Published in final edited form as:

Stat Med. 2016 September 30; 35(22): 3892–3906. doi:10.1002/sim.6971.

MIDAS: A Practical Bayesian Design for Platform Trials with Molecularly Targeted Agents

Ying Yuan^{1,*}, Beibei Guo², Mark Munsell¹, Karen Lu³, and Amir Jazaeri^{3,*}

¹Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, U.S.A

²Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA, U.S.A

³Department of Gynecologic Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, U.S.A

Abstract

Recent success of immunotherapy and other targeted therapies in cancer treatment has led to an unprecedented surge in the number of novel therapeutic agents that need to be evaluated in clinical trials. Traditional phase II clinical trial designs were developed for evaluating one candidate treatment at a time, and thus not efficient for this task. We propose a Bayesian phase II platform design, the Multi-candidate Iterative Design with Adaptive Selection (MIDAS), which allows investigators to continuously screen a large number of candidate agents in an efficient and seamless fashion. MIDAS consists of one control arm, which contains a standard therapy as the control, and several experimental arms, which contain the experimental agents. Patients are adaptively randomized to the control and experimental agents based on their estimated efficacy. During the trial, we adaptively drop inefficacious or overly toxic agents and “graduate” the promising agents from the trial to the next stage of development. Whenever an experimental agent graduates or is dropped, the corresponding arm opens immediately for testing the next available new agent. Simulation studies show that MIDAS substantially outperforms the conventional approach. The proposed design yields a significantly higher probability for identifying the promising agents and dropping the futile agents. In addition, MIDAS requires only one master protocol, which streamlines trial conduct and substantially decreases the overhead burden.

Keywords

Platform design; Platform trials; Phase II trials; Bayesian adaptive design

1. Introduction

The landmark success of drugs such as ipilimumab and nivolumab has rejuvenated the field of immuno-oncology and has led to intense focus on other co-stimulatory and co-inhibitory

*Correspondence to: Ying Yuan, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, U.S.A. yyuan@mdanderson.org; Amir Jazaeri, Department of Gynecologic Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, U.S.A. AAJazaeri@mdanderson.org.

immune pathways. This research has produced a substantial and increasing number of novel immunotherapies, exponentially increasing the number of available molecularly targeted agents available for cancer treatment. This provides medical researchers and the pharmaceutical industry with unprecedented opportunities to make advances against cancer. At the same time, it challenges the traditional one-treatment-at-a-time clinical trial paradigm that has been plagued by high cost, slow progress, and a high failure rate.

Traditional phase II clinical trial designs were developed mainly for evaluating candidate treatments in a one treatment at a time fashion. Commonly used phase II designs include Simon's two-stage design [1] and its extensions [2, 3, 4, 5, 6, 7], and Bayesian phase II designs based on posterior probabilities [8, 9, 10], predictive probabilities [11], and Bayes factors [12], among others. Two-arm randomized phase II trials have increasingly been used to assess new drugs with the progression free survival (PFS) as the endpoint [13, 14].

The traditional phase II trial designs are cumbersome and grossly inefficient when the number of agents to be tested is large and increases over time, due to several reasons. First, the traditional approach requires the separate development of a research protocol, regulatory compliance, funding, and related infrastructure for each trial, which results in excessive “time-to-clinical use”, overhead and overlap of infrastructure. In addition, the “white” space between trials slows down the process of testing candidate agents. Using the traditional approach in multiple randomized trials of different single agents requires the repeated use of the control arm (i.e., the standard therapy), which wastes patient and medical resources. Furthermore, differences among trials (e.g., different patient populations, different treatment procedures, and possibly different treatments as the control) make it difficult, if not impossible, to compare candidate agents across trials without bias. Traditional multi-arm trials can alleviate some of these issues, but still face significant logistic and statistical difficulties. For example, the inherent challenges associated with conducting a large multi-center, multi-agent clinical trial, limit the feasibility of conducting a single multi-arm trial that includes all the candidate agents. Even when such a design is possible, during the course of the trial emerging clinical data on the agents being tested is likely to change the equipoise between the arms and may even make some agents irrelevant. Thus, several independent multiple-arm trials are still needed, resulting in difficulties already described. Finally, standard multi-arm trials do not have a built-in mechanism to allow researchers to adaptively add new agents to the ongoing trial. The ability to accommodate the addition of new agents during a trial is important in the current, rapidly evolving research environment where new agents are developed at unprecedented speeds.

To address the limitations associated with traditional clinical testing approaches, we propose a Bayesian phase II trial design called the MIDAS. This design allows investigators to continuously screen a large number of candidate agents in an efficient and seamless fashion. MIDAS requires only one master protocol (and hence a single regulatory and contractual approval) and one control arm, which streamlines the trial conduct, decreases the overhead, saves resources, and provides a straightforward comparison of the experimental agents. MIDAS consists of one control arm, which contains a standard therapy as the control, and an adjustable number of experimental arms, which contain the experimental agents. Patients are adaptively randomized to the control and experimental agents based on their estimated

efficacy. During the trial, we adaptively drop inefficacious or overly toxic agents from the trial and graduate the promising agents to the next stage of development. Whenever an experimental agent graduates or is dropped from the trial, the corresponding arm opens immediately for testing the next available new agent. In principle, MIDAS allows for an unending screening process that can efficiently select promising agents or combinations for further clinical development.

In order to illustrate the clinical application of MIDAS and contrast it with a traditional phase II approach, we designed a phase II clinical trial aimed at testing a number of new immunotherapy/targeted agents for treating patients diagnosed with recurrent platinum-resistant epithelial ovarian cancer (EOC). The operating characteristics of each approach were determined using simulation. Since the FDA's approval of cisplatin/paclitaxel in 1991, there has been no drug or combination that has improved the overall survival (OS) of patients with EOC beyond that obtained from existing therapies. Thus, the development of new strategies (e.g., immunotherapy) to improve the OS of patients with EOC represents a crucial unmet need. The initial set of experimental agents to be evaluated in this trial includes three novel immunotherapy agents, A, B and C, and their combinations, A+B, B+C, A+C, and A+B+C. We used simulation to test the MIDAS approach to evaluate the efficacy of these agents and their combinations in comparison with the physician's chosen therapy as the control. While only 7 agents/combinations were initially tested in the trial, more new agents or their combinations, for example combining some or all of A, B and C with chemotherapy, may be added during the trial based on the trial data and clinical considerations.

MIDAS belongs to an emerging class of adaptive designs called platform designs, which can be broadly defined as trial designs that aim to find the best treatment for a disease by simultaneously investigating multiple agents in an adaptive way [15, 16]. This is in contrast to the traditional clinical trials that focus on evaluating a single treatment in a single population. Clinical trials, such as BATTLE [17] and I-SPY2 [18], provide early examples of adaptively investigating multiple agents in a trial.

The remainder of this article is organized as follows. We describe MIDAS and the statistical model in Section 2. In Section 3, we present simulation studies to investigate the operating characteristics of the proposed design. In Section 4, we conduct sensitivity analyses to further investigate the properties of our method. We conclude with a brief discussion in Section 5.

2. Methods

2.1. Probability model

We assume that the efficacy endpoint is the progression-free survival T (i.e., PFS), and toxicity is measured by a binary variable Y , with $Y = 1$ indicating that a patient experienced dose-limiting toxicity. We choose to measure efficacy using the PFS rather than a simpler binary outcome (tumor response/no response) because the PFS is a better surrogate for the OS, which is the endpoint of ultimate interest. This is especially true for immunotherapy and other targeted agents, which often delay cancer progression and prolong patient survival

without substantially shrinking the tumor. Compared to the approach of directly using the OS as the endpoint, the use of the PFS allows us to screen a large number of candidate agents in a reasonable time frame because the PFS requires a short followup time than the OS. In the case that the OS is an appropriate endpoint (e.g., for relapsed cancer patients having short survival time), our design is also applicable.

MIDAS evaluates multiple treatments simultaneously. We model each treatment independently without imposing any specific structure across treatments. Specifically, we assume that the time to disease progression, T , follows an exponential distribution with hazard θ ,

$$T \sim \text{Exp}(\theta).$$

Let n denote the number of patients who have been assigned to a treatment arm at the time of any given interim decision. For the i th patient, where $i = 1, \dots, n$, let T_i^o denote the observed time of failure or administrative right censoring, and let $\delta = I(T_i^o = T_i)$ denote the censoring indicator. Let $S(t) = \Pr(T > t)$ denote the survival function, which takes the form of $S(t) = e^{-t\theta}$ under the exponential model. Given the interim data $\mathcal{D}_n = \{(T_i^o, \delta_i), i = 1, \dots, n\}$, let $m = \sum_{i=1}^n \delta_i$ denote the total number of failures and $\tilde{T} = \sum_{i=1}^n T_i^o$ denote the total observation time. The likelihood is given by

$$L(\mathcal{D}_n | \theta) = \prod_{i=1}^n f(T_i^o | \theta)^{\delta_i} S(T_i^o | \theta)^{1-\delta_i} = \theta^m \exp(-\tilde{T}\theta).$$

We assign θ a conjugate gamma prior, $\theta \sim \text{Ga}(a, b)$, with pdf $f(\theta | a, b) = b^a \theta^{a-1} e^{-b\theta} / \Gamma(a)$, where $\Gamma(\cdot)$ is the gamma function and $a, b > 0$ are fixed hyperparameters. The posterior distribution of θ follows a gamma distribution,

$$\theta | \mathcal{D}_n \sim \text{Ga}(a+m, b+\tilde{T}). \quad (1)$$

Caution is needed when choosing hyper-parameters a and b . The straightforward approach of setting a and b at very small values (e.g., 0.001) is not appropriate here. This is because at the early stage of the trial, we may not observe any event (i.e., disease progression) in the control arm, i.e., $m = 0$. The resulting posterior distribution of θ , i.e., $\text{Ga}(a+m, b+\tilde{T})$, has substantial mass at the values close to 0, which causes highly unstable posterior estimate of the hazard ratio of the experimental agent versus the control because the denominator (i.e., the hazard of control) is extremely close to 0. We suggest the following procedure to set hyper-parameters a and b . We first elicit from clinicians the prior estimate of median survival time η and its low limiting $\tilde{\eta}$, which correspond to hazards $\log(2)/\eta$ and $\log(2)/\tilde{\eta}$, respectively, under the exponential distribution. We choose the values of a and b such that

the median of $Ga(a, b)$ matches $\log(2)/\eta$ and the prior probability of $\theta > \log(2)/\tilde{\eta}$ is small (e.g., 5%).

The assumption that the time to disease progression T follows an exponential distribution is a strong parametric assumption, but well serves our design purposes. Our goal is not to precisely estimate the whole survival curve, but to identify the agents that have excessively low or high efficacy based on the median PFS. Thall, Wooten, and Tannir [19] showed that the exponential model is remarkably robust for monitoring efficacy based on the median survival time. Our sensitivity analysis in Section 4 shows that our design is robust to the exponential distribution assumption.

We assume a standard beta-binomial model for toxicity. Let z denote the number of patients who have experienced toxicity out of the n patients at the time of an interim decision. We assume that z follows a binomial distribution,

$$z \sim \text{Binom}(n, p),$$

where p is the toxicity probability. Assigning a conjugate beta prior to p , say $p \sim \text{Beta}(a_T, b_T)$, the posterior distribution of p is given by

$$p|z, n \sim \text{Beta}(a_T + z, b_T + n - z).$$

To set hyperparameters a_T and b_T , we elicit an estimate of toxicity from clinicians, and then choose the values of a_T and b_T such that the prior mean of p matches that elicited estimate, while controlling the prior effective sample size $a + b$ equal a small value, like $a + b = 1$.

We have specified the marginal distributions of T and Y . We choose not to model the joint distribution of T and Y for several reasons. First, our decision rules (described later) rely on only the marginal distributions of T and Y , thus our approach is sufficient for decision making. Second, in a MIDAS trial, the main focus is to evaluate the efficacy of new agents. We assume that the safety of these agents has been established in early phase I trials. In other words, here T is the primary endpoint and Y is the secondary endpoint, and thus the joint modeling of T and Y may not be warranted. Third, unlike traditional cytotoxic agents, because of a different mechanism for treating cancer, high toxicity from immunotherapeutic agents or other novel molecularly targeted agents does not lead to high efficacy. In other words, the correlation between efficacy and toxicity tends to be weak for immunotherapeutic agents. Thus, joint modeling does not provide much efficiency gain for estimating efficacy and toxicity, but makes the estimation much more complicated.

2.2. Trial design

As illustrated in Figure 1, the MIDAS trial includes one control slot/arm and K experimental slots/arms. The control slot contains the standard therapy, which serves as the control arm. Each of the K experimental slots contains a new agent for testing, resulting in K experimental arms. Let arm 0 denote the control arm, and arm k denote the k th experimental

arm, with λ_k denote the true hazard ratio between the k th experimental arm, $k = 1, \dots, K$, versus the control. By definition, $\lambda_0 \equiv 1$ for the control. Let p_k denote the toxicity probability for the agent tested in the k th experimental arm, and ϕ_T denote the upper limit of acceptable toxicity. The MIDAS trial can be described as follows:

1. The trial starts by assigning the standard treatment to the control slot and K new experimental agents to K experimental slots.
2. As a burn-in, the first $(K+1)N_{min}$ enrolled patients are equally randomized to $K+1$ treatment arms, i.e., each treatment arm receives N_{min} patients.
3. For patients enrolled subsequently, the general strategy is to adaptively randomize them to arm k with probability π_k that is proportional to

$$\alpha_k = \text{pr}\{\lambda_k = \min(\lambda_0, \lambda_1, \dots, \lambda_K) | \text{data}\}, k=0, 1, \dots, K, \quad (2)$$

which is the posterior probability that arm k is the best treatment that is associated with the lowest hazard. By doing so, we are more likely to randomize patients to efficacious treatments. To accommodate practical considerations, we include the following three adjustments for π_k .

(3a) *Catch-up rule*: One important feature of the MIDAS trial is its ability to continuously test experimental agents. Once an experimental agent is removed from the trial due to findings of superiority, excessive toxicity, or futility, the slot that contained that agent immediately opens for testing another new agent (details are provided in steps 4 and 5 below). It is desirable to skew the randomization probability for the newly added agent such that the new agent has a higher chance of receiving patients and its sample size thereby catches up with those of the agents that have been studied in the trial for a certain period of time. Let n_k denote the number of patients that have been assigned to arm k , and define $\mathcal{C} = \{k, n_k < N_{min}\}$ as the set of experimental arms that need to catch up. We use N_{min} as the cutoff because that is the number of patients initially treated by each of the first K experimental arms after equal randomization. For an experimental arm $k \in \mathcal{C}$, we set the randomization probability as

$$\pi_k = \max\left(\frac{\alpha_k}{\sum_{i=0}^K \alpha_i}, \frac{1}{K+1}\right), k \in \mathcal{C}.$$

This catch-up rule ensures that the treatment arms with fewer than N_{min} patients have a probability of receiving new patients that is at least $1/(K+1)$. As $\pi_k \geq 1/(K+1)$, it is possible, although very rare, $\sum_{k \in \mathcal{C}} \pi_k > 1$. When that occurs, we normalize them to have a sum of 1. We use $\bar{\mathcal{C}} = \{k, n_k \geq N_{min}\}$ to denote the experimental arms that do not need to catch up in terms of sample size.

(3b) *Curtail rule*: Under MIDAS, the control arm remains active until the trial is terminated. As a result, it is likely that a large number of patients will be assigned to the control arm after several rounds of screening new agents. This is not desirable because treating an excessive number of patients on the control arm wastes precious medical resources and contributes little to learning the efficacy of the experimental agents. In addition, it reduces the chance that patients receive the new treatments that are potentially more effective. Thus, we impose the following rule to curtail π_0 , the probability of a patient being randomized to the control arm:

$$\pi_0 = \begin{cases} (1 - \sum_{k \in \mathcal{C}} \pi_k) \frac{\alpha_0}{\alpha_0 + \sum_{k \in \mathcal{C}} \alpha_k} & \text{if } n_0 < \beta N_{\max} \\ (1 - \sum_{k \in \mathcal{C}} \pi_k) \frac{\alpha_0}{\alpha_0 + \sum_{k \in \mathcal{C}} \alpha_k} \left(\frac{\beta N_{\max}}{n_0} \right)^\gamma & \text{if } n_0 \geq \beta N_{\max}. \end{cases}$$

where $0 < \beta < 1$ is a constant, $N_{\max}$ is the prespecified maximum sample size for each of the experimental arms, and the term $(1 - \sum_{k \in \mathcal{C}} \pi_k)$ reflects that we have “spent” some randomization probabilities, i.e., $\sum_{k \in \mathcal{C}} \pi_k$, for the agents that need to catch up. The first equation of this rule says that before the number of patients treated on the control arm (i.e., n_0) reaches a certain threshold, $\beta N_{\max}$, we randomize patients to the control with the probability proportional to the posterior probability that the control is the best treatment (i.e., α_0). However, once n_0 reaches $\beta N_{\max}$, as defined by the second equation, we tune down the randomization probability π_0 to assign fewer patients to the control via the term $(\beta N_{\max}/n_0)^\gamma$, where γ controls how quickly we tune down the randomization probability. We note that this is not the only way to tune down π_0 ; other approaches are certainly possible. A numerical study shows that $(\beta N_{\max}/n_0)^\gamma$ serves its purpose and controls the possibility of assigning too many patients to the control arm.

(3c) After applying the above two rules, the randomization probability for each of the remaining experimental arms, i.e., $k \in \mathcal{C}$, is proportional to the posterior probability that arm k is associated with the lowest hazard,

$$\pi_k = (1 - \pi_0 - \sum_{k \in \mathcal{C}} \pi_k) \frac{\alpha_k}{\sum_{k \in \mathcal{C}} \alpha_k}, k \in \mathcal{C}.$$

4. During the trial, we conduct the following interim monitoring after each patient or each cohort of patients is randomized.

(4a) Drop the agent from slot k if either of the following two conditions is true:

$$(\text{stopping for futility}) \text{pr}(\lambda_k < \phi_E | \text{data}) < C_{E1}, \quad (3)$$

$$(\text{stopping for toxicity}) \text{pr}(p_k > \phi_T | \text{data}) > C_T, \quad (4)$$

where ϕ_E (1) represents the upper limit of the hazard ratio for the experimental agent to be regarded as promising, ϕ_T represents the upper limit of the toxicity, and C_{E1} and C_T are prespecified cutoffs, which should be calibrated through simulations for good design operating characteristics. The first condition says that, conditional on the observed data, the agent has a small chance to have a hazard ratio lower than ϕ_E , and the second condition says that the agent has a high (posterior) probability of being overly toxic.

(4b) Graduate the agent from slot k if one of the following two conditions is true:

$$(\text{graduate for superiority}) \text{pr}(\lambda_k < \phi_E | \text{data}) > C_{E2} \quad (5)$$

$$(\text{graduate for saturation}) n_k \geq N_{\max}, \quad (6)$$

where C_{E2} is a prespecified cutoff. The first condition says that, conditional on the observed data, the agent has a high probability of having a hazard ratio lower than ϕ_T , thus the agent should graduate from the trial and move forward to further development (e.g., a large confirmative trial). The second condition reflects the practical consideration that when a certain number of patients have been treated with an experimental agent and that agent has not been dropped for being overly toxic or futile, the agent may be worth developing further and should graduate from the current trial. The agents that graduate can be prioritized on the basis of various clinical considerations for further development in large-scale confirmatory trials. For an arm with none of the stopping and graduation rules (3)-(6) being satisfied, we will continue to randomize patients to that arm.

To reduce sampling variation, we add a minimum sample size requirement that an agent must be used to treat at least $N_{\min}$ patients before it graduates for superiority or is dropped for futility. In addition, we require that the last patient treated by an experimental agent be followed at least τ months before that agent is dropped for futility. In other words, when the futility stopping rule (3) is satisfied, we do not immediately drop the agent, but

continue to follow the patients in that arm for at least τ months to confirm its futility. During this confirmation period, no new patients will be enrolled into that arm. If it turns out that the agent does not satisfy the futility stopping rule, the arm will be reopen for enrolling new patients.

When we evaluate the stopping and graduation rules (3) to (6), all control patients are used to calculate hazard ratio λ_k . This approach is efficient, but may introduce biases for the experimental agent introduced later in the trial if patient characteristics change during recruitment. If that is a concern, we may restrict the comparison to the concurrent controls that are enrolled in the same time frame as the experimental arm. In that case, we may set $\gamma = 0$ in the curtail rule such that a reasonable number of patients will be concurrently randomized to the control arm.

5. Whenever an experimental agent is dropped or graduates, the corresponding slot immediately opens for testing a new agent. If needed, a phase I lead-in can be used to confirm the safety of the candidate agent before it enters the trial. We use the Bayesian optimal interval (BOIN) design [20] as the phase I lead-in because of its simplicity and superior performance. The BOIN design is easy to implement in a way similar to the 3+3 design, but yields performance comparable to the model-based continuous reassessment method (CRM) [21].
6. Continue steps 3 to 5 until all experimental agents in the candidate pool have been tested.

As MIDAS continuously screens a large number of candidate agents, one may concern about the multiplicity or type I/II rate. Since MIDAS is intended for phase II trials and aims to fast screen a large number of candidate agents, strictly controlling the type I/II error rate may not be necessary. The multiplicity or type I/II error rate can be empirically controlled by calibrating the design parameters to obtain desirable operating characteristics (e.g., reasonably low false positive and false negative), as shown next in simulation studies. The software (coded in R) to implement the MIDAS design will be freely available for downloading from http://odin.mdacc.tmc.edu/~yyuan/index_code.html.

3. Simulation Studies

We investigated the operating characteristics of MIDAS through simulation studies. We considered a total of 7 experimental agents, tested against the control (i.e., the standard therapy). We labeled the 7 experimental agents as agents 1 to 7 (see Table 1). We set the median PFS = 5 months (i.e., the hazard $\theta = 0.139$) for the control and considered various hazard ratios (HR) for the experimental agents (versus the control). A maximum number of $K = 3$ experimental agents were allowed to be tested simultaneously with $N_{min} = 10$ and $N_{max} = 40$. We assumed that patient's arrival follows a Poisson distribution with the rate of 6 patients per month. We set $\beta = 0.35$ and $\gamma = 2$ such that we started to tune down the randomization probability to the control once the number of patients assigned to the control reached 35% of N_{max} . Before dropping an agent due to futility, we required the last patient in that arm be followed at least 3 months, i.e., $\tau = 3$. The toxicity upper bound $\phi_T = 0.3$ and

the efficacy bound $\phi_E = 0.7$. We took the probability cutoffs $C_{E1} = 0.4$, $C_{E2} = 0.95$, $C_T = 0.94$. These cutoffs were calibrated to achieve overall desirable design operating characteristics. For example, $C_T = 0.94$ was chosen so that the probability of dropping an agent due to toxicity is low (high) for an agent with a low (high) toxicity probability. $C_{E1} = 0.4$ and $C_{E2} = 0.95$ were chosen to yield desirable probabilities of graduating promising agents and dropping futile agents. Specifically, we calibrated C_{E1} and C_{E2} so that the probability of dropping a futile agent with an HR of 0.83 is 65-70% and the probability of graduating a promising agent with an HR of 0.56 is 66-76%. For the toxicity model, we used a vague beta prior, $p \sim \text{Beta}(0.3, 0.7)$, which corresponds to an effective sample size of 1. For the PFS model, following the approach described previously, we used gamma prior $\theta \sim \text{Ga}(0.37, 0.88)$ such that the prior median PFS = 5 months and the prior probability that the median PFS < 0.3 months is 5%, and the prior probability that the median PFS < 2 months is 33%. We updated the adaptive randomization probabilities and monitored toxicity and efficacy by a cohort size of 3.

We constructed 10 scenarios for the 8 agents in the trial. The 10 scenarios cover different cases in terms of the number of promising new agents, the order in which the agents enter the trial, and the toxicity rates of the agents. In scenarios 1-3, there are 2 promising new agents, with respective HR of 0.56 and 0.42. The difference between these three scenarios is the order in which the agents enter the trial. For example, in scenario 1, the two promising agents enter the trial immediately after the initiation of the trial, recalling that we allow a maximum of 3 experimental agents to be tested simultaneously. In scenario 2, the two promising agents enter the trial late. In scenario 3, one promising agent (i.e., agent 1) enters the trial early and the other (i.e., agent 5) enters the trial late. Scenarios 4-8 have 3, 4, 5, 1, and 0 promising agents, respectively. As described, the primary objective of MIDAS is to evaluate the efficacy of the experimental agents, the safety of which has been evaluated in phase I trials. Therefore, in scenarios 1 to 8, we assume that the agents are generally safe, with toxicity rates below the toxicity upper bound of $\phi_T = 0.3$. Nevertheless, due to the small sample sizes of phase I trials, an agent may turn out to be toxic even though it has been tested in a phase I trial. To account for that case, we constructed scenarios 9 and 10, in which some agents are overly toxic. Under each scenario, we simulated 1000 trials.

We compared MIDAS to the conventional approach, which consisted of 7 independent 2-arm trials. In each of these 2-arm trials, a maximum of N patients were equally randomized to the control and the experimental agent. The same graduation and stopping criteria (i.e., equations (3)-(6)) were used to monitor the efficacy and toxicity of the experimental agent. For ease of comparison, we calibrated the sample size N such that the total sample size of the seven 2-arm trials matched that of MIDAS for each of the 10 scenarios. For the 2-arm trials, it is possible that none of the criteria (3)-(6) was satisfied when the maximum sample size was reached. In that case, we graduated the agent since the futility stopping rule was not met.

Table 1 summarizes the operating characteristics of MIDAS and the conventional approach, and Table 2 presents the average trial duration and the average sample size. In scenario 1, agents 2 and 3 are promising, with respective PFS HRs of 0.56 and 0.42. The total sample size was about 161. MIDAS assigned substantially fewer patients to the control arm than the

conventional approach. On average, 44.5 and 81 patients were treated on the control arm under MIDAS and the conventional approach, respectively, which correspond to 28% and 50% of the total patients. The graduation percentages for the two promising agents were about 10% higher under MIDAS than the conventional approach (66.2% & 86.4% vs 57.8% & 75.6%). This is because MIDAS assigned more patients to the effective treatment arms, thereby improving the power to identify the promising agents. For example, for the two promising agents with HR of 0.56 and 0.42, the average numbers of patients were 24.3 and 25.2 under MIDAS and 15.0 and 16.3 under the conventional approach. Agents 1 and 4-7 are futile, with a HR that was greater than or equal to 0.7. Both MIDAS and the conventional approach were able to drop the futile agents with large probabilities. Scenarios 2 and 3 are the same as scenario 1 except that the agents enter the trial in a different order. The operating characteristics of MIDAS are generally similar to those under scenario 1. The slight changes (e.g., 66.2% and 75.8% graduation for an agent with HR 0.56 under scenarios 1 and 2, respectively) were due to the fact that the accrual rate was faster for the agent that entered the trial later because at the later stage there may be fewer active arms to compete for patient accrual as some agents have been dropped/graduated. In scenarios 4-7, there are 3, 4, 5, and 1 promising agents, respectively. The results are generally similar to those obtained under scenarios 1-3. That is, MIDAS outperformed the conventional approach with substantially fewer patients allocated to the control, and with higher probabilities of identifying the promising and futile agents. In scenario 8, none of the new agents is promising. MIDAS had similar or higher probabilities of identifying and dropping the futile agents. For example, for agent 4, which achieved a PFS HR of 1.67, the percentages for dropping that agent due to futility were 97.8% under MIDAS and 91.4% under the conventional design. Scenarios 9 and 10 were constructed to consider that case in which some agents are overly toxic, under which MIDAS also generally outperformed the conventional design. For example, in scenario 9, agents 4 and 6 are associated with promising PFS (PFS HRs of 0.56 and 0.5), but are overly toxic (toxicity rates of 0.6 and 0.55, respectively). MIDAS dropped these two agents 89.3% and 83.5% of the time, whereas the conventional approach dropped these two agents 82.6% and 77% of the time. In scenario 10, agent 7 is effective but overly toxic. MIDAS was 15.3% more likely (85.9% versus 70.6%) to identify and drop that agent.

Another important benefit of using the MIDAS, compared to the conventional approach, is that the MIDAS dramatically shortened the trial duration by removing the “white space” between trials (see Table 2). We assume that, under the conventional approach, the 2-arm trials are done sequentially and there is a 6-month “white space” (i.e., the time to prepare protocols, obtain approval from the IRB, and set up infrastructures for enrolling patients and running the trials) between two consecutive trials. The MIDAS cuts the trial duration by more than half in the 10 scenarios. For example, in scenario 1, the average trial duration was 2.3 years under MIDAS and 5.3 years under the conventional approach. For the purpose of illustration, we considered the extreme case that the seven 2-arm trials are conducted sequentially under the conventional approach. In a real clinical development, some of these trials may be conducted in a simultaneous fashion, and thus the advantage of MIDAS over the conventional approach, in terms of trial duration, may be less striking as these reported in Table 2. However, the conventional approach still requires development of multiple research protocols and related infrastructures, which may result in excessive “time-to-

clinical use” and overhead. Note that the results described previously on the probabilities of graduating effective agents and dropping futile agents (i.e., Table 1) are not affected by whether the seven 2-arm trials are conducted sequentially or simultaneously.

4. Sensitivity Analyses

In our design, we assume that PFS follows an exponential distribution that has a constant hazard. To examine the robustness of our design to violations of this assumption, we carried out sensitivity analyses by simulating data from distributions with non-constant hazard functions. We considered four shapes of non-constant hazard functions: hump-shaped, increasing, decreasing, and linearly increasing shapes (see Figure 2). For the control arm, the hump-shaped hazard functions were generated from a log-logistic distribution with a shape parameter of 2, and the increasing, decreasing, and linearly increasing hazard functions were generated from Weibull distributions, with respective shape parameters of 1.2, .75 and 2. The scale parameters of the Weibull distribution and log-logistic distribution were selected such that the median survival time was 5 months. Assuming proportional hazard, the hazard functions for the experimental arms were generated with HRs listed in Table 1. Figure 3 depicts the percentage of times agents were dropped due to futility, the percentage of times agents graduated, and the number of patients assigned to each of the 7 new agents for two representative scenarios, scenarios 1 and 4. We can see that the operating characteristics of the proposed design are generally robust across different shapes of the hazard functions, with slight differences.

We also evaluated the performance of the proposed design with different values of γ , i.e., the parameter that controls the rate of tuning down the randomization probability for the control arm. Tables 3 and 4 show the operating characteristics of the proposed design with $\gamma = 1$ and $\gamma = 3$, respectively, for the first five scenarios. The results were very similar to these with $\gamma = 2$. Based on the numerical study (results for other scenarios are not shown), we recommend $\gamma = 2$ for practical use.

Lastly, we briefly evaluated the performance of the proposed design when each experimental agent was compared only to its concurrent controls (i.e., the patients who were concurrently randomized to the control arm), rather than all cumulative patients in the control arm. This approach avoids the potential bias caused by the drift of patient prognosis, i.e., patient prognosis changes over time. Because of using the curtail rule and adaptive randomization as described in Section 2.2, it is likely that at interims there are no concurrent controls. Thus we used approximate concurrent controls, which contained the concurrent controls, if existed, and the latest 5 patients enrolled into the control arm just before the concurrent controls. As shown in Table 5, the operating characteristics of the proposed design using the concurrent controls were generally similar to those reported in Table 1, but we did see some variations. For example, in Scenario 1, the futility stopping for agent 5 dropped from 97.1% to 93.8%. As patient prognosis drifting often occurs in practice, several important issues warrant further investigation: for example, how does the MIDAS perform when patient prognoses drift, and how to calibrate the curtail rule and randomization to allow reliable comparison between experimental arms and the concurrent controls.

5. Discussion

We have proposed a Bayesian adaptive platform trial design, MIDAS, which is able to continuously test a large number of agents. This work was motivated by the urgent need for flexible trial designs to efficiently screen a large number of immunotherapeutic agents and their combinations. MIDAS consists of one control slot and several experimental slots. During the trial conduct, MIDAS adaptively drops a futile or toxic agent, graduates a promising agent, and adds new agents for testing. A significant advantage of MIDAS is that replacement agents can be selected in a “just in time” fashion based on the most up-to-date preclinical and clinical data. This data may in fact be generated during the earlier iteration of the same trial. For example, treatment with agent A may result in the activation of compensatory drug resistant pathway b, providing the rationale for a proposed combination of A + anti-b as a replacement regimen.

Although we have focused on immunotherapy and targeted agents, MIDAS can be directly used to evaluate other treatment agents. One limitation of MIDAS is that it does not account for potential subgroup effects. A straightforward approach to address that issue is to include each patient's biomarker or characteristics as the covariates in the PFS model, based on which we adaptively and selectively drop or graduate experimental agents for each subgroup. This is a topic for further research. Simulation results demonstrate that MIDAS substantially outperforms conventional approaches that evaluate one agent at a time.

Acknowledgments

Yuan's research was partially supported by Award Number 5P50CA098258 and R01CA154591, and Munsell and Lu's research was partially supported by Award Number 5P50CA098258 from the National Cancer Institute.

References

1. Simon RM. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials*. 1989; 10:1–10. [PubMed: 2702835]
2. Chen TT. Optimal three-stage designs for phase II cancer clinical trials. *Statistics in Medicine*. 1997; 16:2701–2711. [PubMed: 9421870]
3. Ensign LG, Gehan EA, Kamen DS, Thall PF. An optimal three-stage design for phase II clinical trials. *Statistics in Medicine*. 1994; 13:1727–1736. [PubMed: 7997706]
4. Hanfelt JJ, Slack RS, Gehan EA. A modification of Simon's optimal design for phase II trials when the criterion is median sample size. *Controlled Clinical Trials*. 1999; 20:555–566. [PubMed: 10588296]
5. Jung SH, Carey M, Kim KM. Graphical search for two-stage designs for phase II clinical trials. *Controlled Clinical Trials*. 2001; 22:367–372. [PubMed: 11514038]
6. Shuster J. Optimal two-stage designs for single arm phase II cancer trials. *Journal of Biopharmaceutical Statistics*. 2002; 12:39–51. [PubMed: 12146719]
7. Lin Y, Shih WJ. Adaptive two-stage designs for single-arm phase IIA cancer clinical trials. *Biometrics*. 2004; 60:482–490. [PubMed: 15180674]
8. Thall PF, Simon RM. Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics*. 1994; 50:337–349. [PubMed: 7980801]
9. Thall PF, Simon RM, Estey EH. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statistics in Medicine*. 1995; 14:357–379. [PubMed: 7746977]
10. Heitjan DF. Bayesian interim analysis of phase II cancer clinical trials. *Statistics in Medicine*. 1997; 16:1791–1802. [PubMed: 9280033]

11. Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. *Clinical Trials*. 2008; 5:93–106. [PubMed: 18375647]
12. Johnson VE, Cook JD. Bayesian design of single-arm phase II clinical trials with continuous monitoring. *Clinical Trials*. 2009; 6:217–226. [PubMed: 19528131]
13. Korn EL, Arbuck SG, Pluda JM, Simon R, Kaplan RS, Christian MC. Clinical trial designs for cytostatic agents: are new approaches needed? *Journal of Clinical Oncology*. 2001; 19:265–272. [PubMed: 11134222]
14. Rubinstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, Smith MA. Design issues of randomized phase II trials and a proposal for phase ii screening trials. *Journal of Clinical Oncology*. 2005; 23:7199–7206. [PubMed: 16192604]
15. Mullard A. Multicompany trials adapt todisciplines beyond cancer. *Nature Medicine*. 2014; 20:3.
16. Berry SM, Connor JT, Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. *Journal of the American Medical Association*. 2015; 313:1619–1620. [PubMed: 25799162]
17. Zhou X, Liu S, Kim ES, Herbst RS, Lee JJ. Bayesian adaptive design for targeted therapy development in lung cancer—a step toward personalized medicine. *Clinical Trials*. 2008; 5:181–193. [PubMed: 18559407]
18. Barker AD, Sigman CC, Kelloff GJ, et al. I-spy 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clinical Pharmacology & Therapeutics*. 2009; 86:97–100. [PubMed: 19440188]
19. Thall PF, Wooten LH, Tannir NM. Monitoring event times in early phase clinical trials: some practical issues. *Clinical Trials*. 2005; 2:467–478. [PubMed: 16422307]
20. Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. *Journal of the Royal Statistical Society: Series C*. 2015; 64:507–523.
21. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: A practical design for phase I clinical trials in cancer. *Biometrics*. 1990; 46:33–48. [PubMed: 2350571]

Figure 1. Diagram of the MIDAS trial

Figure 2. Four shapes (decreasing, increasing, linear, and hump-shaped) of hazard functions for PFS in the sensitivity analyses.

Figure 3. Sensitivity analyses when hazard functions for PFS follow a non-constant shape. The three panels on the left are for scenario 1, and the three panels on the right are for scenario 4. In each panel, the x-axis contains the 7 new agents considered in the trial. For each agent, the five bars represent constant, decreasing, increasing, linear, and hump-shaped hazard functions, respectively, from left to right.

Operating characteristics of MIDAS and the conventional approach to trial conduct. The results of the conventional approach are shown in parentheses.

Table 1

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
Scenario 1							
Control	1.00	0.15	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	44.5 (81.0)
1	0.83	0.03	0.0	0.0 (0.0)	69.4 (68.8)	30.6 (31.2)	19.1 (13.2)
2	0.56	0.04	0.0	0.0 (0.4)	33.8 (41.8)	66.2 (57.8)	24.3 (15.0)
3	0.42	0.03	0.0	0.0 (0.2)	13.6 (24.2)	86.4 (75.6)	25.2 (16.3)
4	1.25	0.05	9.3	0.4 (0.2)	90.9 (90.2)	8.7 (9.6)	14.3 (10.5)
5	1.67	0.04	12.7	0.1 (0.4)	97.1 (96.8)	2.8 (2.8)	12.0 (9.2)
6	2.50	0.04	16.3	0.0 (0.2)	100.0 (99.6)	0.0 (0.2)	10.7 (8.5)
7	2.50	0.03	19.5	0.2 (0.0)	99.3 (99.8)	0.5 (0.2)	11.0 (8.5)
Scenario 2							
Control	1.00	0.15	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	38.8 (76.3)
1	1.67	0.03	0.0	0.0 (0.0)	98.7 (96.0)	1.3 (4.0)	11.5 (9.2)
2	2.50	0.04	0.0	0.0 (0.0)	99.9 (99.4)	0.1 (0.6)	10.2 (8.4)
3	1.25	0.03	0.0	0.0 (0.0)	93.7 (88.0)	6.3 (12.0)	14.6 (10.3)
4	0.42	0.05	9.2	0.1 (0.4)	10.7 (24.4)	89.2 (75.2)	21.8 (14.5)
5	0.83	0.04	10.0	0.1 (0.2)	64.5 (64.2)	35.4 (35.6)	20.5 (12.0)
6	0.56	0.04	11.4	0.0 (0.2)	24.2 (42.8)	75.8 (57.0)	25.5 (13.2)
7	2.50	0.03	16.4	0.0 (0.0)	99.8 (99.4)	0.2 (0.6)	10.8 (8.5)
Scenario 3							
Control	1.00	0.15	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	38.6 (78.6)
1	0.42	0.03	0.0	0.0 (0.2)	13.9 (28.2)	86.1 (71.6)	25.1 (15.2)
2	2.50	0.04	0.0	0.0 (0.2)	100.0 (99.6)	0.0 (0.2)	10.2 (8.5)
3	2.50	0.03	0.0	0.0 (0.0)	100.0 (99.6)	0.0 (0.4)	10.3 (8.4)
4	0.83	0.05	8.7	0.0 (0.2)	65.4 (65.2)	34.6 (34.6)	19.7 (12.7)
5	0.56	0.04	9.7	0.0 (0.0)	25.9 (43.8)	74.1 (56.2)	24.3 (14.2)
6	1.67	0.04	12.7	0.2 (0.0)	97.6 (97.2)	2.2 (2.8)	11.6 (9.1)
7	1.25	0.03	17.1	0.0 (0.0)	89.6 (89.4)	10.4 (10.6)	15.5 (10.4)

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
Scenario 4							
Control	1.00	0.15	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	40.0 (83.9)
1	0.45	0.03	0.0	0.0 (0.0)	19.7 (30.8)	80.3 (69.2)	24.2 (15.1)
2	0.83	0.04	0.0	0.0 (0.2)	70.3 (70.2)	29.7 (29.6)	18.4 (12.4)
3	0.50	0.03	0.0	0.0 (0.2)	26.6 (33.6)	73.4 (66.2)	24.3 (14.8)
4	0.56	0.05	9.2	0.2 (0.4)	26.8 (40.2)	73.0 (59.0)	24.0 (14.6)
5	1.67	0.04	12.4	0.3 (0.2)	98.4 (97.4)	1.3 (2.4)	11.5 (9.1)
6	1.25	0.04	16.5	0.1 (0.2)	89.2 (90.6)	10.7 (9.2)	14.9 (10.4)
7	2.50	0.03	20.1	0.1 (0.4)	99.4 (98.8)	0.5 (0.8)	10.8 (8.3)
Scenario 5							
Control	1.00	0.1	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	25.0 (83.9)
1	0.50	0.04	0.0	0.0 (0.4)	24.6 (33.2)	75.4 (66.4)	25.9 (13.0)
2	0.83	0.03	0.0	0.0 (0.0)	70.7 (57.6)	29.3 (42.4)	19.1 (11.9)
3	2.50	0.04	0.0	0.0 (0.0)	99.6 (99.4)	0.4 (0.6)	10.2 (8.5)
4	0.42	0.02	8.8	0.0 (0.0)	11.4 (24.8)	88.6 (75.0)	23.2 (13.4)
5	1.25	0.06	10.9	0.4 (0.2)	89.9 (79.8)	9.7 (19.8)	13.7 (10.5)
6	0.56	0.05	14.6	0.0 (0.4)	23.9 (36.0)	76.1 (63.6)	25.4 (13.1)
7	0.45	0.04	18.3	0.0 (0.6)	12.3 (30.4)	87.7 (69.0)	25.7 (13.1)
Scenario 6							
Control	1.00	0.1	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	22.7 (84.7)
1	0.56	0.04	0.0	0.0 (0.4)	33.0 (38.0)	67.0 (61.6)	24.0 (12.9)
2	1.67	0.03	0.0	0.0 (0.0)	98.7 (92.8)	1.3 (7.2)	10.9 (9.4)
3	0.45	0.04	0.0	0.0 (0.2)	18.4 (25.8)	81.6 (74.0)	24.6 (13.4)
4	0.56	0.02	8.7	0.1 (0.0)	27.1 (41.0)	72.8 (59.0)	23.9 (12.7)
5	1.25	0.06	11.4	0.4 (0.6)	91.1 (84.8)	8.5 (14.6)	13.9 (10.0)
6	0.42	0.05	15.4	0.0 (0.2)	8.9 (22.6)	91.1 (77.2)	24.4 (13.8)
7	0.50	0.04	18.9	0.1 (0.0)	13.6 (33.6)	86.3 (66.4)	25.9 (13.1)
Scenario 7							
Control	1.00	0.1	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	50.7 (71.8)
1	5.00	0.04	0.0	0.0 (0.0)	100.0 (100.0)	0.0 (0.0)	10.0 (8.0)

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to fatality	Graduation	
2	1.67	0.03	0.0	0.0 (0.0)	98.8 (97.8)	1.2 (2.2)	12.0 (9.4)
3	2.50	0.04	0.0	0.0 (0.4)	99.9 (99.6)	0.1 (0.0)	10.2 (8.5)
4	1.25	0.02	9.1	0.0 (0.0)	93.8 (94.8)	6.2 (5.2)	13.3 (11.2)
5	0.50	0.06	9.7	0.2 (0.6)	19.0 (40.0)	80.8 (59.4)	25.2 (17.1)
6	2.50	0.05	10.5	0.2 (0.6)	99.4 (99.4)	0.4 (0.0)	10.4 (8.5)
7	1.67	0.04	16.2	0.1 (0.0)	98.1 (98.6)	1.8 (1.4)	12.6 (9.4)
Scenario 8							
Control	1.00	0.1	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	63.4 (69.0)
1	5.00	0.04	0.0	0.0 (0.4)	100.0 (99.4)	0.0 (0.2)	10.0 (9.0)
2	2.50	0.03	0.0	0.0 (0.0)	99.9 (99.8)	0.1 (0.2)	10.3 (9.5)
3	2.50	0.04	0.0	0.0 (0.2)	99.9 (99.0)	0.1 (0.8)	10.2 (9.6)
4	1.67	0.02	9.1	0.0 (0.4)	97.8 (91.4)	2.2 (8.2)	11.7 (11.3)
5	1.67	0.06	9.7	0.3 (0.6)	98.6 (93.2)	1.1 (6.2)	11.5 (11.2)
6	2.50	0.05	10.1	0.3 (0.2)	99.2 (99.4)	0.5 (0.4)	10.6 (9.6)
7	5.00	0.04	16.5	0.0 (0.2)	99.9 (99.8)	0.1 (0.0)	10.5 (9.0)
Scenario 9							
Control	1.00	0.07	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	26.3 (70.8)
1	0.45	0.05	0.0	0.0 (0.2)	21.6 (30.2)	78.4 (69.4)	23.7 (14.3)
2	0.50	0.04	0.0	0.0 (0.2)	24.5 (35.4)	75.5 (64.4)	24.2 (13.7)
3	2.50	0.5	0.0	37.1 (52.2)	62.9 (47.6)	0.0 (0.2)	10.1 (5.9)
4	0.56	0.6	7.9	89.3 (82.6)	4.9 (9.6)	5.8 (7.8)	7.7 (6.5)
5	1.67	0.04	9.6	0.2 (0.4)	97.4 (95.6)	2.4 (4.0)	11.7 (9.4)
6	0.50	0.55	12.3	83.5 (77.0)	4.3 (10.8)	12.2 (12.2)	9.2 (7.2)
7	0.56	0.03	14.4	0.2 (0.4)	21.0 (36.0)	78.8 (63.6)	27.3 (13.8)
Scenario 10							
Control	1.00	0.07	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	30.4 (66.5)
1	1.67	0.03	0.0	0.0 (0.0)	98.7 (93.6)	1.3 (6.4)	11.4 (9.1)
2	0.50	0.05	0.0	0.0 (0.6)	22.2 (30.6)	77.8 (68.8)	25.9 (13.0)
3	1.67	0.05	0.0	0.0 (0.4)	98.5 (91.4)	1.5 (8.2)	11.5 (9.4)
4	2.50	0.5	8.8	52.1 (53.2)	47.4 (46.4)	0.5 (0.4)	7.7 (6.0)

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
5	2.50	0.04	9.9	0.1 (0.0)	99.6 (99.6)	0.3 (0.4)	10.5 (8.4)
6	0.56	0.02	12.3	0.0 (0.2)	19.6 (33.0)	80.4 (66.8)	26.5 (13.2)
7	0.45	0.55	15.4	85.9 (70.6)	1.7 (11.6)	12.4 (17.8)	9.5 (7.1)

Table 2

Average trial duration and average sample size

scenario	1	2	3	4	5	6	7	8	9	10
	Average trial duration (months)									
MIDAS	27.3	26.2	26.4	28.6	28.5	28.9	24.5	23.6	23.9	22.7
2-arm trials	63.3	62.2	62.4	64.6	64.5	64.9	60.5	59.6	59.9	58.7
	Average sample size									
	161.1	153.6	155.3	168.1	168.1	170.3	144.4	138.4	140.1	133.4

Table 3Operating characteristics of MIDAS with $\gamma = 1$.

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
Scenario 1							
Control	1.00	0.15	0.0	0.0	0.0	0.0	44.6
1	0.83	0.03	0.0	0.0	67.7	32.3	19.3
2	0.56	0.04	0.0	0.0	32.3	67.7	24.2
3	0.42	0.03	0.0	0.0	13.2	86.8	25.3
4	1.25	0.05	9.3	0.5	90.4	9.1	14.2
5	1.67	0.04	12.7	0.0	97.5	2.5	12.0
6	2.50	0.04	16.4	0.2	99.6	0.2	10.7
7	2.50	0.03	19.6	0.1	99.3	0.6	10.9
Scenario 2							
Control	1.00	0.15	0.0	0.0	0.0	0.0	38.9
1	1.67	0.03	0.0	0.0	99.2	0.8	11.3
2	2.50	0.04	0.0	0.0	100.0	0.0	10.2
3	1.25	0.03	0.0	0.0	94.3	5.7	14.2
4	0.42	0.05	9.2	0.1	11.2	88.7	21.9
5	0.83	0.04	10.0	0.0	64.8	35.2	20.3
6	0.56	0.04	11.3	0.0	24.8	75.2	25.4
7	2.50	0.03	16.3	0.0	99.8	0.2	10.7
Scenario 3							
Control	1.00	0.15	0.0	0.0	0.0	0.0	38.7
1	0.42	0.03	0.0	0.0	13.6	86.4	25.8
2	2.50	0.04	0.0	0.0	99.9	0.1	10.1
3	2.50	0.03	0.0	0.0	100.0	0.0	10.2
4	0.83	0.05	8.6	0.3	64.5	35.2	19.6
5	0.56	0.04	9.7	0.0	25.8	74.2	24.8
6	1.67	0.04	12.9	0.0	97.0	3.0	11.7
7	1.25	0.03	17.1	0.1	90.2	9.7	15.4

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
Scenario 4							
Control	1.00	0.15	0.0	0.0	0.0	0.0	40.5
1	0.45	0.03	0.0	0.0	20.6	79.4	25.1
2	0.83	0.04	0.0	0.0	73.3	26.7	18.7
3	0.50	0.03	0.0	0.0	26.4	73.6	24.5
4	0.56	0.05	9.4	0.0	28.2	71.8	24.8
5	1.67	0.04	12.9	0.5	97.3	2.2	11.4
6	1.25	0.04	16.7	0.0	90.2	9.8	14.8
7	2.50	0.03	20.4	0.1	99.7	0.2	10.9
Scenario 5							
Control	1.00	0.10	0.0	0.0	0.0	0.0	25.6
1	0.50	0.04	0.0	0.0	26.3	73.7	25.2
2	0.83	0.03	0.0	0.0	73.3	26.7	18.7
3	2.50	0.04	0.0	0.0	99.9	0.1	10.1
4	0.42	0.02	8.8	0.0	12.2	87.8	23.6
5	1.25	0.06	10.8	0.2	90.8	9.0	13.6
6	0.56	0.05	14.5	0.2	26.9	72.9	24.7
7	0.45	0.04	18.1	0.2	10.4	89.4	26.6

Table 4

Operating characteristics of MIDAS with $\gamma = 3$.

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
Scenario 1							
Control	1.00	0.15	0.0	0.0	0.0	0.0	44.6
1	0.83	0.03	0.0	0.0	69.3	30.7	19.2
2	0.56	0.04	0.0	0.0	33.8	66.2	24.4
3	0.42	0.03	0.0	0.0	13.3	86.7	25.0
4	1.25	0.05	9.3	0.2	91.1	8.7	14.3
5	1.67	0.04	12.7	0.0	97.3	2.7	11.9
6	2.50	0.04	16.4	0.0	100.0	0.0	10.8
7	2.50	0.03	19.6	0.1	99.4	0.5	11.0
Scenario 2							
Control	1.00	0.15	0.0	0.0	0.0	0.0	39.6
1	1.67	0.03	0.0	0.0	98.8	1.2	11.6
2	2.50	0.04	0.0	0.0	99.9	0.1	10.2
3	1.25	0.03	0.0	0.0	94.1	5.9	14.4
4	0.42	0.05	9.2	0.0	11.2	88.8	23.5
5	0.83	0.04	10.1	0.3	66.8	32.9	19.5
6	0.56	0.04	11.4	0.0	25.4	74.6	24.2
7	2.50	0.03	16.5	0.1	99.7	0.2	10.8
Scenario 3							
Control	1.00	0.15	0.0	0.0	0.0	0.0	38.7
1	0.42	0.03	0.0	0.0	15.6	84.4	24.9
2	2.50	0.04	0.0	0.0	99.9	0.1	10.2
3	2.50	0.03	0.0	0.0	99.9	0.1	10.2
4	0.83	0.05	8.6	0.0	62.0	38.0	20.2
5	0.56	0.04	9.6	0.1	23.2	76.7	24.5
6	1.67	0.04	12.7	0.1	97.8	2.1	11.5
7	1.25	0.03	17.0	0.1	90.0	9.9	15.7

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
Scenario 4							
Control	1.00	0.15	0.0	0.0	0.0	0.0	39.5
1	0.45	0.03	0.0	0.0	20.5	79.5	24.5
2	0.83	0.04	0.0	0.0	70.2	29.8	19.3
3	0.50	0.03	0.0	0.0	25.0	75.0	24.5
4	0.56	0.05	9.3	0.1	26.7	73.2	24.3
5	1.67	0.04	12.7	0.2	97.2	2.6	11.6
6	1.25	0.04	16.7	0.3	88.3	11.4	14.9
7	2.50	0.03	20.4	0.1	99.4	0.5	10.9
Scenario 5							
Control	1.00	0.10	0.0	0.0	0.0	0.0	25.5
1	0.50	0.04	0.0	0.0	23.0	77.0	25.0
2	0.83	0.03	0.0	0.0	70.6	29.4	19.3
3	2.50	0.04	0.0	0.0	100.0	0.0	10.1
4	0.42	0.02	8.8	0.0	12.2	87.8	23.4
5	1.25	0.06	10.8	0.6	90.9	8.5	14.0
6	0.56	0.05	14.4	0.2	24.5	75.3	24.8
7	0.45	0.04	18.2	0.0	12.4	87.6	25.9

Table 5

Operating characteristics of MIDAS using concurrent controls.

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
Scenario 1							
Control	1.00	0.15	0.0	0.0	0.0	0.0	45.7
1	0.83	0.03	0.0	0.0	69.6	30.4	19.4
2	0.56	0.04	0.0	0.0	34.1	65.9	24.0
3	0.42	0.03	0.0	0.0	12.9	87.1	24.0
4	1.25	0.05	9.2	0.2	86.7	13.1	15.7
5	1.67	0.04	12.7	0.0	93.8	6.2	13.6
6	2.50	0.04	16.7	0.1	98.8	1.1	11.3
7	2.50	0.03	20.1	0.1	97.6	2.3	11.7
Scenario 2							
Control	1.00	0.15	0.0	0.0	0.0	0.0	38.5
1	1.67	0.03	0.0	0.0	98.5	1.5	11.4
2	2.50	0.04	0.0	0.0	99.9	0.1	10.2
3	1.25	0.03	0.0	0.0	93.9	6.1	14.3
4	0.42	0.05	9.1	0.1	13.4	86.5	24.1
5	0.83	0.04	10.0	0.0	59.3	40.7	21.7
6	0.56	0.04	11.4	0.1	25.6	74.3	25.7
7	2.50	0.03	16.9	0.1	98.5	1.4	11.3
Scenario 3							
Control	1.00	0.15	0.0	0.0	0.0	0.0	37.6
1	0.42	0.03	0.0	0.0	11.4	88.6	24.6
2	2.50	0.04	0.0	0.0	99.9	0.1	10.2
3	2.50	0.03	0.0	0.0	100.0	0.0	10.1
4	0.83	0.05	8.7	0.6	58.2	41.2	21.6
5	0.56	0.04	9.7	0.0	25.0	75.0	25.9
6	1.67	0.04	12.8	0.3	93.5	6.2	12.9
7	1.25	0.03	17.6	0.0	80.1	19.9	18.1

Agent	Hazard Ratio	True toxicity rate	Entry Time (Months)	Percentage of			Number of patients
				Dropped due to toxicity	Dropped due to futility	Graduation	
Scenario 4							
Control	1.00	0.15	0.0	0.0	0.0	0.0	40.0
1	0.45	0.03	0.0	0.0	17.8	82.2	24.5
2	0.83	0.04	0.0	0.0	70.6	29.4	18.6
3	0.50	0.03	0.0	0.0	25.4	74.6	24.3
4	0.56	0.05	9.3	0.1	25.2	74.7	24.2
5	1.67	0.04	12.7	0.0	95.4	4.6	12.8
6	1.25	0.04	16.6	0.1	82.0	17.9	16.8
7	2.50	0.03	20.4	0.0	98.2	1.8	11.7
Scenario 5							
Control	1.00	0.1	0.0	0.0	0.0	0.0	27.2
1	0.50	0.04	0.0	0.0	25.2	74.8	25.2
2	0.83	0.03	0.0	0.0	73.3	26.7	19.2
3	2.50	0.04	0.0	0.0	99.9	0.1	10.2
4	0.42	0.02	8.9	0.1	14.0	85.9	23.9
5	1.25	0.06	11.0	0.4	84.5	15.1	15.9
6	0.56	0.05	14.8	0.0	22.1	77.9	25.7
7	0.45	0.04	18.7	0.1	11.6	88.3	25.5