Published in final edited form as:

Stat Methods Med Res. 2018 April ; 27(4): 1219–1229. doi:10.1177/0962280216657853.

A Bayesian multi-stage cost-effectiveness design for animal studies in stroke research

Chunyan Cai^{1,2}, Jing Ning³, and Xuelin Huang³

¹Division of Clinical and Translational Sciences, Department of Internal Medicine, McGovern Medical School; The University of Texas Health Science Center at Houston, Houston, TX, USA

²Biostatistics/Epidemiology/Research Design Core, Center for Clinical and Translational Sciences, The University of Texas Health Science Center at Houston, Houston, TX, USA

³Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

Much progress has been made in the area of adaptive designs for clinical trials. However, little has been done regarding adaptive designs to identify optimal treatment strategies in animal studies. Motivated by an animal study of a novel strategy for treating strokes, we propose a Bayesian multi-stage cost-effectiveness design to simultaneously identify the optimal dose and determine the therapeutic treatment window for administering the experimental agent. We consider a non-monotonic pattern for the dose-schedule-efficacy relationship and develop an adaptive shrinkage algorithm to assign more cohorts to admissible strategies. We conduct simulation studies to evaluate the performance of the proposed design by comparing it with two standard designs. These simulation studies show that the proposed design yields a significantly higher probability of selecting the optimal strategy, while it is generally more efficient and practical in terms of resource usage.

Keywords

Admissible set; Animal study; Bayesian approach; Cost-effectiveness design; Multistage design

1. Introduction

Animal studies in preclinical research are important sources of the initial data on the safety and effectiveness of experimental treatments before they advance to clinical trials.¹

Unfortunately, the therapeutic effects observed in a large proportion of animal studies may not successfully translate into similar results in clinical studies.² One possible reason is that the animal studies were not rigorously designed and reported.^{3,4,5,6} To ensure the translation

Corresponding author: Division of Clinical and Translational Sciences, Department of Internal Medicine, McGovern Medical School; Biostatistics/Epidemiology/Research Design Core, Center for Clinical and Translational Sciences; The University of Texas Health Science Center at Houston, Houston, TX, USA; chunyan.cai@uth.tmc.edu.

of animal research results into greater clinical success and impact, it is imperative that researchers appropriately apply advanced statistical methodology for animal studies.

Adaptive designs, which have been developed and implemented in recent clinical trials,^{7,8} are different from traditional trial designs in that they allow the accumulating data to modify aspects of the trial in a pre-planned manner. Popular adaptations of adaptive designs include the adaptive dose-finding algorithm, modification of a randomization scheme, sample size re-estimation, stopping the trial early for futility or superiority, dropping or adding new treatment arms, biomarker-guided treatment allocation, and seamless phase I/II or II/III designs. A comprehensive review of adaptive strategies used in clinical trials can be found in Chow et al.,⁹ Berry et al.,¹⁰ and Chow and Chang.⁷ The benefits of adaptive trials include reducing sample size and total cost, assigning more patients to more efficacious treatments, and improving the efficiency without compromising the integrity of the study. In an evaluation of Bayesian decision-theoretic designs, Lewis and Berry¹¹ discussed their potential applicability and advantages when applied to animal studies. However, the literature offers few strategies for accommodating the adaptive strategies widely used in clinical trials to animal studies for the purpose of efficiently utilizing available resources, saving costs and yielding better translation validity.

The motivation for this paper is an animal study of a novel agent for treating stroke. Stroke is the leading cause of serious long-term disability and death in the USA.¹² The current approved treatment for acute ischemic stroke is thrombolysis with intravenous recombinant tissue plasminogen activator (rtPA).¹³ Despite its effectiveness in improving neurological outcomes, the therapeutic benefit of rtPA is time-dependent and limited to a short therapeutic window. Most of the guidelines on the administration of rtPA recommend its use within a 3-to-4.5-hour window after the onset of symptoms.¹⁴ In fact, the median time from stroke onset to hospital arrival is between 3 and 6 hours, and therefore treating patients within 3 to 4.5 hours since the onset of a stroke is a difficult target to meet. As a result, a substantial number of patients who suffer an ischemic are not eligible for rtPA treatment.¹⁵ Recently, researchers have investigated the use of advanced technologies to apply rtPA as quickly as possible to optimize its benefit, such as putting the emergency room in the ambulance¹⁶ and using telemedicine.¹⁷ However, speeding up the treatment is not an easy task in real practice, and requires substantial cost and personnel with specialized training. In our motivating study, the physicians aim to assess a novel agent that is expected to have an effective benefit within an extended therapeutic window. An animal study is planned to determine the optimal dose that gives maximal safety protection and the therapeutic time window based on efficacy. Here, the therapeutic window indicates the maximum treatment time since the onset of symptoms at which the agent is still effective. With this potentially extended window, the experimental agent would allow for the possibility of treating more patients who do not arrive at the hospital within a short time frame.

Toward this goal, we propose a Bayesian multi-stage design to simultaneously identify the optimal dose and determine the therapeutic window for the experimental agent. We define a utility score for each investigated strategy (i.e., a combination of dose level and treatment window) based on the efficacy rates and administration costs associated with the treatment window. Our proposed design consists of a run-in stage and an iterative selection stage. The

run-in stage starts with cohort assignments to the strategies with different dose levels given within the shortest treatment window. Then, using the selected dose level, it switches to assigning cohorts to the strategies given within different treatment windows. The run-in stage provides an initial estimation of the efficacy and toxicity rates for the systematic finding in the next stage. In the iterative selection stage, we develop an adaptive shrinkage algorithm to encourage the cohort assignments within admissible strategies. Based on the data collected in both the run-in and iterative selection stages, we calculate the utility scores and select the optimal strategy as the one with the highest utility score in the admissible set.

The remainder of this paper is organized as follows. In Section 2, we introduce our proposed multi-stage Bayesian design to simultaneously identify the optimal dose and determine the therapeutic window for the experimental agent. We apply logistic regression models for both toxicity and efficacy endpoints for statistical inference. In Section 3, we examine the operating characteristics of the proposed design through extensive simulation studies and compare it to two standard approaches. We conclude with a brief discussion in Section 4.

2. Methods

2.1 Notations

Assume there are K dose levels denoted as $d_1 < d_2 < \dots < d_K$, and L pre-specified treatment windows denoted as $S_1 < S_2 < \dots < S_L$, resulting in a total of KL strategy combinations to be investigated. Here, S_l denotes the time from symptom onset to the administration of medication. Let the treatment strategy using dose level d_k within time window S_l as T_{kl} . Suppose both toxicity and efficacy endpoints are binary and define the toxicity and efficacy rates of receiving T_{kl} by p_{kl} and q_{kl} respectively. In our motivating study, the toxicity endpoint can be whether there are any adverse effects on physiology or haematology, and the efficacy endpoint can be whether there is any improvement in the functional outcome. We define a strategy as admissible if its toxicity and efficacy rates satisfy the following criteria:

18

$$Pr(p_{kl} < \phi) > c_1 \ \& \ Pr(q_{kl} > \psi) > c_2, \quad (1)$$

where ϕ is the pre-specified highest tolerable toxicity level, ψ is the pre-specified lowest acceptable efficacy level, and c_1 and c_2 are the calibrated cut-offs. The set containing all admissible strategies is called an admissible set. Then the optimal treatment strategy is the one having the highest efficacy rate among the strategies given within the longest treatment window within the admissible set, and this longest treatment window is called the therapeutic window.

To incorporate the cost of administering each strategy into our algorithm, we assign a weight $0 < w_l \leq 1$ to a treatment window S_l for $l = 1, \dots, L$, which is inversely related to the cost required to meet the window S_l . The shorter the treatment window, the higher the administration cost. Therefore, we assume $0 < w_1 < w_2 < \dots < w_L = 1$ for L different treatment windows. On that basis, we define the utility score U_{kl} , $k = 1, \dots, K$, $l = 1, \dots, L$, for strategy T_{kl} as follows,

$$U_{kl} = w_l \times Pr(q_{kl} > \psi). \quad (2)$$

Using the defined utility score as the basis for the comparison of all admissible strategies, we propose a multi-stage cost-effectiveness design to simultaneously identify the optimal dose and determine the therapeutic window for the experimental agent.

2.2 Trial Design

Our design consists of a run-in stage and an iterative selection stage. The goal of the run-in stage is to explore the two-dimensional space quickly and collect preliminary data so that the initial admissible set can be identified for the next stage. In the iterative selection stage, we propose an adaptive shrinkage algorithm with iterative steps to determine the allocation of cohort assignments and identify the final admissible set. In each iterative step g , we aim to determine the number of cohorts to be used in the current step, denoted as M_g , by comparing the number of strategies in the admissible set, denoted as Q_g , to the number of cohorts we have treated in the previous step, denoted as M_{g-1} .

2.2.1 Run-in stage

- (a) Assign one cohort to each strategy T_{k1} , $k = 1, \dots, K$, i.e., K strategies administered with K dose levels within the shortest treatment window S_1 . From these K strategies, identify the dose level d_{k*} that has the highest value of $Pr(q_{k1} > \psi)$ with tolerable toxicity, i.e., $Pr(p_{k1} < \phi) \leq c_1$. If none of these strategies is tolerable, select the lowest dose level d_1 as d_{k*} .
- (b) Assign one cohort to each strategy with the selected dose level d_{k*} given within S different treatment windows, i.e., T_{k*s} , $s = 1, \dots, S$.

The run-in stage is complete after the assignment of $M_0 = K + S$ cohorts, and the iterative selection stage then starts.

2.2.2 Iterative selection stage—Let $0 < f_l < f_u < 1$ denote the lower and upper shrinkage parameters, respectively. The adaptive shrinkage algorithm is described as follows:

- (c) Start with $g = 1$.
- (d) Identify the admissible set. Suppose there are Q_g strategies that satisfy the criteria (1). If $Q_g = 0$, terminate the study and deem the trial inconclusive. Otherwise,
 - i. if $Q_g > \lceil f_u M_{g-1} \rceil$, set $M_g = \lceil f_u M_{g-1} \rceil$ and select the top M_g strategies by ranking U_{kl} within the admissible set. Here, $\lceil a \rceil$ denotes the smallest integer not less than a real value of a .
 - ii. if $Q_g < \lceil f_l M_{g-1} \rceil$, set $M_g = \lceil f_l M_{g-1} \rceil$ and select all Q_g admissible strategies as well as the top $M_g - Q_g$ strategies by ranking U_{kl} among the non-admissible strategies.

- iii. if $\lceil f_l M_{g-1} \rceil \leq Q_g \leq \lceil f_u M_{g-1} \rceil$, set $M_g = Q_g$ and select all admissible strategies.
- (e) Assign one cohort to each of the M_g selected strategies.
- (f) Stop the iteration if the pre-specified criterion is met. Otherwise, repeat steps (d) and (e) with $g = g + 1$.

Based on the data collected in both the run-in and iterative selection stages, we update the admissible set. When none of the strategies is admissible, we terminate the study and deem the trial inconclusive. Otherwise, we update the utility score and select the optimal strategy as the one with the highest utility score in the admissible set.

Our proposed shrinkage algorithm adaptively calculates the number of cohorts to be used in the current step based on the number of admissible strategies and the number of cohorts used in the previous step. If the admissible set is small, the algorithm encourages the exploration of non-admissible doses; if the admissible set is large, the algorithm focuses on assigning cohorts to potentially “optimal” strategies while saving resources. The shrinkage speed of

our algorithm can be controlled by the values of f_l and f_u . For example, we can set $f_l = \frac{1}{3}$ and $f_u = \frac{2}{3}$. The choice of the iteration stopping criteria can be calibrated through simulation studies or determined on the basis of the available resources. For example, we can stop the iteration if $M_g = 1$ or when a pre-specified number of iterations is reached or the resources are exhausted. There are several unique features of our proposed design based on the defined utility score. First, the selected optimal strategy is always admissible. Second, if there is more than one strategy with the same treatment window that is the longest window in the admissible set, the one with a larger value of $\Pr(q_{kl} > \psi)$ is selected. Third, within the admissible set, the strategies with longer treatment windows have a higher chance of being selected than the strategies with shorter treatment windows.

2.3 Modeling Toxicity and Efficacy

We consider the following dose-schedule-toxicity model to capture the relationship between the toxicity rate, dose level and treatment schedule,

$$\text{logit}(p_{kl}) = \beta_0 + \beta_1 d_k + \beta_2 S_l. \quad (3)$$

We assume $\beta_1 > 0$ and $\beta_2 > 0$ such that the toxicity rate monotonically increases with both the dose level and treatment window. If the toxicity rate does not change much as the treatment window increases, our model can still capture it if β_2 is close to zero.

In our motivating study, clinicians expect that the efficacy may not change much when the strategy is given within a potential time period, referred to as a “non-change” window. If the strategy is given outside of this “non-change” window, the treatment efficacy may decrease when the time to treatment increases. For the relationship between the dose level and the efficacy rate, we consider a model that includes the quadratic term of the dose level to capture the potential non-monotonic dose-efficacy relationship. That is, the efficacy may

monotonically change or may first increase and then decrease when the dose level changes. To incorporate such a complex pattern of the dose-schedule-efficacy relationship, we assume that the efficacy rate of treatment T_{kl} follows a logistic model

$$\text{logit}(q_{kl}) = \begin{cases} \gamma_0 + \gamma_1 d_k + \gamma_2 d_k^2 & s_l < \omega \\ \gamma_0 + \gamma_1 d_k + \gamma_2 d_k^2 - \gamma_3(s_l - \omega) & s_l \geq \omega, \end{cases} \quad (4)$$

where the parameter ω captures the boundary of the potential “non-change” window during which the efficacy may not change much. Under this model, the efficacy rate depends only on the dose level but not the treatment window if the experimental agent is given within the window ω . If given outside of the window ω , the efficacy rate decreases when the treatment window increases by assuming $\gamma_3 > 0$.

2.4 Prior Specification and Likelihood

We standardize the dose d_k and treatment time S_l with mean 0 and standard deviation 0.5 and use a weakly informative prior recommended by Gelman et al.¹⁹ Specifically, we assume that intercepts in models (3) and (4), i.e., γ_0 and β_0 follow $\text{Cauchy}(0, 10)$, and that the regression coefficients γ_1 and γ_2 in model (4) follow $\text{Cauchy}(0, 2.5)$. Here, $\text{Cauchy}(c, d)$ denotes a Cauchy distribution with the center parameter c and the scale parameter d . We assign γ_3 in the efficacy model (4) an independent gamma prior distribution with the shape parameter of 0.5 and the rate parameter of 0.5 to ensure the monotonic relationship between the efficacy and treatment window when the treatment is given outside of the “non-change” window. Similarly, we assign β_1 and β_2 in the toxicity model (3) the same prior as γ_3 to ensure the monotonicity of the dose-schedule-toxicity relationship. For the boundary of the “non-change” window, ω , we choose a uniform prior with a sufficient coverage of all possible values of the standardized S_l .

Suppose that at a certain stage of the trial, among n_{kl} patients who are given the treatment strategy T_{kl} , x_{kl} and y_{kl} patients have respectively experienced dose-limiting toxicity and efficacy within the follow-up period, where $k = 1, \dots, K$ and $l = 1, \dots, L$. Let $\beta = \{\beta_0, \beta_1, \beta_2\}$ and $\gamma = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$ denote the regression coefficients in models (3) and (4). The likelihood function of the observed data $\mathcal{D} = \{x_{kl}, y_{kl}\}$ can be expressed as follows,

$$L(\mathcal{D} | \beta, \gamma, \omega) \propto \prod_{k=1}^K \prod_{l=1}^L p_{kl}^{x_{kl}} (1 - p_{kl})^{n_{kl} - x_{kl}} \times q_{kl}^{y_{kl}} (1 - q_{kl})^{n_{kl} - y_{kl}}.$$

Let $f(\beta)$, $f(\gamma)$, $f(\omega)$ denote the prior distributions for β , γ , and ω , respectively. Assuming prior independence among β , γ , and ω , we write the joint posterior distribution as

$$f(\beta, \gamma, \omega | \mathcal{D}) \propto L(\mathcal{D} | \beta, \gamma, \omega) f(\beta) f(\gamma) f(\omega),$$

from which the full conditional distributions can be obtained. The Gibbs sampler is used to obtain posterior draws of unknown parameters for statistical inferences.

3. Simulation Studies

3.1 Simulation Settings

The performance of the proposed design is evaluated through extensive simulation studies and compared to two standard designs: one (referred to as the factorial design) that assigns an equal number of cohorts to each strategy under investigation and another (referred to as the simple design) that identifies the optimal dose first and then determines the therapeutic window for the selected dose. For the factorial design, when none of the strategies is admissible, we deem the trial inconclusive. Similarly, we deem the trial that uses the simple design inconclusive when all the strategies at the identified dose level are not admissible. We consider $K = 5$ dose levels with $L = 5$ different treatment windows, such as 3 hours, 4 hours, 6 hours, 8 hours, and 10 hours since the onset of symptoms. For the weight score assigned to each treatment window, we set $w_1 = 0.2$, $w_2 = 0.4$, $w_3 = 0.6$, $w_4 = 0.8$, $w_5 = 1.0$, respectively. We set the highest tolerable toxicity level at $\phi = 0.3$ and the lowest acceptable efficacy level at $\psi = 0.4$. The cut-off values used in the admissible criteria (1) are set at $c_1 = 0.5$ and $c_2 = 0.5$. For the factorial design, we assign one cohort to each strategy, which results in the utilization of a sample size of $5 \times 5 \times 3 = 75$ if we consider a cohort size of 3. The simple design is equivalent to a design that consists of only the run-in stage of our proposed design, with $M_0 = K + S = 10$ cohorts. To obtain a similar sample size with the factorial design, we consider a cohort size of 8 to have a sample size of 80. For our proposed design, 10 cohorts are used in the run-in stage. Three iterative steps are conducted in the iterative selection stage, and the maximum number of cohorts used for each step is 7, 5, and 4, respectively. Therefore, the maximum sample size in our proposed design is $26 \times 3 = 78$, which is also close to that of the factorial design. We draw 2000 posterior samples to make inference after 1000 burn-in iterations. We investigate 12 scenarios with different toxicity and efficacy profiles, as shown in Table 1. In scenarios 1–8, the admissible strategies are shaded and the optimal strategy is shown in boldface. In contrast, in scenarios 9 to 12, none of the strategies is admissible and no optimal strategy exists.

3.2 Simulation Results

The selection percentages for each strategy under the proposed, factorial, and simple designs for all scenarios are summarized in Table 2. The average sample sizes for each scenario under the three designs are presented in Table 3. The proposed design is an adaptive design in which the sample size is not fixed. Therefore, we also present the standard deviation of the required sample size for the proposed design in Table 3. In scenario 1, the toxicity rate increases when the dose level increases, but is independent of the treatment window. For the schedule-efficacy relationship, the efficacy rate decreases when the treatment window increases. For the dose-efficacy relationship, the efficacy rate first increases and then decreases when the dose level increases. According to our admissible criteria (1), there are 13 admissible strategies, and T_{33} is the optimal strategy. Among the three designs, the proposed design correctly selects the optimal strategy with the highest percentage (30.8% for the proposed design vs. 25.7% for the factorial design vs. 20.9% for the simple design). In addition, the simple design utilizes the largest sample size and performs the worst among the three designs. These results suggest that the proposed design outperforms the two standard designs. In scenario 2, we consider the same efficacy profile as in scenario 1, but

with a slightly different toxicity profile. The results similarly show that our proposed design outperforms the two standard designs.

In scenario 3, we consider the same toxicity profile as in scenario 2, but with a different efficacy profile. Specifically, the efficacy rates are the same when the strategy is given within S_1 , S_2 and S_3 . When the treatment is administered outside of the window S_3 , the efficacy rate decreases with an increase in the treatment window. For the dose-efficacy relationship, the efficacy rate monotonically decreases with an increase in the dose level. There are 9 admissible doses and the optimal strategy is T_{13} . The selection percentage for T_{13} in the proposed design is 11.7% higher than that in the factorial design. In addition, the sample size used in the proposed design has 5 fewer animals than that in the factorial design. The simple design performs poorly and incorrectly selects T_{23} with the highest probability among all the strategies. Its selection percentage for the optimal strategy T_{13} is 20.2% lower compared with that of the proposed design. The poor performance of the simple design may be due to the incorrect identification of the dose level in the first step, which dramatically affects the decision of optimal strategy selection in the second step. In scenario 4, we consider the same efficacy profile as in scenario 3, but a different toxicity profile, which assumes that the toxicity rate monotonically increases when both the dose level and treatment window increase. The optimal strategy is T_{13} . In this scenario, the results suggest that the proposed design outperforms the other two designs. Our proposed design selects the optimal strategy with the highest selection percentage of the three designs, which is 8.2% higher than that of the standard design and 19.6% higher than that of the simple design. In the scenarios 5–8, the proposed design yields the best performance. In contrast, the simple design selects the optimal strategy with the lowest percentage among the three designs, especially in scenarios 7 and 8. These results demonstrate that our proposed adaptive shrinkage algorithm is powerful for allocating cohorts within the neighborhood of the optimal strategy, which results in the efficient use of resources and improves the overall performance of the trial design. The factorial design performs worse than the proposed design in selecting the optimal strategy, but performs better than the simple design.

In scenario 9, we consider the same toxicity profile as in scenario 2 and low efficacy rates for all strategies. Therefore, none of the strategies is admissible, and the trial should be terminated and deemed inconclusive. Our results show that the proposed design successfully deems the trial inconclusive all the time and only costs an average sample size of 30.5. The factorial and simple designs also successfully deem the trial inconclusive more than 97% of the time, but with the fixed sample size, they cost much larger sample sizes than the proposed design. In scenario 10, we consider the same toxicity profile as in scenario 1 and the same efficacy profile as in scenario 9. All of the strategies are non-admissible in scenario 10, and we observe findings similar to those in scenario 9. In scenario 11, although there are some strategies with high efficacy rates, the investigated strategies are too toxic and the trial should be terminated. The proposed design successfully deems the trial inconclusive 96.7% of the time, which is higher than the factorial (90.3% of the time) and simple designs (94.7% of the time). In scenario 12, we consider a more extreme case in which all the strategies are too toxic and have low efficacy. All these three designs deem the trial inconclusive 100% of the time, but our proposed design requires a much smaller sample size than the other two

designs. These results further confirm that our proposed design outperforms the factorial and simple designs when there is no strategy that satisfies our admissible criteria.

3.3 Sensitivity Analysis

We conduct a sensitivity analysis to evaluate the effect of the cut-offs c_1 and c_2 in the admissible criteria (1) on the design performance. Specifically, we consider $c_1 = c_2 = 0.05, 0.2, 0.4, 0.6$ for all 12 scenarios listed in Table 1 and compare their operating characteristics to that with $c_1 = c_2 = 0.5$, i.e., the simulations described in Section 3.2. We summarize the results in Table S1 in the supplementary materials, which lists the correct selection percentages of the optimal strategy for scenarios 1–8 and reports the percentages of inconclusive trials for scenarios 9–12. We also report the average sample size with the standard deviation of the proposed design for each scenario.

Our simulation results show that when the values of c_1 and c_2 increase from 0.4 to 0.6, both the percentages of correct decisions and the required sample sizes decrease, because the values of c_1 and c_2 control the stringency of the admissible criteria. Higher values of c_1 and c_2 lead to a more stringent admissible criteria, which results in exploring fewer admissible strategies in the iterative selection stage and a smaller required sample size. When $c_1 = c_2 = 0.2$, the percentages of correct decisions are slightly higher than those with $c_1 = c_2 = 0.4$ for most scenarios, but with larger sample sizes. When $c_1 = c_2 = 0.05$, although the percentages of the correct decisions and the required sample sizes are similar to those with $c_1 = c_2 = 0.2$, the design performance would not be satisfied in some scenarios, e.g., scenarios 6 and 11. The possible reason is that when $c_1 = c_2 = 0.05$, all the investigated strategies may meet the admissible criteria, and then in the iterative selection stage, the proposed design could be easily trapped in the strategies with high utilities and high toxicity rates. According to the simulation results, we recommend choosing the values for c_1 and c_2 from 0.4 to 0.6, which would provide a high correct decision rate and a relatively lower cost as well. In addition, for the recommended range of c_1 and c_2 , the proposed design consistently outperforms the factorial and simple designs with the highest percentages of making correct decisions.

4. Discussion

We propose a Bayesian multi-stage design to simultaneously identify the optimal dose and determine the therapeutic window for the experimental agent. By using the proposed utility score as a basis, our goal of identifying the optimal strategy by comparing both the treatment window and efficacy rate within the admissible set can be simplified as a comparison of the utility scores among all admissible strategies. By exploring the two-dimensional space quickly, the run-in stage of the proposed design provides an initial estimation of the admissible set for the iterative selection stage. In the iterative selection stage, an adaptive shrinkage algorithm is proposed to encourage the assignment of cohorts within the admissible strategies while requiring fewer cohorts than the previous iterative step. Our simulation results demonstrate that the proposed design performs well in identifying the optimal dose and determining the appropriate therapeutic window under different patterns of dose-schedule-response relationships while also saving resources. The shrinkage speed and the stopping criteria in the iterative selection stage can be calibrated through simulation

studies or on the basis of available resources. For example, given the number of animals available and the number of strategies to be investigated, we can calculate how many cohorts to use in the run-in stage and how many cohorts to use in the iterative selection stage. Such information can guide the determination of the shrinkage speed and the stopping criteria. Alternatively, given the stopping criteria and the number of strategies to be investigated, we can determine the maximum number of cohorts required and the maximum sample size. For the cohort size, we recommend using three to ensure the estimation precision while saving costs. A cohort size larger than 3 may be not necessary given the limited resources for the animal study.

Due to complex dose-schedule-response patterns, our assumed toxicity and efficacy models may be not able to precisely estimate the dose-schedule-response surface given the limited sample size in animal studies. This may result in an unreliable admissible set. Hence, instead of incorporating the posterior mean of the efficacy rate directly, we incorporate the posterior probability that the efficacy rate is acceptable into our utility score. This provides our algorithm with the flexibility to select the other strategies that are given within a shorter treatment window, but with strong evidence of their efficacy rates compared to the strategies given within the longest treatment window, but with low evidence of efficacy. In addition, the choice of both toxicity and efficacy models is flexible in our framework and can be decided according to the previous clinical knowledge of the experimental agent.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Hackam DG. Translating animal research into clinical benefit. *BMJ*. 2007; 7586:163.
2. Aban IB, George B. Statistical considerations for preclinical studies. *Exp Neurol*. 2015; 270:82–87. [PubMed: 25725352]
3. Kilkenny C, Parsons N, Kadyszewski E, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PloS one*. 2009; 4(11):e7824. [PubMed: 19956596]
4. Landis SC, Amara SG, Asadullah K, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012; 490(7419):187–191. [PubMed: 23060188]
5. Henderson VC, Kimmelman J, Fergusson D, et al. Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for in vivo animal experiments. *PLoS Med*. 2013; 10(7):e1001489. [PubMed: 23935460]
6. Perrin S. Preclinical research: Make mouse studies work. *Nature*. 2014; 507(7493):423–425. [PubMed: 24678540]
7. Chow, SC., Chang, M. Adaptive design methods in clinical trials. London: Chapman & Hall; 2011.
8. Sverdlov, O. Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects. London: Chapman & Hall; 2015.
9. Chow SC, Chang M, et al. Adaptive design methods in clinical trials-a review. *Orphanet J Rare Dis*. 2008; 3(11):169–90.
10. Berry, SM., Carlin, BP., Lee, JJ., et al. Bayesian adaptive methods for clinical trials. London: Chapman & Hall; 2010.
11. Lewis RJ, Berry DA. Group sequential clinical trials: a classical evaluation of bayesian decision-theoretic designs. *J Am Stat Assoc*. 1994; 89(428):1528–1534.

12. Di Carlo A. Human and economic burden of stroke. *Age Ageing*. 2009; 38(1):4–5. [PubMed: 19141505]
13. Hacke W, Kaste M, Fieschi C, et al. Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke: the european cooperative acute stroke study (ecass). *JAMA*. 1995; 274(13):1017–1025. [PubMed: 7563451]
14. Del Zoppo GJ, Saver JL, Jauch EC, et al. Expansion of the time window for treatment of acute ischemic stroke with intravenous tissue plasminogen activator a science advisory from the american heart association/american stroke association. *Stroke*. 2009; 40(8):2945–2948. [PubMed: 19478221]
15. Evenson KR, Rosamond WD, Morris DL. Prehospital and in-hospital delays in acute stroke care. *Neuroepidemiology*. 2001; 20(2):65–76. [PubMed: 11359072]
16. Walter S, Kostopoulos P, Haass A, et al. Diagnosis and treatment of patients with stroke in a mobile stroke unit versus in hospital: a randomised controlled trial. *Lancet Neurol*. 2012; 11(5): 397–404. [PubMed: 22497929]
17. Levine SR, Gorman M. telestroke the application of telemedicine for stroke. *Stroke*. 1999; 30(2): 464–469. [PubMed: 9933289]
18. Thall PF, Russell KE. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase i/ii clinical trials. *Biometrics*. 1998:251–264. [PubMed: 9544520]
19. Gelman A, Jakulin A, Pittau MG, et al. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008:1360–1383.

Table 1

Toxicity and efficacy probabilities of each strategy for 12 scenarios. The admissible strategies are shaded and the optimal one is shown in boldface.

Scenario	Treatment Window	Dose Level									
		Toxicity probability					Efficacy probability				
		1	2	3	4	5	1	2	3	4	5
1	5	0.05	0.08	0.12	0.18	0.25	0.15	0.20	0.25	0.20	0.15
	4	0.05	0.08	0.12	0.18	0.25	0.25	0.28	0.32	0.28	0.22
	3	0.05	0.08	0.12	0.18	0.25	0.35	0.38	0.50	0.38	0.30
2	2	0.05	0.08	0.12	0.18	0.25	0.45	0.50	0.60	0.50	0.40
	1	0.05	0.08	0.12	0.18	0.25	0.50	0.60	0.72	0.60	0.45
	5	0.06	0.08	0.12	0.18	0.25	0.15	0.20	0.25	0.20	0.15
3	4	0.06	0.08	0.12	0.18	0.25	0.25	0.28	0.32	0.28	0.22
	3	0.05	0.06	0.09	0.13	0.20	0.35	0.38	0.50	0.38	0.30
	2	0.05	0.06	0.09	0.13	0.20	0.45	0.50	0.60	0.50	0.40
4	1	0.05	0.06	0.09	0.13	0.20	0.50	0.60	0.72	0.60	0.45
	5	0.06	0.08	0.12	0.18	0.25	0.24	0.20	0.15	0.10	0.05
	4	0.06	0.08	0.12	0.18	0.25	0.32	0.25	0.20	0.15	0.10
5	3	0.05	0.06	0.09	0.13	0.20	0.65	0.55	0.45	0.35	0.20
	2	0.05	0.06	0.09	0.13	0.20	0.65	0.55	0.45	0.35	0.20
	1	0.05	0.06	0.09	0.13	0.20	0.65	0.55	0.45	0.35	0.20
6	5	0.25	0.40	0.45	0.55	0.55	0.24	0.20	0.15	0.10	0.05
	4	0.22	0.25	0.38	0.50	0.55	0.32	0.25	0.20	0.15	0.10
	3	0.15	0.20	0.26	0.45	0.50	0.65	0.55	0.45	0.35	0.20
7	2	0.10	0.15	0.23	0.36	0.45	0.65	0.55	0.45	0.35	0.20
	1	0.05	0.10	0.18	0.28	0.35	0.65	0.55	0.45	0.35	0.20
	5	0.15	0.22	0.22	0.35	0.52	0.10	0.15	0.20	0.25	0.30
8	4	0.12	0.18	0.20	0.32	0.48	0.20	0.25	0.30	0.35	0.55
	3	0.10	0.16	0.18	0.26	0.45	0.30	0.35	0.50	0.60	0.70
	2	0.08	0.14	0.16	0.24	0.42	0.40	0.50	0.60	0.70	0.80
9	1	0.05	0.12	0.15	0.22	0.38	0.40	0.50	0.60	0.70	0.80
	5	0.21	0.24	0.40	0.45	0.55	0.10	0.15	0.20	0.25	0.30

Scenario	Treatment Window	Dose Level									
		Toxicity probability					Efficacy probability				
		1	2	3	4	5	1	2	3	4	5
7	4	0.18	0.20	0.25	0.40	0.45	0.20	0.25	0.30	0.35	0.55
	3	0.12	0.15	0.20	0.25	0.40	0.30	0.35	0.50	0.60	0.70
	2	0.08	0.10	0.15	0.20	0.25	0.40	0.50	0.60	0.70	0.80
	1	0.02	0.05	0.10	0.15	0.21	0.40	0.50	0.60	0.70	0.80
	5	0.21	0.24	0.40	0.45	0.55	0.20	0.25	0.30	0.25	0.20
	4	0.18	0.20	0.25	0.40	0.45	0.25	0.30	0.35	0.30	0.25
	3	0.12	0.15	0.20	0.25	0.40	0.35	0.40	0.60	0.45	0.30
	2	0.08	0.10	0.15	0.20	0.25	0.55	0.62	0.75	0.60	0.45
	1	0.02	0.05	0.10	0.15	0.21	0.55	0.62	0.75	0.60	0.45
	5	0.25	0.40	0.45	0.55	0.55	0.20	0.25	0.30	0.25	0.20
8	4	0.22	0.25	0.38	0.50	0.55	0.25	0.30	0.35	0.30	0.25
	3	0.15	0.20	0.26	0.45	0.50	0.35	0.40	0.60	0.45	0.30
	2	0.10	0.15	0.23	0.36	0.45	0.55	0.62	0.75	0.60	0.45
	1	0.05	0.10	0.18	0.28	0.35	0.55	0.62	0.75	0.60	0.45
	5	0.06	0.08	0.12	0.18	0.25	0.01	0.02	0.10	0.05	0.04
	4	0.06	0.08	0.12	0.18	0.25	0.02	0.05	0.15	0.08	0.06
	3	0.05	0.06	0.09	0.13	0.20	0.05	0.10	0.20	0.12	0.08
	2	0.05	0.06	0.09	0.13	0.20	0.08	0.15	0.25	0.16	0.10
	1	0.05	0.06	0.09	0.13	0.20	0.10	0.20	0.30	0.22	0.15
	5	0.05	0.08	0.12	0.18	0.25	0.01	0.02	0.10	0.05	0.04
10	4	0.05	0.08	0.12	0.18	0.25	0.02	0.05	0.15	0.08	0.06
	3	0.05	0.08	0.12	0.18	0.25	0.05	0.10	0.20	0.12	0.08
	2	0.05	0.08	0.12	0.18	0.25	0.08	0.15	0.25	0.16	0.10
	1	0.05	0.08	0.12	0.18	0.25	0.10	0.20	0.30	0.22	0.15
	5	0.05	0.08	0.12	0.18	0.25	0.01	0.02	0.10	0.05	0.04
	4	0.05	0.08	0.12	0.18	0.25	0.02	0.05	0.15	0.08	0.06
	3	0.05	0.08	0.12	0.18	0.25	0.05	0.10	0.20	0.12	0.08
	2	0.05	0.08	0.12	0.18	0.25	0.08	0.15	0.25	0.16	0.10
	1	0.05	0.08	0.12	0.18	0.25	0.10	0.20	0.30	0.22	0.15
	5	0.40	0.45	0.50	0.55	0.60	0.15	0.20	0.25	0.20	0.15
11	4	0.40	0.45	0.50	0.55	0.60	0.25	0.28	0.32	0.28	0.22
	3	0.40	0.45	0.50	0.55	0.60	0.35	0.38	0.50	0.38	0.30
	2	0.40	0.45	0.50	0.55	0.60	0.45	0.50	0.60	0.50	0.40
	1	0.40	0.45	0.50	0.55	0.60	0.50	0.60	0.72	0.60	0.45
	5	0.40	0.45	0.50	0.55	0.60	0.15	0.20	0.25	0.20	0.15
	4	0.40	0.45	0.50	0.55	0.60	0.25	0.28	0.32	0.28	0.22
	3	0.40	0.45	0.50	0.55	0.60	0.35	0.38	0.50	0.38	0.30
	2	0.40	0.45	0.50	0.55	0.60	0.45	0.50	0.60	0.50	0.40
	1	0.40	0.45	0.50	0.55	0.60	0.50	0.60	0.72	0.60	0.45
	5	0.40	0.45	0.50	0.55	0.60	0.15	0.20	0.25	0.20	0.15

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Scenario	Treatment Window	Dose Level									
		Toxicity probability					Efficacy probability				
		1	2	3	4	5	1	2	3	4	5
12	5	0.40	0.45	0.50	0.55	0.60	0.01	0.02	0.10	0.05	0.04
	4	0.40	0.45	0.50	0.55	0.60	0.02	0.05	0.15	0.08	0.06
	3	0.40	0.45	0.50	0.55	0.60	0.05	0.10	0.20	0.12	0.08
	2	0.40	0.45	0.50	0.55	0.60	0.08	0.15	0.25	0.16	0.10
	1	0.40	0.45	0.50	0.55	0.60	0.10	0.20	0.30	0.22	0.15

Table 2

Percentage of selecting each strategy as the optimal one for twelve scenarios. The admissible strategies are shaded and the optimal one is shown in boldface.

Scenario	Treatment Window	Dose Level														
		Proposed Design					Factorial Design					Simple Design				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	5	0.0	0.1	1.1	0.1	0.1	0.7	0.7	3.3	0.5	0.4	0.0	1.2	4.2	1.1	0.0
	4	1.0	1.1	7.0	1.5	0.5	2.1	2.9	11.8	3.5	1.4	0.2	6.0	12.9	3.4	0.1
	3	5.1	9.0	30.8	10.2	1.8	5.9	11.1	25.7	11.5	3.4	1.3	17.2	20.9	12.4	0.3
2	2	1.8	6.0	8.8	5.3	1.1	1.1	2.2	4.9	1.7	0.7	1.1	7.8	2.4	5.0	1.0
	1	0.1	0.0	0.2	0.1	0.1	0.0	0.1	0.2	0.1	0.0	0.1	0.1	0.0	0.1	0.1
	5	0.1	0.1	0.8	0.1	0.0	0.5	0.7	3.2	0.7	0.5	0.0	1.3	3.6	0.9	0.0
3	4	0.5	1.0	7.2	1.8	0.3	2.4	2.9	11.8	3.4	1.6	0.2	6.3	12.2	3.9	0.1
	3	5.6	8.6	31.9	9.7	2.3	5.4	10.9	25.9	11.3	3.4	1.1	18.1	21.1	12.2	0.4
	2	1.9	6.9	7.6	3.9	1.4	1.3	2.8	4.7	1.5	0.7	1.2	7.7	2.4	5.1	0.9
4	1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.1
	5	2.1	0.0	0.0	0.0	0.0	7.8	0.5	0.0	0.0	0.0	4.1	1.9	0.1	0.0	0.0
	4	20.6	1.6	0.2	0.0	0.0	22.6	5.2	0.7	0.0	0.0	16.4	10.1	0.2	0.0	0.0
5	3	43.8	13.2	2.5	0.1	0.1	32.1	19.1	3.3	0.1	0.1	23.6	30.3	2.0	0.1	0.0
	2	1.7	1.7	0.9	0.1	0.0	2.6	2.1	0.4	0.1	0.1	1.1	5.7	1.7	0.2	0.1
	1	0.0	0.0	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.2	0.0	0.0
6	5	1.6	0.0	0.0	0.0	0.0	4.5	0.3	0.0	0.0	0.0	3.4	1.1	0.1	0.0	0.0
	4	20.5	1.7	0.2	0.0	0.0	20.4	4.0	0.2	0.0	0.0	17.0	7.3	0.2	0.0	0.0
	3	43.8	12.0	2.1	0.1	0.0	35.6	20.2	2.1	0.0	0.0	24.2	32.3	1.5	0.0	0.0
7	2	2.5	3.0	1.4	0.1	0.0	4.3	3.6	0.7	0.1	0.1	1.1	6.6	2.4	0.2	0.0
	1	0.1	0.0	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.0
	5	0.0	0.1	0.5	1.2	0.3	0.0	0.0	0.8	1.4	2.3	0.0	0.0	0.7	2.5	0.1
8	4	0.1	0.4	4.2	10.7	1.5	0.9	0.0	5.1	13.7	7.0	0.1	0.4	4.5	13.8	0.1
	3	0.9	1.5	20.0	32.9	4.1	2.0	0.9	18.4	29.9	6.3	0.3	1.8	9.1	29.0	0.8
	2	0.4	0.7	4.3	8.7	1.8	0.2	0.4	3.8	5.3	0.4	0.9	1.6	1.9	10.2	3.0
9	1	0.0	0.1	0.1	0.4	0.1	0.1	0.0	0.0	0.2	0.1	0.2	0.0	0.0	2.1	1.8

Scenario	Treatment Window	Dose Level														
		Proposed Design					Factorial Design					Simple Design				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
6	5	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.4	0.8	1.4	0.0	0.0	0.3	1.5	0.2
	4	0.2	0.1	3.2	6.2	3.7	0.7	0.0	4.2	11.3	10.4	0.0	0.3	1.5	11.2	2.3
	3	0.4	0.9	14.5	41.2	14.1	0.8	0.5	12.1	33.6	16.2	0.1	1.5	3.9	38.3	7.7
	2	0.2	0.4	1.6	7.5	5.1	0.1	1.3	4.7	1.2	1.2	0.2	0.9	0.5	9.4	13.2
	1	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.2	3.2
7	5	0.1	0.1	0.9	0.0	0.0	0.7	1.8	4.0	0.4	0.0	0.1	2.5	2.6	0.4	0.0
	4	1.3	2.6	12.6	1.7	0.0	2.6	7.8	18.9	2.5	0.2	0.4	10.8	13.5	2.5	0.1
	3	4.6	13.6	38.2	9.8	0.8	3.1	14.1	28.4	10.2	0.8	1.7	23.1	19.1	10.5	0.1
	2	0.8	4.3	3.9	2.2	0.4	0.5	0.7	1.4	0.7	0.3	1.1	5.6	0.4	4.2	0.7
	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0
8	5	0.1	0.5	0.4	0.0	0.0	0.7	1.7	1.1	0.0	0.0	0.1	1.8	1.1	0.0	0.0
	4	1.1	4.7	5.8	0.1	0.0	2.6	9.1	6.9	0.4	0.0	0.3	9.5	7.8	0.1	0.0
	3	5.5	22.9	34.5	1.4	0.0	5.6	25.7	28.6	2.7	0.1	2.1	25.4	18.6	1.6	0.0
	2	0.9	7.7	9.6	2.1	0.2	0.9	5.1	5.5	1.1	0.1	0.9	6.9	9.0	5.9	0.1
	1	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.3	0.1	0.0	0.0	0.0	1.0	3.2	0.1
9	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.8	0.1	0.1	0.0	0.0	0.2	0.0	0.0
	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.1	0.2	0.0	0.0	0.2	0.1	0.0
10	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.1	0.1	0.0	0.0	0.2	0.0	0.0
	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.1	0.2	0.0	0.0	0.2	0.1	0.0
11	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	0.0	0.0	0.2	0.0	0.0	0.1	0.0	0.9	0.1	0.1	0.0	0.0	0.3	0.0	0.0
	1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.4	0.1	0.2	0.0	0.0	0.3	0.1	0.0

Scenario	Treatment Window	Dose Level														
		Proposed Design					Factorial Design					Simple Design				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
12	1	0.6	0.1	0.0	0.0	0.0	1.7	0.5	0.1	0.0	0.0	1.2	0.4	0.2	0.0	0.0
	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 3

Average sample sizes $\pm$ standard deviations of the three designs.

Scenario	Average Sample Size Standard $\pm$ Deviation		
	Proposed Design	Factorial Design	Simple Design
1	73.8 $\pm$ 11.2	75	80
2	73.6 $\pm$ 11.5	75	80
3	69.5 $\pm$ 15.0	75	80
4	68.7 $\pm$ 14.8	75	80
5	71.8 $\pm$ 11.5	75	80
6	74.6 $\pm$ 8.5	75	80
7	75.7 $\pm$ 8.4	75	80
8	72.8 $\pm$ 10.9	75	80
9	30.5 $\pm$ 2.7	75	80
10	30.6 $\pm$ 3.2	75	80
11	33.8 $\pm$ 8.5	75	80
12	30.0 $\pm$ 0.8	75	80

The sample sizes of the factorial and simple designs are fixed.