pattern, or stepwise time changes. In this setting, *NCC* borrowing cannot be justified by power gains alone. Instead, it requires prespecified calibration simulations covering plausible trend shapes, magnitudes of drift, and enrollment patterns. Researchers should report both CC-separate analyses and alternative sensitivity approaches, with final interpretations grounded in the comprehensive profile of Type I error, bias and MSE.

However, *NCC* borrowing should be restricted from serving as the primary analysis particularly high in the red zone when the treatment and control groups may follow different or oppositely directed temporal trends, substantial little time overlapping, or large stepwise time trends oCCur over long enrollment intervals. The CC analysis should remain as primary analysis with *NCC* borrowing as sensitivity analyses, but their results should not replace the concurrent-control comparison.

It is worth noting that the boundaries between the three zones should not be interpreted as universal numerical thresholds. They depend on the endpoint, sample size scales, event rate, enrollment pattern, plausible magnitude and shape of temporal drift, and the assumptions underlying the analytical method. accordingly, the framework is intended to support a calibration-based decision process. Investigators should define plausible temporal-trend scenarios before the trial, evaluate candidate methods under these scenarios, and prespecified the conditions under which *NCC* borrowing will be permitted, restricted, or excluded from the primary analysis.

Discussion

This study evaluated whether and under what conditions non-concurrent controls can be safely used in platform trials. Rather than assuming that *NCC* borrowing is universally beneficial or that one

borrowing model is uniformly superior, we systematically evaluated multiple frequentist and Bayesian approaches under a broad range of temporal drift patterns and treatment-entry patterns across binary and continuous endpoints.

Across all scenarios, one conclusion was consistent: the feasibility of *NCC* borrowing depends primarily on whether time trends are homogeneous between the treatment and control arms. The most favorable results were observed under homogeneous temporal trends. When time drift was shared across arms, several methods, including regression adjustment, spline regression, BTM, and MAP prior, successfully improved efficiency while maintaining acceptable type I error control. In contrast, when time trends are heterogeneous, especially when drift is present in the control arm alone or differs systematically between control and treatment arms, most borrowing strategies become vulnerable to bias and type I error inflation.

A second important finding is that the shape of the time trend matters. Smooth linear and inverted-U trends were generally less problematic than stepwise trends. Stepwise drift consistently created the strongest departures from nominal operating characteristics, especially for pooled analysis and power prior methods. Abrupt changes combined with long treatment-entry intervals created substantial discrepancies between *NCC* and *CC* data. This is methodologically important because abrupt changes in control outcomes may arise in practice from protocol amendments, site activation waves, operational learning curves, or changes in standard of care. Our results therefore suggest that methods appearing robust under smooth drift may still fail when the drift is discontinuous.

In addition, the simulations indicate that entry pattern alone is not the main driver of performance. The difference between equidistant and staggered entry schedules was modest when time trends were homogeneous, but became more consequential under stepwise drift because larger temporal gaps increase the mismatch between older *NCC* observations and the current concurrent control. This implies that the effect of enrollment structure is largely mediated by the extent to which it amplifies exposure to temporal drift, rather than acting independently, demonstrating that the location of heterogeneity and the borrowing mechanism both matters.

The most important limitation of *NCC* borrowing emerged when the control arm had a trajectory different from that of the treatment arms. In the control-only drift setting, no method simultaneously

maintained nominal Type I error, low bias, and adequate power across all directions and magnitudes of drift. Negative or oppositely directed control trends produced Type I error inflation, especially under stepwise patterns. Strong positive control drift instead led to conservative inference and severe power loss. These apparently different failures have the same practical implication: strong control-specific temporal drift can make the historical control data fundamentally non-exchangeable with the current control data. This result is relevant to long-running platform trials. Although EMAP and MAP were often effective in limiting bias and MSE, but low estimation error did not guarantee nominal Type I error or adequate power. Conversely, methods with higher apparent power could achieve that power partly through inflated Type I error. Therefore, the quality of *NCC* borrowing must be assessed using multiple operating characteristics rather than a single metric. This paper only explores continuous and binary data. For survival data, similar findings may be obtained after exponential transformation, but further methodological exploration is still needed.

Our findings align with and extend previous work on *NCC* use in platform trials. Lee and Wason showed that regression models adjusting for time can provide unbiased inference when time trends are common across arms and correctly specified, especially under block randomization and stepwise adjustment assumptions[17]. Roig et al. further emphasized that valid inference using step-function adjustment critically depends on equal time trends across arms and correct model specification[18]. Our results confirm these conclusions but also show that their robustness is limited once this assumption is violated. In particular, under heterogeneous control-arm drift, regression adjustment and spline-based approaches no longer fully protected against bias and type I error inflation, despite performing well in homogeneous settings. This distinction is important because real platform trials may not offer strong evidence that the control trajectory evolves identically to all active treatment arms. This study also complements recent work by Krotka et al., who demonstrated that spline regression can flexibly accommodate smooth temporal changes and may outperform simpler parametric models when drift is nonlinear[19]. We similarly found that spline models worked well under linear and inverted-U trends and remained competitive in homogeneous stepwise settings. However, their performance deteriorated once trend heterogeneity emerged between arms, suggesting that improved flexibility in modeling time does not by itself solve the more fundamental problem of arm-specific temporal non-exchangeability.

Among these time-adjusted methods, regression adjustment and spline regression performed well under homogeneous or mildly heterogeneous trends, particularly when the trend was adequately represented by the model. Their performance deteriorated when the treatment and control arms followed different temporal trajectories. This confirms that flexible modeling of calendar time cannot, by itself, resolve arm-specific non-exchangeability. A more complex time function may model the shape of the drift more accurately, but it does not identify a treatment effect when the time effect differs fundamentally between arms.

For Bayesian regression-type methods, BTM proposed by Saville et al. was one of the best-performing approaches in homogeneous settings[20]. In our simulations, BTM maintained nominal or near-nominal type I error while improving power under common trends, consistent with its intended role as a smoothing model for temporal drift. Nevertheless, its robustness was not universal. When the control arm alone drifted, BTM showed mild-to-moderate inflation and increasing bias as heterogeneity intensified. This suggests that smoothing time effects across all data may be effective when there is a common temporal structure, but less so when the relationship between time and outcome differs fundamentally between arms.

For DPP, our findings help clarify the role of prior discounting approaches. Power priors have been widely studied for integrating historical information, including adaptive variants in which the borrowing weight is estimated from the data [21,22,31]. In principle, this should protect against prior-data conflict. In practice, however, our simulations showed that the DPP remained sensitive to abrupt drift and to control-specific heterogeneity. Even when its performance was better than naïve pooling, it frequently displayed inflation and bias in difficult scenarios. This suggests that a single scalar discounting mechanism may not be sufficient when incompatibility arises from structured temporal drift rather than global disagreement between *NCC* and concurrent controls.

EMAP and MAP were often effective in limiting bias and MSE, but low estimation error did not guarantee nominal Type I error or adequate power. Conversely, methods with higher apparent power could achieve that power partly through inflated Type I error. Therefore, the quality of *NCC* borrowing must be assessed using multiple operating characteristics rather than a single metric.The comparison of EMAP and MAP should be interpreted in this broader context. MAP provided a conventional Bayesian borrowing benchmark, whereas EMAP introduced a more flexible mechanism for adapting

the borrowing strength according to the degree of compatibility between *NCC* and CC data. In several settings, EMAP provided strong control of bias and MSE and showed favorable performance under local heterogeneity or long entry intervals. In other settings, MAP was equally stable or more conservative. Neither method was uniformly superior. The practical choice between them should therefore depend on the anticipated temporal-trend structure, the endpoint type, the degree of prior-data conflict, and the desired balance between power and error control.

This study may have the following highlights. We conduct a clinically motivated stress-test scenarios closely reflecting the operational complexity of platform trials incorporating sequential treatment entry, varying degrees of enrollment overlap, different lengths of entry intervals, and a comprehensive range of clinically plausible temporal-trend patterns. Therefore, a unified methodological comparison of conventional frequentist regression approaches and Bayesian information-discounting strategies for jointly evaluation of validity, efficiency, and estimation accuracy. Finally, we translate the simulation evidence into a practical decision zones framework to guide whether *NCC* borrowing should be acceptable, conditional borrowing and restricted, or used only as a sensitivity analysis in platform trial. This framework is intended to support the pre-specification of *NCC* borrowing strategies and provide practical guidance for statisticians, trialists, and regulatory researchers designing or analyzing platform trials.

However, several limitations should also be noted. Although we focused on relatively stylized temporal structures and equal or prespecified recruitment patterns, real platform trials may involve more complex accrual dynamics, treatment discontinuation, delayed outcomes, or interactions between site effects and calendar time. Besides, the simulations assumed correctly specified outcome models within each method class and did not consider covariate imbalance or missing data, both of which may further complicate *NCC* borrowing. Bayesian approaches may depend to some extent on prior specification and tuning parameters, even though we used reasonable default values informed by the literature, alternative calibrations might yield somewhat different operating characteristics. In addition, we focused on fixed-sample comparisons for newly added arms rather than full decision-theoretic platform incorporating interim adaptation, response-adaptive randomization, or time-to-event outcomes, all of which may interact with *NCC* borrowing in important ways.

Conclusion

In summary, this study shows that robust borrowing of *NCC* in platform trials is possible, but only when the structure of temporal drift is carefully considered. Homogeneous drift permits safe and efficient borrowing by several methods, whereas heterogeneous drift, particularly in the control arm, poses a serious threat to validity. These findings support a cautious, context-dependent approach to *NCC* borrowing and provide practical guidance for the design and analysis of platform trials with potential time trends.

List of abbreviations

BaSe-MAP: Bayesian semiparametric meta-analytic-predictive

BHM: Bayesian hierarchical model

BoS: borrowing-of-strength

BTM: Bayesian Time Machine

CC: concurrent control

DPP: dynamic power prior

EMAP: elastic meta-analytic-predictive

ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

MAP: meta-analytic-predictive

MCMC: Markov chain Monte Carlo

MSE: mean squared error

NCC: non-concurrent control

PS: propensity score

RCT: randomized controlled trial

R-MAP: robust meta-analytic-predictive

NDLM: normal dynamic linear model

Declarations

Consent for publication

Not applicable

Availability of data and materials

All data generated and analysed during this study are included within this article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 82273728) and the Shannxi Province Natural Science Foundation of China (Grant No. 2026JC-YXQN-284).

Authors' contributions

CL proposed the conception, designed the work and drafted the manuscript. BK and SX performed data analysis, results interpretation and figures creation. All the authors reviewed and approved the final manuscript.

Ethics approval and consent to participate: The text is based on methodological simulation research. The case study is a hypothetical scenario, which does not involve actual research on humans, and hence exempt from approval by ethics committees and informed consent.

Clinical trial number

Not applicable.

Additional files

Additional file 1-2.docx contains the parameter calibration and sensitivity analysis of the MAP prior for binary endpoints (Appendix 1), and the calibration of elastic function parameters for the EMAP prior (Appendix 2). Additional file 3.xlsx contains the raw simulation results across all evaluated scenarios.

Acknowledgements

Not applicable

Reference

1. Berry SM, Connor JT, Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. *JAMA*. 2015;313(16):1619-1620. https://doi.org/10.1001/jama.2015.2316.

2. Adaptive Platform Trials Coalition. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat Rev Drug Discov*. 2019;18(10):797-807. https://doi.org/10.1038/s41573-019-0034-3.

3. Park JJH, Harari O, Dron L, Lester RT, Thorlund K, Mills EJ. An overview of platform trials with a checklist for clinical readers. *J Clin Epidemiol*. 2020;125:1-8. https://doi.org/10.1016/j.jclinepi.2020.04.025.

4. Angus DC, Derde L, Al-Beidh F, Annane D, Arabi Y, Beane A, et al. Effect of Hydrocortisone on Mortality and Organ Support in Patients With Severe COVID-19: The REMAP-CAP COVID-19 Corticosteroid Domain Randomized Clinical Trial. *JAMA*. 2020;324(13):1317-1329. https://doi.org/10.1001/jama.2020.17022.

5. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther*. 2009;86(1):97-100. https://doi.org/10.1038/clpt.2009.68.

6. European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol - PubMed. Accessed September 23, 2026. https://pubmed.ncbi.nlm.nih.gov/30782589/

7. Adaptive Global Innovative Learning Environment for Glioblastoma: GBM AGILE - PubMed. Accessed September 23, 2026. https://pubmed.ncbi.nlm.nih.gov/28814435/

8. Pocock SJ. The combination of randomized and historical controls in clinical trials. *J Chronic Dis*. 1976;29(3):175-188. https://doi.org/10.1016/0021-9681(76)90044-8.

9. ICH E10, Choice of Control Group and Related Issues in Clinical Trials [https://database.ich.org/sites/default/files/E10_Guideline.pdf]

10. Jahanshahi M, Gregg K, Davis G, Ndu A, Miller V, Vockley J, et al. The Use of External Controls in FDA Regulatory Decision Making. *Ther Innov Regul Sci*. 2021;55(5):1019-1035. https://doi.org/10.1007/s43441-021-00302-y.

11. Bofill Roig M, Burgwinkel C, Garczarek U, Koenig F, Posch M, Nguyen Q, et al. On the use of non-concurrent controls in platform trials: a scoping review. *Trials*. 2023;24(1):408. https://doi.org/10.1186/s13063-023-07398-7.

12. Yuan Y, Guo B, Munsell M, Lu K, Jazaeri A. MIDAS: a practical Bayesian design for platform trials with molecularly targeted agents. *Stat Med*. 2016;35(22):3892-3906. https://doi.org/10.1002/sim.6971.

13. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat*. 2014;13(1):41-54. https://doi.org/10.1002/pst.1589.

14. Ventz S, Alexander BM, Parmigiani G, Gelber RD, Trippa L. Designing Clinical Trials That Accept New Arms: An Example in Metastatic Breast Cancer. *J Clin Oncol*. 2017;35(27):3160-3168. https://doi.org/10.1200/JCO.2016.70.1169.

15. Ren Y, Li X, Chen C. Statistical considerations of phase 3 umbrella trials allowing adding one treatment arm mid-trial. *Contemp Clin Trials*. 2021;109:106538. https://doi.org/10.1016/j.cct.2021.106538.

16. Chen WC, Wang C, Li H, Lu N, Tiwari R, Xu Y, et al. Propensity score-integrated composite likelihood approach for augmenting the control arm of a randomized controlled trial by incorporating real-world data. *J Biopharm Stat*. 2020;30(3):508-520. https://doi.org/10.1080/10543406.2020.1730877.

17. Lee KM, Wason J. Including non-concurrent control patients in the analysis of platform trials: is it worth it? *BMC Med Res Methodol*. 2020;20(1):165. https://doi.org/10.1186/s12874-020-01043-6.

18. Bofill Roig M, Krotka P, Burman CF, Glimm E, Gold SM, Hees K, et al. On model-based time trend adjustments in platform trials with non-concurrent controls. *BMC Med Res Methodol*. 2022;22(1):228. https://doi.org/10.1186/s12874-022-01683-w.

19. Krotka P, Posch M, Gewily M, Höglinger G, Bofill Roig M. Statistical Modeling to Adjust for Time Trends in Adaptive Platform Trials Utilizing Non-Concurrent Controls. *Biom J*. 2025;67(3):e70059. https://doi.org/10.1002/bimj.70059.

20. Saville BR, Berry DA, Berry NS, Viele K, Berry SM. The Bayesian Time Machine: Accounting for temporal drift in multi-arm platform trials. *Clin Trials*. 2022;19(5):490-501. https://doi.org/10.1177/17407745221112013.

21. Neuenschwander B, Branson M, Spiegelhalter DJ. A note on the power prior. *Stat Med*. 2009;28(28):3562-3566. https://doi.org/10.1002/sim.3722.

22. Gravestock I, Held L, COMBACTE-Net consortium. Adaptive power priors with empirical Bayes for clinical trials. *Pharm Stat*. 2017;16(5):349-360. https://doi.org/10.1002/pst.1814.

23. Banbeta A, van Rosmalen J, Dejardin D, Lesaffre E. Modified power prior with multiple historical trials for binary endpoints. *Stat Med*. 2019;38(7):1147-1169. https://doi.org/10.1002/sim.8019.

24. Bennett M, White S, Best N, Mander A. A novel equivalence probability weighted power prior for using historical control data in an adaptive clinical trial design: A comparison to standard methods. *Pharm Stat*. 2021;20(3):462-484. https://doi.org/10.1002/pst.2088.

25. Normington J, Zhu J, Mattiello F, Sarkar S, Carlin B. An efficient Bayesian platform trial design for borrowing adaptively from historical control data in lymphoma. *Contemp Clin Trials*. 2020;89:105890. https://doi.org/10.1016/j.cct.2019.105890.

26. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*. 2014;70(4):1023-1032. https://doi.org/10.1111/biom.12242.

27. Hupf B, Bunn V, Lin J, Dong C. Bayesian semiparametric meta-analytic-predictive prior for historical control borrowing in clinical trials. *Stat Med*. 2021;40(14):3385-3399. https://doi.org/10.1002/sim.8970.

28. Wang C, Lin M, Rosner GL, Soon G. A Bayesian model with application for adaptive platform trials having temporal changes. *Biometrics*. 2023;79(2):1446-1458. https://doi.org/10.1111/biom.13680.

29. Jiang L, Nie L, Yuan Y. Elastic priors to dynamically borrow information from historical data in clinical trials. *Biometrics*. 2023;79(1):49-60. https://doi.org/10.1111/biom.13551.

30. Zhang W, Pan Z, Yuan Y. Elastic meta-analytic-predictive prior for dynamically borrowing information from historical data with application to biosimilar clinical trials. *Contemp Clin Trials*. 2021;110:106559. https://doi.org/10.1016/j.cct.2021.106559.

31. Duan Y, Ye K, Smith EP. Evaluating water quality using power priors to incorporate historical information. *Environmetrics*. 2006;17(1):95-106. https://doi.org/10.1002/env.752.

32. Treatment-control comparisons in platform trials including non-concurrent controls - X-MOL. Accessed September 23, 2026. https://www.x-mol.com/paper/1929832370830974976

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- table1.docx
- table2.docx
- tables1.docx
- tables2.docx
- Additionalfile3.xlsx
- FigureS1.tif
- FigureS2.tif
- Additionalfile12.docx
- Figurelegends.docx