Article

CLINICAL TRIALS

Clinical Trials
1–9

DOI: 10.1177/1740774515626362
ctj.sagepub.com
SAGE

Efficiencies of platform clinical trials: A vision of the future

Benjamin R Saville[1,2] and Scott M Berry[1,3]

Abstract

Background: A "platform trial" is a clinical trial with a single master protocol in which multiple treatments are evaluated simultaneously. Adaptive platform designs offer flexible features such as dropping treatments for futility, declaring one or more treatments superior, or adding new treatments to be tested during the course of a trial.

Methods: A simulation study explores the efficiencies of various platform trial designs relative to a traditional two-arm strategy.

Results: Platform trials can find beneficial treatments with fewer patients, fewer patient failures, less time, and with greater probability of success than a traditional two-arm strategy.

Conclusion: In an era of personalized medicine, platform trials provide the innovation needed to efficiently evaluate modern treatments.

Keywords

Platform trial, master protocol, multi-arm, adaptive, Bayesian, clinical trial design

Introduction

The introduction of the randomized clinical trial (RCT) in 1946 laid the foundation for modern pharmaceutical and medical device development.[1] Subsequent developments in ethics, statistical methods, and further scientific discovery led to a boon in clinical development, with hundreds of therapies being accepted for general use in the United States in the latter half of the 20th century. However, since the mid-1990s, the annual number of new drugs and biologics approved by the Food and Drug Administration (FDA) has stagnated despite a massive increase in spending on research and development and in the number of compounds under development.[2–4] Recent estimates of clinical trial failure rates are 36%, 68%, and 40% for phase I, phase II, and phase III trials, respectively,[3] and the estimated average cost of a successful New Molecular Entity (NME) in a big pharmaceutical company is a staggering US$5 billion.[5]

Traditionally, clinical trials have been designed by a single sponsor to evaluate a single treatment in a homogeneous group of patients. For example, a trial may be designed to determine whether "Drug A" is effective in a group of "Type 1" patients (see Figure (1a)). However, biomarker development and personalized medicine are leading to a future in which the vast majority of diseases are "rare" diseases. This will make slow, large-scale clinical trials with a single hypothesis within a single disease impractical to conduct, with the speed of medical discovery outpacing the planned completion of such trials. One potential solution is the use of "umbrella," "basket," or "indication finder" studies, which evaluate the effect of a single drug (e.g. "Drug A") in many different types of patients ("Type 1," "Type 2," etc.; see Figure 1(b)).[6]

Advances in personalized medicine are also leading to increasingly complex treatment regimens. This is forcing researchers to address a different question. That is, "which treatment or combination of treatments is best for each type of patient?" To efficiently address this question, we advocate the use of multi-arm platform trials.[7–9] Platform trials have master protocols that evaluate multiple treatments (e.g. "Drug A," "Drug B") across one or more types of patients ("Type 1," "Type 2," etc.; see Figures 1(c) and 1(d)). Platform trials are

[1]Berry Consultants, Austin, TX, USA
[2]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, USA
[3]Adjunct faculty, University of Kansas Medical Center, Department of Biostatistics, KS, USA

Corresponding author:
Benjamin R Saville, Berry Consultants, Austin, TX 78746, USA.
Email: ben@berryconsultants.com

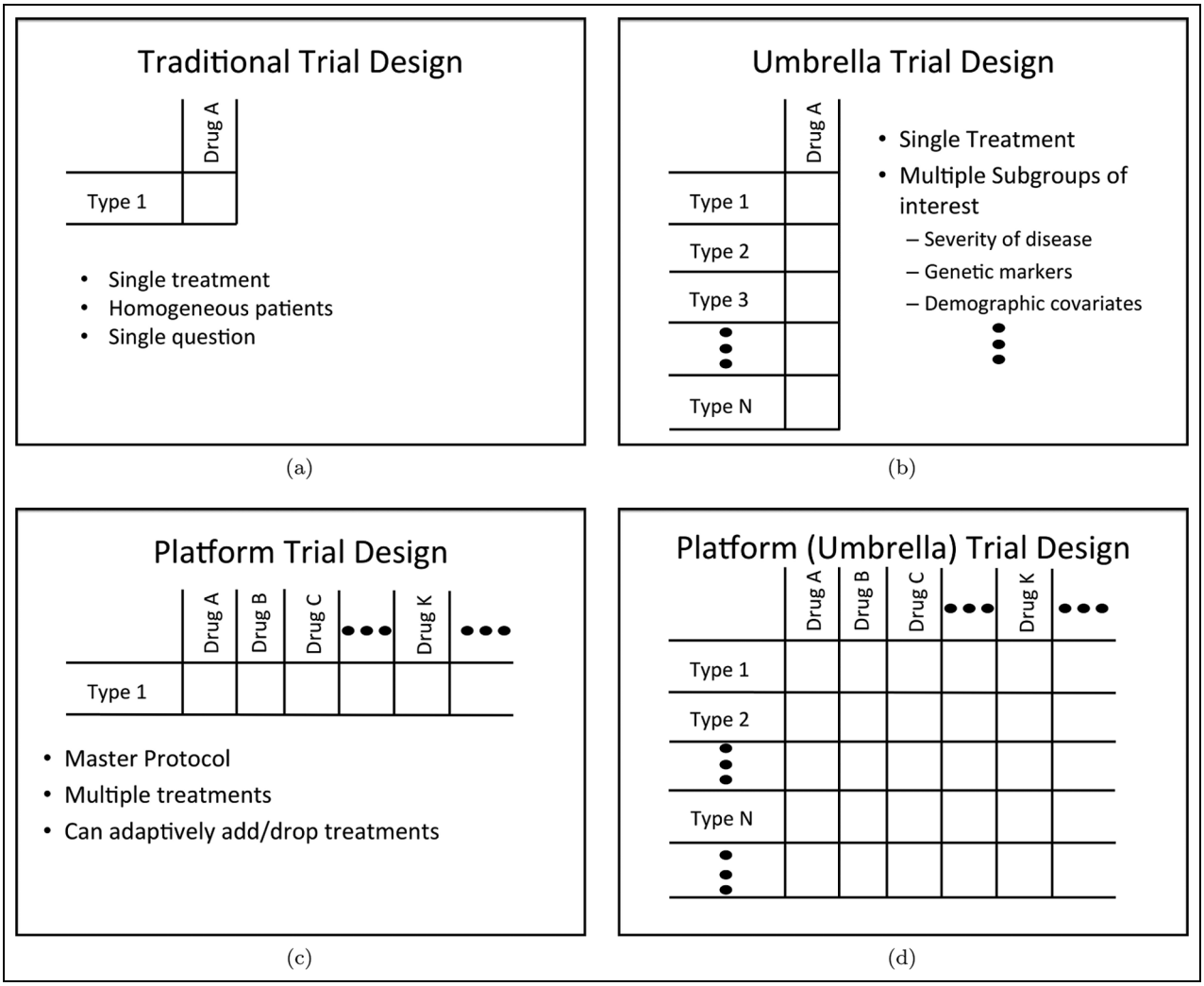

Figure 1. Illustrations of (a) traditional, (b) umbrella, and (c and d) platform designs.

especially useful for exploring combinations of treatments and for direct comparisons between competing treatments, both of which are often neglected in premarket settings. In addition, the sharing of resources in platform trials possibly between multiple sponsors can greatly reduce costs and increase statistical efficiency.

Platform trials can either consist of a fixed number of treatments (e.g. a network trial with five treatments from different sponsors) or an adaptive number of treatments in which treatments can be dropped and/or added during the course of the trial. The latter is what we refer to as an open or perpetual platform trial, in that the trial continues in a perpetual nature provided that new experimental treatments are available to enter the trial. Such platform trials can find effective treatments much more quickly and with fewer resources compared to traditional strategies investigating one treatment per trial. In addition, they do not require a new trial infrastructure for every treatment under investigation.

Platform trials have been successfully implemented or are currently being planned in a variety of diseases, including breast cancer,[10] lung cancer,[11,12] brain cancer, pandemic influenza and community-acquired pneumonia,[13] Alzheimer's,[14] Ebola, melanoma, glioblastoma, and sclererderma. These include both phase II and phase III settings. A recent report released by the President's Council of Advisors on Science and Technology (PCAST) included a call for platform trials for the evaluation of antibiotic drugs.[15] The most prominent platform trial, I-SPY 2, is a phase II platform trial used to screen drugs in neoadjuvant breast cancer.[10] Several drugs are compared to a control with adaptive randomization by biomarker subtypes; drugs that lack sufficient activity are replaced by new drugs, and drugs that show promise are able to move more quickly through the trial and "graduate" to smaller focused phase III studies. Despite the clear benefits of platform trials, they remain relatively rare in practice, and we believe many researchers do not fully comprehend the benefits

of platform trials. The purpose of this article is to quantitatively assess the efficiencies of platform trials in a simple setting in order to illustrate the potential benefits. We present three general strategies for comparison via simulation.

Methods

Competing strategies

Consider a disease with high morbidity (e.g. cancer) in which a very large number or population of treatments exist, of which only a small proportion are effective. We assume a binary outcome for patient response (yes/no), where "yes" indicates a successful treatment and that patients in the study population are homogeneous, meaning there is only one "type" of patient. In this setting, a reasonable goal of a phase 2 or phase 3 clinical trial is to find an effective treatment with as few patients as possible. A design that is able to find an effective therapy with a smaller number of patients will be less expensive, take less development time, and will result in fewer patient failures. In this context, we compare three general strategies in their ability to find an effective therapy:

1. *A sequence of traditional two-arm trials comparing a single treatment versus control.*
   (a) This strategy involves the comparison of a single treatment to a control with equal 1:1 randomization. The success criterion is a p-value from a one-sided chi-square test less than 0.025. If the treatment does not meet the success criterion, a new 1:1 trial will be conducted. The process is repeated indefinitely until a treatment is found that meets the success criterion, at which point the program ends; see Figure 2(a).
   (b) A 1:1 randomized design similar to (1a), but includes interim monitoring for success and futility at equally spaced intervals. For consistency with other designs, success and futility criteria are based on the Bayesian posterior probability that the treatment proportion is greater than the control proportion, as detailed in section "Adaptive simulation assumptions." If the treatment does not meet success criteria by the end of the trial, a new 1:1 trial will be conducted. The process is repeated indefinitely until a treatment is found that meets success criteria, at which point the program ends.
2. *A sequence of "closed platform trials" comparing several treatments versus a control.*
   (a) This strategy is based on a shared control design, which is essentially the most basic platform trial design. This strategy involves a single platform trial with a fixed sample of treatments (e.g. $N = 5$) and a control, with equal randomization and fixed sample sizes in each arm. The trial is a "closed" platform trial, meaning no additional treatments are added beyond those included at the start of the trial. Each treatment is compared to the

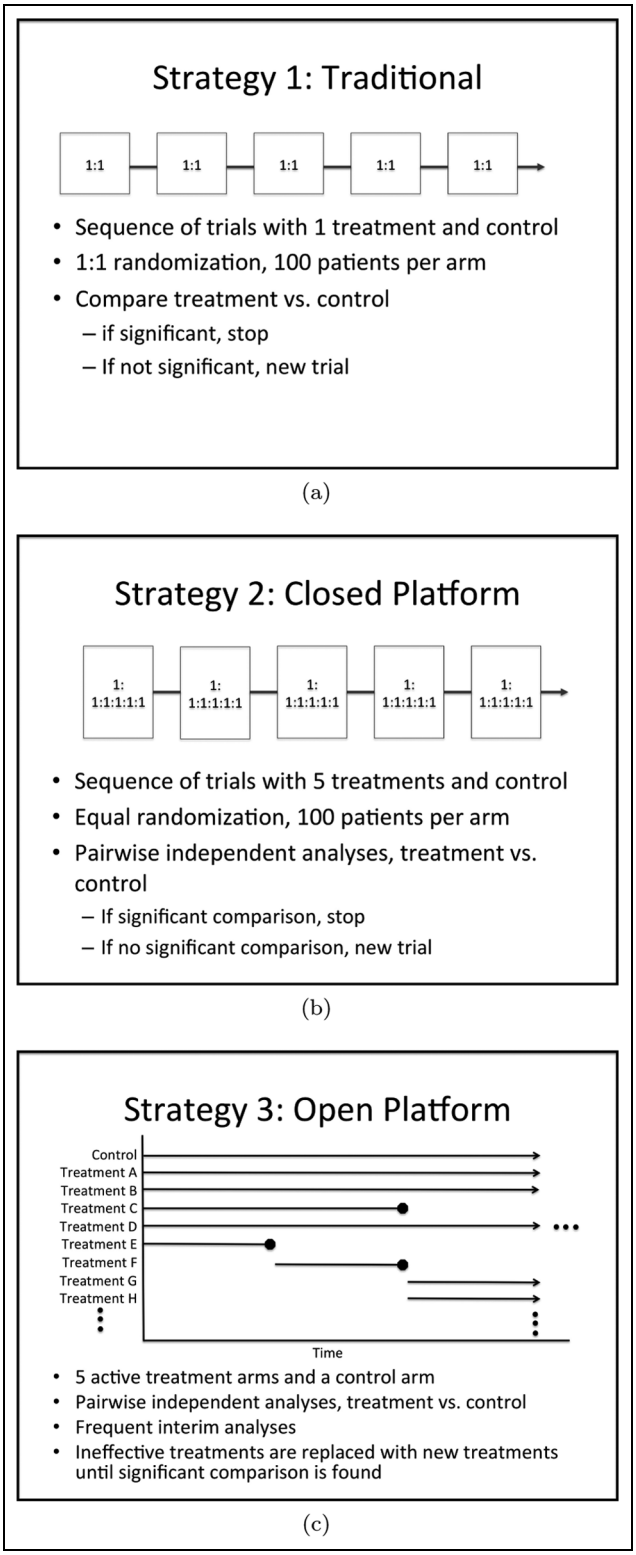

Figure 2. Illustrations of three competing strategies.

control using pairwise independent one-sided chi-square tests; success is claimed for a treatment if its p-value is less than 0.025. If none of the treatments meets the success criterion, a completely new platform trial is conducted with five new treatments and a new set of control patients. The process is repeated indefinitely until success is claimed for a treatment, at which point the program ends (see Figure 2(b)).

(b) A platform strategy similar to (2a) with five treatments versus a control, but includes interim monitoring for success and futility at equally spaced intervals. For consistency with other designs, success and futility criteria are based on the Bayesian posterior probability that the treatment proportion is greater than the control proportion, as detailed in section "Adaptive simulation assumptions." The overall accrual rate is constant; hence, if a treatment is dropped, it will lead to greater accrual on the remaining active arms. If none of the treatments meets the success criterion, a completely new platform trial is conducted with five new treatments and a new set of control patients. The process is repeated indefinitely until success is claimed for a treatment, at which point the program ends.

3. *Open (perpetual) adaptive platform trial.*
This strategy is an open adaptive platform clinical trial with a fixed number of active treatments (e.g. $N = 5$) and a control. The trial is "open" with respect to adding new treatments to replace ineffective treatments during the trial. Bayesian methods are used to adaptively drop (and add) treatment arms until the success criterion is met by one of the treatments, at which point the program ends (see Figure 2(c)). The success criterion is based on the Bayesian posterior probability that a treatment response rate is greater than the control response rate, as detailed in section "Adaptive simulation assumptions." Various adaptive features of the open platform trial are investigated, such as frequency of interim analyses, a maximum number of patients per treatment, response adaptive randomization (RAR), and comparison of treatment arms to either concurrent or all control patients.

We use the term "program" to denote the strategy of repeating a trial indefinitely until statistical significance is obtained. The term "trial" will refer to the individual clinical trials within a program. For strategy (3), the two terms are interchangeable because the program is a single perpetual trial.

For all strategies, we control the Type I error of labeling an ineffective treatment as "statistically significant" at 0.025, that is, each treatment has its own 0.025 alpha. There are multiplicity issues associated with this for all strategies (i.e. platform trials and sequences of traditional trials), which cause inflation in the Type I error of the "program," which we define as the proportion of programs that incorrectly identify an ineffective treatment as "statistically significant." Simulations are used to quantify the mean number of trials, number of treatments evaluated, number of patients enrolled, number of patient failures, and number of years of duration of the program. In addition, we calculate the proportion of programs that correctly identify an "effective" treatment (power), the proportion of effective treatments that are correctly identified as effective, and the proportion of ineffective treatments that are mistakenly identified as effective.

The three strategies chosen for evaluation are far from an exhaustive set of strategies available for the design of clinical trials. For example, trials may have different types of endpoints (continuous/survival), different effect sizes, different distributions of available effective therapies, different statistical methods of interim monitoring, or time lags in between patient enrollment and the observation of the primary outcome. The difficult task of exploring all possible strategies and combinations of simulation parameters is beyond the scope of this article. Rather, the purpose of this article is to demonstrate efficiencies that could be gained via platform trials in a basic setting. The principles underlying the benefits in this simple setting can be generalized to a variety of settings, strategies, or outcomes.

General simulation assumptions

To simulate trials for strategies (1)–(3), we make the following general assumptions:

- Patients who receive a control or ineffective treatment have a $p = 0.30$ response rate.
- Patients who receive an effective treatment have a $p = 0.50$ response rate.
- A total of 10% of available treatments are effective.
- Treatments are randomly selected from the pool of available treatments, and individual responses are simulated according to the response rate of that treatment.
- Enrollment rate is 200 patients/year, and there is a 3-month lag in between trials for strategies (1) and (2).
- Fixed sample sizes of $N = 100$ per arm for strategies (1a) and (2a).
- All chi-square tests are one-sided tests for two independent proportions, testing $H_0 : p_1 \leq p_0$ versus $H_a : p_1 > p_0$ via a Pearson chi-square.

Adaptive simulation assumptions

Programs which incorporate adaptive features (strategies (1b), (2b), and (3)) have the following additional assumptions:

- Patients are allocated to the control arm in a fixed 1:1 or 1:5 ratio relative to all active treatment arms. The randomization among the active treatment arms can vary depending on the strategy and scenario.
- All outcomes are observed on currently enrolled patients before conducting an interim analysis.
- Efficacy criteria are based on the Bayesian posterior probability that the treatment is superior to the control, that is, $\Pr(p_j > p_0|x)$, where p_j is the probability of success for active treatment j, p_0 is the probability of success for controls, and x represents the observed data. If $\Pr(p_j > p_0|x)$ is greater than a success boundary S, the treatment is considered effective.
- Futility criteria for dropping a treatment are based on the Bayesian posterior probability that the treatment success probability is greater than the control success probability by a clinically meaningful difference of 10 percentage points, that is, $\Pr(p_j > p_0 + 0.10|x)$. If this quantity is less than a futility boundary F, then the treatment is dropped from the program.
- The efficacy and futility criteria are varied by scenario so as to preserve the one-sided Type I error at 0.025 per treatment while maintaining reasonable power of the program.
- Priors for all probabilities are Beta (1, 1) which are uniform priors on the interval [0, 1].
- Each platform trial has exactly N active treatment arms and 1 control arm at all times.
- At each interim analysis, one of the following decisions will be made for a given active treatment:

1. If a treatment meets the success criterion, stop the program and claim efficacy;
2. Else if a treatment meets the futility criterion or the group maximum has been met, drop the treatment (and replace with new treatment if implementing open platform strategy);
3. Else continue enrolling patients on that treatment.

Open adaptive platform strategies

To give further understanding about the impact of adaptive features of the open platform strategy with respect to efficiency metrics, we provide six different strategies with various combinations of features, which are summarized in Table 1. We vary the following parameters for these six strategies:

- RAR (yes or no): specifies whether trial uses response adaptive randomization (RAR); if no, then the trial uses fixed equal randomization;
- Active Tmts (5 or 10): number of active treatments;
- Cohort size (300 or 150): number of patients enrolled between each interim analysis;
- Group max (100 or none): specifies whether there is a cap on group maximum sample size. If cap equals 100, and a success criterion has not been met by 100 patients within an arm, the treatment is removed from the trial;
- Controls (all or concurrent): specifies whether each treatment is compared to all control patients or concurrent controls.

For strategies (3d)–(3f), RAR is used, in which the randomization probabilities are based on θ_j or the probability that each active treatment (j) is the optimal treatment

$$\theta_j = \Pr(p_j > p_k \text{ for all } k \neq j), \quad j = 1, 2, \ldots, 5 \qquad (1)$$

Table 1. Scenario features.

Strategy	Success criterion	RAR	Active Tmts	Cohort size	Group max	Success boundary	Futility boundary	Controls
1a. Traditional two-arm	p-value	–	1	200	100	0.025	–	Concurrent
1b. Traditional adaptive two-arm	Bayes	–	1	50	100	0.991	0.10	Concurrent
2. Closed platform	p-value	–	5	600	100	0.025	–	Concurrent
2b. Closed adaptive platform	Bayes	–	5	150	100	0.99	0.10	Concurrent
3a. Open adaptive platform	Bayes	–	5	300	–	0.985	0.10	All
3b. Open adaptive platform	Bayes	–	5	150	–	0.99	0.10	All
3c. Open adaptive platform	Bayes	–	5	150	100	0.99	0.10	Concurrent
3d. Open adaptive RAR platform	Bayes	Yes	5	150	100	0.99	0.10	Concurrent
3e. Open adaptive RAR platform	Bayes	Yes	5	150	100	0.985	0.10	All
3f. Open adaptive RAR platform	Bayes	Yes	10	150	–	0.991	0.10	All

RAR: response adaptive randomization.

Table 2. Simulation summaries and population of treatments (5000 "Programs").

Strategy	Number trials	Number treatments	Number patients	Number failures	Years	Prob program success	Prob effective treatment success	Prob ineffective treatment success
1a. Traditional two-arm	9.8	9.8	1966	1357	12.0	0.78	0.82	0.025
1b. Traditional adaptive two-arm	10.7	10.7	1338	922	9.1	0.78	0.73	0.023
2. Closed platform	2.5	12.7	1528	1045	8.0	0.86	0.83	0.027
2b. Closed adaptive platform	3.4	13.7	971	663	5.5	0.82	0.76	0.025
3a. Open adaptive platform	1.0	11.7	1085	742	5.4	0.85	0.95	0.025
3b. Open adaptive platform	1.0	13.1	849	579	4.2	0.85	0.91	0.022
3c. Open adaptive platform	1.0	14.7	935	638	4.7	0.83	0.77	0.025
3d. Open adaptive RAR platform	1.0	14.4	912	616	4.6	0.83	0.75	0.025
3e. Open adaptive RAR platform	1.0	13.4	769	521	3.8	0.85	0.85	0.022
3f. Open adaptive RAR platform	1.0	18.2	640	431	3.2	0.87	0.85	0.023

RAR: response adaptive randomization; Number trials: mean number of trials; Number treatments: mean number of experimental treatments evaluated; Number patients: mean total number of patients enrolled; Number failures: mean total number of failures across all treatments; Years: number of years to claim efficacy with 200 patients/year and 3 months between trials; Prob program success: proportion of programs that correctly identify an "effective" treatment; Prob effective treatment success: proportion of effective treatments that are correctly identified as effective; Prob ineffective treatment success: proportion of ineffective treatments that are mistakenly identified as effective.

The randomization probability for each active arm is proportional to

$$q_j \propto \sqrt{\frac{\theta_j V(p_j)}{n_j + 1}} \tag{2}$$

where $V(p_j)$ is the estimated variance of the response rate and n_j is the sample size for treatment j. The RAR probability typically randomizes more patients to the treatment that is most likely to be the optimal treatment and does this by minimizing the change in the estimated variance of the optimal treatment through randomization. This accounts for both the probability that each treatment is the best and the precision of the estimated response rate. The randomization probabilities at a given interim are normalized so that they sum to 1. If any treatment is dropped from the program based on the futility criterion, then an initial fixed number of patients ($n_j = 12, 25,$ or 50, depending on scenario) are assigned to a replacement treatment, and another renormalization is done to allocate the remaining patients to the ongoing treatments.

Simulation results

Table 2 shows the simulation results for the various strategies. For comparison purposes, a traditional design under the above assumptions with a single (non-repeating) stand-alone trial has 0.83 power for a two-group comparison. Our proposed traditional strategy (1a) involves the repetition of the stand-alone trials until statistical significance is obtained, at which point the program ends. On average, the traditional strategy (1a) requires 9.8 trials (and 9.8 treatments) to claim statistical significance. The mean number of patients is 1966 with 1357 mean failures. The proportion of programs that correctly identify an effective treatment is 0.78 (i.e. power of the program), meaning 22% of the programs ultimately label an ineffective treatment as effective. The proportion of effective treatments that are identified as effective is 0.82 (0.83 is the analytical solution but only differs here due to simulation variability), and the proportion of ineffective treatments mistakenly identified as effective is 0.025. The mean duration to finding a significant treatment is 12 years. The traditional adaptive strategy (1b) reduces the mean number of patients and failures to 1338 and 922, respectively. The proportion of programs that correctly identify an effective treatment is 0.78, and the proportion of effective treatments that are correctly identified as effective is 0.73.

On average, the closed platform strategy without interims (2a) requires 2.5 trials and 12.7 treatments to claim statistical significance. The mean number of patients is 1528 with 1045 mean failures. The proportion of programs that correctly identify an effective treatment is 0.86, the proportion of effective treatments correctly identified as effective is 0.83, and the proportion of ineffective treatments mistakenly identified as effective is 0.027 (0.025 is the analytical solution). The mean duration to finding a significant treatment is 8 years. The closed platform strategy (2a) is more efficient at finding effective treatments than the traditional strategy (1a) primarily because of the sharing of controls within a trial. Interim monitoring for success and futility leads to improved performance for both the traditional (1b) and closed platform (2b) strategies. The

closed adaptive trial with interim monitoring (2b) outperforms strategies (1a), (1b), and (2a) by a substantial margin, requiring on average 971 patients, 663 failures, and 5.5 years to claim statistical significance, with 0.82 probability correctly identifying an effective treatment.

In general, the open adaptive platform programs demonstrate better performance than either the traditional or closed platform strategies. The magnitude of performance benefits vary by the features included in the open platform trial. Factors associated with greater efficiencies included more frequent interim looks, comparison of treatments to all controls, and including RAR.

The most basic open platform (3a) with cohorts of 300 patients (50 per arm) and equal randomization gives a mean of 1085 patients, 742 failures, 11.7 treatments, 5.4 years of development, and a 0.85 probability of successfully finding an effective treatment. This is a substantial improvement over the closed platform strategy (2a), primarily a result of interim futility analyses that allowed dropping of ineffective treatments, but does not improve over the closed adaptive platform strategy (2b) due to a larger cohort size between interims. With a comparable cohort size of 150 patients between interims (strategy (3b)), the open platform strategy has a smaller number of patients (849), fewer failures (579), and shorter program duration (4.2 years), with faster cycling through the large number of ineffective treatments (13.1 active treatments evaluated).

When treatment comparisons are limited to only concurrent controls with a group maximum sample size (3c), the mean number of patients increases to 935 with 638 failures, 4.7 years of development, and 14.7 treatments. RAR produces minimal benefit for the open platform strategy in the context of concurrent controls (3d vs 3c), but RAR provides more substantial benefit for the open platform strategy using all controls (3e vs 3b). Adaptive randomization using all controls (3e) is the most efficient strategy among open platform strategies with 5 active treatments, requiring on average 769 patients, 521 failures, 13.4 treatments, 3.8 years of duration, and 0.85 probability of obtaining an effective treatment. Compared to the traditional strategy, this results in average savings of 1197 patients, 836 patient failures, and 8.2 years of drug development compared to the traditional strategy, while having a greater probability of correctly identifying an effective treatment (0.85 vs 0.78, respectively). This is greater than a 60% reduction in the number of patients and failures!

Finally, the open adaptive platform strategy with 10 active treatments outperformed all of the above strategies, requiring on average 640 patients, 431 failures, 18.2 treatments, 3.2 years of duration, and 0.87 probability of obtaining an effective treatment or nearly a 70% reduction in the number of patients and failures compared to the traditional strategy.

In addition to the paradigm of a very large number (i.e. a population) of treatments available, we also considered the paradigm of a smaller fixed number of treatments available ($N = 10$), of which only one is effective. This would be consistent with a disease such as Ebola, in which there are limited treatments available. In this paradigm, simulations for a single "program" continue until either (1) a treatment meets the success criterion or (2) all available treatments meet the futility criteria. Simulations showed similar differences in performance between the competing strategies, with open adaptive platform strategies achieving similar power with greater than a 50% reduction in number of patients and failures compared to the traditional strategy (results not shown).

Discussion

A key concept of the platform trial is that it has a single master protocol and shared infrastructure across experimental treatments. Hence, all treatment arms share a common statistical analysis plan with predefined futility and success criteria. In practice, not all treatments need to share the same inclusion criteria, meaning different treatments may target different subtypes of patients for testing.

In a simple setting, we have demonstrated improved statistical efficiency of open adaptive platform trials over traditional two-arm trials. In these simulations, much of the benefit gained from the platform strategies is due to the fast cycling through ineffective treatments, which is useful in the setting in which only 10% of treatments are effective. If the proportion of effective treatments were much larger, for example, 50% or 60%, these benefits would likely be reduced because any of the strategies could quickly identify effective treatments. However, given that the vast majority of treatments that enter development fail to obtain regulatory approval, one could make a strong argument that effective treatments are elusive in most clinical settings.

Although platform trials need not implement RAR, our simulations show benefit for RAR versus fixed randomization in a platform strategy with interim monitoring when active treatments are compared to all controls. This contrasts with the views of some authors who claim that RAR provides no benefit relative to fixed randomization with interim monitoring.[16] Relatedly, debates on the ethics of RAR are prevalent in the literature.[17–19] Regardless of one's view on RAR, it is clear from our simulation study that either type of open platform strategy (RAR or fixed randomization) offers substantial benefits relative to traditional two-arm studies or closed platform studies in the context presented.

We note our simulations did not explore the full potential of platform trials. For example, we did not

assess treatment benefit in patient subgroups, such as implemented by I-SPY 2 and BATTLE. Additionally, the illustrated platform strategies do not take advantage of the capability to directly compare experimental treatments (each treatment is regarded as independent of the other treatments in the simulations), nor do the illustrated strategies allow the control arm to be replaced with a superior treatment to continue evaluation of future treatments. To increase efficiency and reliability of results, shrinkage methods could be used to borrow information across treatment arms and account for regression to the mean when appropriate.[6]

There are many challenging issues that accompany the use of adaptive platform trials. For example, delays between enrollment and outcome observation can have an impact and potentially negate the benefit of adaptations. In many delayed outcome settings, Bayesian predictive probabilities can be used to stop accrual for expected success,[20] with a decisive analysis conducted after full follow-up on all patients enrolled at the interim analysis. In such settings, longitudinal modeling of primary outcomes using auxiliary variables and predictive probabilities can be beneficial.[21] For example, I-SPY 2 uses magnetic resonance imaging (MRI) sequentially to predict 6-month response rates.

Another potential issue is population drift, which can occur through a variety of mechanisms, including improved general care and enrollment of either more or less healthy patients. There are two general strategies for dealing with population drift: (1) modeling drift (which has been implemented in trials such as I-SPY 2) and (2) comparing treatments to concurrent controls. These are complicated issues that need discussion beyond this article.

In a platform trial with many experimental agents it may be difficult or impossible to blind patients to every possible arm. They may have different modes of administration or dosing in such ways that blinding to all arms becomes incredibly difficult and burdensome. It is not uncommon that patients in platform trials are unblinded to which possible treatment arm they receive, but remain blinded to whether they receive active or placebo of that treatment.

Additionally, control of Type I error can be a challenge for adaptive platform trials because non-analytic methods (i.e. simulations) may be required. The degree of control is typically determined on a case-by-case basis per regulatory and/or scientific review. General strategies include exploring a wide range of scenarios, with particular emphasis on extremes of the assumptions involved (per context) to determine the range of plausible Type I error. With complexities of delayed outcomes, population drift, and unknown pattern of available treatments, control of Type I error is even more difficult to demonstrate. Furthermore, knowing exactly how to define Type I error for a single arm in a trial with multiple arms is difficult. If two drugs were explored in separate trials, they may each have individual 2.5% Type I error; if the two drugs are in a common platform trial, should the experiment have 2.5% Type I error, or should each arm have 2.5% Type I error? All these issues should be discussed thoroughly (with regulatory agencies where applicable) before the running of a platform trial.

Platform trials also bring complexity in trial implementation and planning. For example, platform trials may require complex collaborations across sponsors and timely communication between participating sites and data coordinating units. They typically require sponsors to sacrifice some autonomy in running the trial, oftentimes to a third party that designs and executes the master protocol. Related issues include determining the shared costs of running the trial and whether there are sufficient incentives for sponsors to initiate the designing of a platform trial. Adding new treatments may require modified eligibility criteria or additional diagnostic/biomarker data or may change the appeal of the study potentially affecting accrual rates and the study population. Avoiding operational bias can also be a challenge. We recommend further discussion in the literature on these issues.

As demonstrated, platform trials can be much more efficient strategies for finding effective therapies than traditional stand-alone trials. In practice, platform trials have typically used Bayesian methods to implement innovative features such as RAR, hierarchical and longitudinal modeling, predictive probabilities, and the assessment of various treatment combinations across multiple subgroups. Such designs require extensive simulations in the design phase to explore performance. These prospective simulations allow investigators to see how such a trial would progress under a variety of different scenarios and thus provide great flexibility in making decisions that closely align with the scientific goals of the study. This flexibility is essential to the innovation needed to keep pace with the rapid developments in medical research and personalized patient care. To meet this challenge, we advocate increase usage and consideration of multi-arm adaptive platform trials.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. MRC Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis. *Br Med J* 1948; 2: 769–783.

2. Scannell JW, Blanckley A, Boldon H, et al. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov* 2012; 11: 191–200.
3. Hay M, Thomas DW, Craighead JL, et al. Clinical development success rates for investigational drugs. *Nat Biotechnol* 2015; 32(1): 40–51.
4. Woodcock J and Woosley R. The FDA critical path initiative and its influence on new drug development. *Annu Rev Med* 2008; 59: 1–12.
5. Herper M. The cost of creating a new drug now $5 billion, pushing big pharma to change, http://www.forbes.com/sites/matthewherper/2013/08/11/how-the-staggering-cost-of-inventing-new-drugs-is-shaping-the-future-of-medicine (2015, accessed 16 April 2015).
6. Berry SM, Broglio KR, Groshen S, et al. Bayesian hierarchical modeling of patient subpopulations: efficient designs of Phase II oncology clinical trials. *Clin Trials* 2013; 10: 720–734.
7. Berry SM, Connor JT and Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. *JAMA* 2015; 313: 1619–1620.
8. Berry DA. The Brave New World of clinical cancer research: adaptive biomarker-driven trials integrating clinical practice with clinical research. *Mol Oncol* 2015; 9(5): 951–959.
9. Angus DC. Fusing randomized trials with big data: the key to self-learning health care systems? *JAMA* 2015; 314(8): 767–768.
10. Esserman LJ and Woodcock J. Accelerating identification and regulatory approval of investigational cancer drugs. *JAMA* 2011; 306(23): 2608–2609.
11. Zhou X, Liu S, Kim ES, et al. Bayesian adaptive design for targeted therapy development in lung cancer—a step toward personalized medicine. *Clin Trials* 2008; 5(3): 181–193.
12. Steuer CE, Papadimitrakopoulou V, Herbst RS, et al. Innovative clinical trials: the LUNG-MAP study. *Clin Pharmacol Ther* 2015; 97: 488–491.
13. European Society of Intensive Care Medicine. PREPARE—Platform foR European Preparedness Against (Re) emerging Epidemics, http://www.esicm.org/research/prepare (2015, accessed 16 April 2015).
14. European Prevention of Alzheimer's Dementia Consortium. 35 partners from industry and academia to join European research initiative for the prevention of Alzheimer's dementia, http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/EPAD-master-press-release-FINAL-1.pdf (2015, accessed 16 April 2015).
15. President President's Council of Advisors on Science and Technology. Report to the President on Combating Antibiotic Resistance, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_carb_report_sept2014.pdf (2014, accessed 29 April 2015).
16. Korn EL and Freidlin B. Commentary on Hey and Kimmelman. *Clin Trials* 2015; 12(2): 122–124.
17. Hey SP and Kimmelman J. Are outcome-adaptive allocation trials ethical? *Clin Trials* 2015; 12(2): 102–106.
18. Berry DA. Commentary on Hey and Kimmelman. *Clin Trials* 2015; 12(2): 107–109.
19. Lee JJ. Commentary on Hey and Kimmelman. *Clin Trials* 2015; 12(2): 110–112.
20. Saville BR, Connor JT, Ayers GD, et al. The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. *Clin Trials* 2014; 11(4): 485–493.
21. Berry SM, Carlin BP, Lee JJ, et al. *Bayesian adaptive methods for clinical trials*. Boca Raton, FL: Taylor & Francis Group, 2011.