Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

BBECT: BANDIT BASED ETHICAL CLINICAL TRIALS

Mohammed Shahid Abdulla

Information Systems Area
Indian Institute of Management Kozhikode
Kozhikode 673570, INDIA

L Ramprasath

Finance and Accounting Area
Indian Institute of Management Kozhikode
Kozhikode 673570, INDIA

ABSTRACT

The aim of an Ethico-Optimal clinical trial is to randomly allocate the new drug (ND) and the standard of care (SOC) to patients in the sample, but with a greater fraction being administered ND if doing so is statistically justified. Such an adaptation is not possible in static trials, in which approximately half the patients would receive ND and the remaining patients SOC, despite evidence within the trial that ND is efficacious. We adapt a canonical stochastic multi-armed bandit algorithm named UCB1 to a clinical trials setting and analyse the resulting type-2 error (β), as also the minimum sample size required by such a trial for a certain β level. We also present simulations to establish that the ethical properties of such a trial are higher, both to verify our analysis and demonstrate an empirical advantage when compared to an existing method.

1 INTRODUCTION

Consider the stochastic multi-armed bandit (SMAB) problem, with K arms or levers, each producing an outcome with mean reward $\{\mu_k\}_{k=1}^K$, arranged such that $\mu_k > \mu_{k+1}$. After pulling each of these K arms in the first K rounds, the single player is permitted to pull any one of these K arms in each succeeding round $K+1 \leq t \leq N$, with each pull yielding a reward $X_t \in [0,1]$. Note here that the random variable X_t belongs to probabilty distribution function $\mathscr{F}_{a_t}$ with support $[0,1]$, where $a_t \in \{1,2,...,K\}$ is the arm or action chosen to be pulled at t. Also note that $E(X_t) = \mu_{a_t}$, with X_t being used to update empirical means $\{\bar{X}_t^k\}_{k=1}^K$ that the player maintains for each arm a^k. If $a_t = a^k$, then empirical mean $\bar{X}_t^k$ is updated as follows: $s_t^k := s_{t-1}^k + 1$, followed by $\bar{X}_t^k := \frac{s_{t-1}^k \bar{X}_{t-1}^k + X_t}{s_t^k}$, where s_t^k is the count of pulls of arm k till (and including) t. The information $\{s_t^k\}_{k=1}^K$ is also retained by the player as she goes to round $t+1$. A common and efficient algorithm for the SMAB algorithm is UCB1 (Upper Confidence Bound variant 1) (Auer et al. 2002) which infers a_{t+1} as follows:

$$a_{t+1} \quad = \quad \arg\max_{1\leq k\leq K} \left\{ \bar{X}_t^k + \sqrt{\frac{2\log(t)}{s_t^k}} \right\}.$$

UCB1 is a *logarithmic* regret SMAB algorithm i.e., it has been established that,

$$E(s_N^k) \quad \leq \quad \frac{8\log(N)}{\Delta^2} + 1 + \frac{\pi^2}{3}, \text{ for } k > 1, \text{ for all } N > \tfrac{8\log(N)}{\Delta^2}, \qquad (1)$$

where $\Delta = \mu_1 - \mu_2$. Such an assurance indicates that only $O(\log T)$ of T opportunities to pull arms were lost to the sub-optimal choice a^k, $k > 1$.

In a simple 2−arm binary-response Phase 3 clinical trial, a sample size N of patients with a particular condition is decided using certain measurements made in the preceding Phase 2 trials. A key input into deciding N is the statistical significance required in the Phase 3 trial's conclusion, most notably the clinical trial's Type-1 error α. This error is the probability of recommending ND when said drug's performance is not statistically different from current standard of care (SOC) or *placebo*.

Also key to deciding N is the trial's Type-2 error β (alternatively, the power of the trial $1-\beta$) which is the probability of recommending SOC when ND is better than SOC. In a randomized controlled trial, which we call a static clincal trial also, roughly $\frac{N}{2}$ patients are administered ND, while the other half are administered SOC. This is done in such a way that patients do not know and cannot reliably infer which of the 2 drugs, a^1 or a^2, they have received.

It is considered ethical within such a trial to administer more number of patients with the new drug if there is statistical evidence *till that point* in the trial of better outcomes. Such ethico-optimal trials have been investigated in (Biswas and Bhattacharya 2011) and a review of such designs can be found in (Villar et al. 2015). If we assume $K=2$ in the earlier description of UCB1, then being able to observe the outcome X_t of administering drug a_t to the $t-$th patient helps decide a_{t+1}. Further, the new drug likely has the greater mean outcome $\mu_1 > \mu_2$, as observed by the investigators in a Phase 2 trial, and hence from UCB1's analysis $s_N^2 = O(\log(N))$. The size of the trial can be set to have N subjects, such that

$$N \quad = \quad \min_{N' \in \mathscr{Z}} \; N' \geq 2 \times \left\lceil \frac{8\log(N')}{\Delta^2} \right\rceil . \qquad (2)$$

Then, the number of patients administered SOC would be $O(\log(N))$, a quantity with promise of being lower than the $\frac{N}{2}$ in a static clinical trial.

Our aim is to employ a variant of UCB1, UCB1-MPA, which is described below, as the basic unit of generic bandit-based clinical trial algorithms. This is referred to as BBECT throughout this article. From (Bubeck et al. 2011), the bandit-based recommendation algorithm UCB1-MPA (Most Played Arm):

$$a_{t+1} \quad = \quad \arg\max_{k \in \{1,2\}} \left\{ \bar{X}_t^k + \sqrt{\frac{\alpha \log(t)}{s_t^k}} \right\} .$$

The method to *allocate* the $(t+1)-$th arm is similar to UCB1, except for the constant α in place of 2. However, at the end of N pulls, UCB1-MPA also *recommends* as the best therapy the arm $a* = \arg_{k=1,2}\max\{s_N^k\}$.

UCB1-MPA modifies UCB1's regret expression to derive an upper-bound on the probability that a^2 would be recommended in place of a^1. Such a probability would thus be the *Type-2 error* β of the SMAB-based clinical trial: the probability that SOC will be recommended despite a significant difference between ND and SOC. Related to α, we come up with a further modification to UCB1-MPA such that inference in favour of a^1 is drawn only if it is used for a fraction larger than $\frac{1}{2}$ of the total pulls. We report the appropriate fraction using simulations for $\alpha = 5\%$, such that 95% of the simulations have insufficient information to reject the null hypothesis.

Note, however, that N in (2) above is typically much larger than the N recommended by Z-statistic based methods used in static clinical trials. It is clarified here that $\Delta = \mu_1 - \mu_2$ is an approximate quantity known to the clinical trial investigators on account of Phase-2 findings, that precede Phase-3 for which BBECT is being proposed. An example here is that for $\Delta = 0.2$, the lowest possible N if using native UCB1 is 3233. The N calculated from formula in (Sullivan), as required for static binary response clinical trials, is s.t. $N = 326$.

After a literature survey, in Section 2 below we describe BBECT using UCB1-MPA with a proof that obtains an N much lower than in the proof of UCB1-MPA itself. Further, in Section 3, we run a series of experiments with BBECT based on UCB1-MPA to compare with both static clinical trials and ethico-optimal clinical trials.

1.1 Literature Survey

As the central module in BBECT, there also are other bandit algorithms apart from UCB1 that have various advantages:

- relaxation of $X_t \in [0,1]$ can take place, such algorithms are called 'heavy-tailed bandit' algorithms, and
- regret quantity s_N^2 can be lower due to different scaling constants and offsets, but are still $O(\log(N))$.

We have, however, chosen UCB1's variant UCB1-MPA as the module for proposed BBECT due to UCB1 being a canonical and easy-to-analyse SMAB algorithm. Indeed, in recent work such as (Williamson and Villar 2020) *forward-looking* multi-armed bandit algorithms have been drafted for clinical trials. Such algorithms calculate indices at each step of the trial, also involving information such as number of enrolled trial participants left, to decide the allocation of patients. However, the work there deals with continuous and normally-distributed outcomes, assumes a prior distribution for parameters of both arms in the trial, and also has a worse tradeoff of trial's power vis-a-vis ethical outcome.

When using BBECT with UCB1-MPA, the sample size N can be calculated in advance based on β required (in that sense it is not myopic as defined by (Williamson and Villar 2020)), and does not require any priors other than an estimate of Δ. Note also the theoretical formulation of regret in the Gittins Index method (Lattimore 2016, (17)) where regret in N steps is greater than $\frac{1600}{\Delta} \cdot \log N$, an unfavourable scale coefficient compared to (1) above. The substitution for this is as follows: $\Delta + \frac{128}{\Delta} \log(2N^2) + 21 \cdot \Delta \lceil \frac{32}{\Delta^2} \log(4N^2) \rceil + 10N \cdot \frac{c''(\log N + \log_+(N\Delta^2))}{N\Delta}$, where we neglect the last term in our calculation. We thus get $\frac{(256+21 \cdot 64+20c'')\log N}{\Delta}$ and hence the assumption of $\frac{1600}{\Delta} \cdot \log N$ earlier. In (Lattimore 2015), an improved UCB1 named 'Optimally-Confident UCB' is presented whose coefficients of regret are difficult to compare directly. However, it is observed there empirically that a Gittins-index -based strategy is the winner in regret terms among a large set of algorithms for small horizons e.g. $N \leq 1000$.

An original investigation using bandit algorithms in binary response trials, that identifies multiple ethical criteria, was performed in (Press 2009). The work in (Press 2009) rules out Gittins- and Whittle-index -based methods due to the infinite horizon formulation (since clinical trials necessarily have a finite horizon). Yet, Gittins-index based methods were established for clinical trials in later investigations such as (Villar et al. 2015), (Smith and Villar 2018) and (Williamson and Villar 2020). A statistic t for the clinical trial is then compared with a t_{crit} as each subject is presented, in the heuristic-based method proposed in (Press 2009), and an allocation to either SOC or ND is decided. While multiple ethico-optimality critieria are considered, incl. cost of treatment and number of treatment failures, the power of the designed trial is not considered, neither is this captured in simulations. The simulations in (Press 2009) capture the results for minimization of 'expected successes lost' criteria, i.e. $(1 - P(\text{allocation to ND})) \times (\mu_1 - \mu_2)$, which is as low as 1.75 for $N = 100$. Note however that the paper does not mention the μ_1, μ_2 used in the simulation.

Even within Management research literature, a shift from quantitative experiments to bandit-based adaptive experiments is being proposed in (Kaibel and Biemann 2021). Their work advises that an ethical outcome in sequential field experiments within organizational behaviour would be allocating more subjects to the more effective intervention. The authors claim that this would align investigations to the goals of researcher syndicates such as Academy of Management (AOM) or the American Psychological Association (APA). Their work, however, does not propose deriving a sample size for any particular bandit approach (though it employs the Thompson Sampling bandit in its simulations). Note that a bandit approach based on Thompson Sampling was tested via simulation in (Villar et al. 2015) and was found to have lesser treatment successes than Gittins- and Whittle- index-based methods. Further, as an illustration, a large N in the simulated trial in (Kaibel and Biemann 2021) results in an Efficient Allocation Proportion - i.e. the fraction of subjects randomized to the better therapy - of 83%. Our algorithm BBECT with UCB1-MPA achieves 80% for even the $1-$st percentile of outcomes from $10,000$ simulations.

The work adapting Bayesian principles to the Bernoulli MABP (BB-MABP, (Villar et al. 2015)) assumes an a-priori distribution for each arm, parametrized by (s_0^k, f_0^k), $k \in \{1,2\}$, where s_0^k represents successes in Phase-2 for arm k. The Gittins (infinite horizon) and Whittle (finite horizon) indices are then used to guide the sampling of the arms during the clinical trial. The authors mention that their proposals, those based on indices, observe both better $p*$ and expected number of successes (ENS), i.e. higher proportion of patients allocated to ND, and a higher level of favourable outcomes for entire cohort. Both $p*$ and ENS are obtained from simulation for a number of patients N that is fixed beforehand as being appropriate for a randomized trial. However, all 3 index-based methods that the authors propose have low power $1-\beta$, around 3.5 times lower than the designed power for the randomized trial. The hybrid controlled Gittins index method proposed in that paper is demonstrated via simulations to have a higher statistical power than the 3 index-based methods proposed earlier. The proposed BBECT may be considered as a method to both design a trial for assured $1-\beta$ values (theoretically), as well as obtain high $p*$ and ENS metrics as demonstrated via simulations.

Note the important difference between BBECT and BB-MABP that inference in the latter is made using Z-statistic, whilst in the former it is made using Most Played Arm (MPA) or a variation therein. Such a variation would be relevant to decide the α of the BBECT trial, e.g. an inference using 'arm played at least 60% of the time'. The scheme of obtaining a cutoff such as 60% for the inference mirrors the different C_α obtained via simulation for different methods in (Villar et al. 2015).

The BBECT with UCB1-MPA that we propose also has the advantage of being a compact scheme, requiring no table look-ups (even different tables at different t) unlike BB-MABP. Also an advantage with BBECT is that a design with target $1-\beta$ may have a larger N_{BBECT} than the N_{FR} in the FR scheme, but will also have better $p*$ and ENS. The latter advantage holds for both N_{BBECT} - which is a theoretical prediction due to logarithmic regret - but also for the much lower threshold N_{FR}.

A valid criticism by (Williamson and Villar 2020) and (Villar et al. 2015) is that most proposed methods of allocation are 'myopic' and do not consider horizon till N. While the Gittins-index based method, the finite-horizon MDP formulated for the Whittle-index based method does consider the remaining opportunities to sample till N is hit, neither the allocation rule nor the inference rule in BBECT has explicit consideration of N. Note that the N is itself designed based on desired type-2 error tolerance $1-\beta$. Contrary to the authors claim in the abstract of (Villar et al. 2015), no index-based allocation rule improves power beyond the situation of being 3.5 times lower than FR's power. The proposed algorithm BBECT may be criticised for obtaining a N_{BBECT} that is much higher than N_{FR} for certain Δ, thus the resulting trial is overpowered in comparison. Yet, index-based rules produce trials that are under-powered with N_{FR} when observed in simulation and have equivalent $p*$ and ENS properties.

2 ANALYSIS OF BBECT FOR STATISTICAL POWER

An outline of UCB's proof of logarithmic regret (Theorem 1 in (Auer et al. 2002)) would serve as a useful illustration. The proof there requires that at least one of the following 3 events occur at an index $t < N$:

$$\begin{aligned} \bar{X}_t^1 &\leq \mu_1 - c_t^1 \\ \bar{X}_t^2 &\geq \mu_2 + c_t^2 \\ \mu_2 + 2c_t^2 &\geq \mu_1. \end{aligned} \tag{3}$$

The probability of these 3 events forms the upper bound for the probability of the event $\{\bar{X}_t^1 + c_t^1 \leq \bar{X}_t^2 + c_t^2\}$, which in turn indicates that at index $t+1$ the suboptimal arm 2 was pulled. Notice that the event in (3) does not occur if $s_t^2 > \frac{8\log(t)}{\Delta^2}$ due to the form of $c_t^2 = \sqrt{\frac{2\log(t)}{s_t^2}}$. It is thus sufficient if $s_t^2 > \frac{8\log(N)}{\Delta^2}$.

Lemma 1 Consider N plays of 2 arms, subject to the condition that $s_N^2 > L_0$, where $L_0 \overset{\Delta}{=} \frac{8\log(N)}{\Delta^2}$. Then, the number of times arm 2 is pulled, s_N^2, is s.t. $E(s_N^2) \leq L_0 + 2\zeta(4)$, where ζ is the Reimannian zeta function.

Proof. Conditioning on $s_N^2 > L_0$, we have that

$$\begin{aligned}
s_N^2 &\le L_0 + \sum_{t>L_0+1}^{N} \sum_{s_t^1=1, s_t^2=t-s_t^1}^{s_t^1=t-(L_0+1)} I\{\bar{X}_t^1 + c_t^1 \le \bar{X}_t^2 + c_t^2 | s_N^2 > L_0\} \\
E(s_N^2) &\le L_0 \\
&+ \sum_{t>L_0+1}^{N} \sum_{s_t^1=1, s_t^2=t-s_t^1}^{s_t^1=t-(L_0+1)} (P\{\bar{X}_t^1 \le \mu_1 - c_t^1\} + P\{\bar{X}_t^2 \ge \mu_2 + c_t^2 | s_N^2 > L_0\}) \cdot P(s_t^1, s_t^2 | s_N^2 > L_0) \\
E(s_N^2) &\le L_0 + \sum_{t>L_0+1}^{N} \sum_{s_t^1=1, s_t^2=t-s_t^1}^{s_t^1=t-(L_0+1)} 2t^{-4} \cdot P(s_t^1, s_t^2 | s_N^2 > L_0) \\
E(s_N^2) &\le L_0 + \sum_{t>L_0+1}^{N} 2t^{-4} \sum_{s_t^1=1, s_t^2=t-s_t^1}^{s_t^1=t-(L_0+1)} \cdot P(s_t^1, s_t^2 | s_N^2 > L_0) \\
&\le L_0 + 2\zeta(4)
\end{aligned}$$

Note the application of the Hoeffding concentration inequality, s.t. $P(\bar{X}_t^1 \le \mu_1 - c_t^1) \le e^{-2s_t^1(c_t^1)^2} = t^{-4}$. The same holds for $P(\bar{X}_t^2 \ge \mu_2 + c_t^2 | s_N^2 > L_0)$ above since the conditioning has no effect. □

The result on $E(s_N^2)$ also holds without conditioning on $s_N^2 > L_0$, since in that case $s_N^2 < L_0$ and therefore $s_N^2 < L_0 + 2\zeta(4)$. Note in the proof above that the marginal probability has been used, where $P(s_t^1, s_t^2 | s_N^2 > L_0)$ is the probability of the pair $(s_t^1, s_t^2) \in \mathcal{N}_+^2$, $s_t^2 > L_0$ occurring as counts of the arms 1 and 2. This change makes the proof different from UCB1's proof (Auer et al. 2002, (6)) where the final bound is $L_0 + 2\zeta(3)$, and allows us to design a lower L_0 in Theorem 1 below. This in turn results in a lower trial size $N \ge 2L_0$ such that adverse probability of outcome is bounded by a small β.

Theorem 1 For each $\beta \in (0,1)$, $\exists L_0$ and N such that $N \ge 2L_0$, $L_0 = \frac{2(r_0+1)^2 \log(N)}{\Delta^2}$, where $r_0 < 1$, s.t. $\sum_{t=L_0+1}^{N} P\{\bar{X}_t^1 + c_t^1 \le \bar{X}_t^2 + c_t^2\} < 1 - \beta$

Proof. We rewrite the 3 events indicated in (3) and earlier, as follows:

$$\begin{aligned}
\bar{X}_t^1 &\le \mu_1 + x_t c_t^1 - c_t^1 &(4)\\
\bar{X}_t^2 &\ge \mu_2 + r_t c_t^2 &(5)\\
\mu_2 + (r_t+1)c_t^2 &\ge \mu_1 + x_t c_t^1. &(6)
\end{aligned}$$

Just as the event (3) above, event (6) requires to be ruled out for s_t^2 exceeding a certain threshold. Thus we require that x_t, r_t and s_t^2 be such that the following holds:

$$\begin{aligned}
\mu_1 - \mu_2 &\ge -x_t c_t^1 + (r_t+1)c_t^2 \text{ where, we solve} \\
\Delta &= -x_t \sqrt{\frac{2\log t}{S_t^1}} + (r_t+1)\sqrt{\frac{2\log t}{S^2}} \text{ as a sufficient condition.}
\end{aligned}$$

$$\begin{aligned}
\text{In the above, } S^2 &= \frac{2(r_0+1)^2 \log N}{\Delta^2} \text{ where we assume } s_t^2 > S^2. \\
\text{Similarly, } S_t^1 &= t - \frac{2(r_0+1)^2 \log N}{\Delta^2} \text{ with } s_t^1 < S_t^1 \text{, next choose a low enough } r_0 \text{ s.t.} \\
\frac{2(r_0+1)^2 \log N}{\Delta^2} &\ll \frac{8 \log N}{\Delta^2}.
\end{aligned}$$

Now notice from (4)-(5) that we can apply conditions similar to Lemma 1. Thus, we set the constraint that $1 - x_t = r_t$, and obtain the solution:

$$x_t = \frac{-1 + \frac{2}{r_0+1}\sqrt{\frac{\log(t)}{\log(N)}}}{\sqrt{\frac{2\log(t)}{t\Delta^2 - 2(1+r_0)^2\log(N)}} + \frac{1}{r_0+1}\sqrt{\frac{\log(t)}{\log(N)}}}$$

Further $L_0 = \frac{2(r_0+1)^2\log N}{\Delta^2}$. Use x_t, r_t to obtain N such that $\sum_{t=L_0+1}^{N} 2t^{-4r_t^2} < 1-\beta$, where β is the desired Type-2 error of the statistical test. □

To illustrate a calculation, we obtain $N = 668$ with $L_0 = 334$ (corresponding to $r_0 = 0.013$) when we input $1-\beta = 0.9$ and $\Delta = 0.2$. Similarly, we obtain $N = 588$ (corresponding to $r_0 = -0.04$) when we input a different power $1-\beta = 0.8$ for same Δ. We have assumed that the process $\{\bar{X}_t^i\}$ is an empirical mean of s_t^i events of type i with Bernoulli parameter p_i, where $i \in \{0,1\}$.

For the static clinical test, there exists a combination $p_1 = 0.6$, $p_2 = 0.4$, for which the $N = 326$ obtained is much lesser. Yet, we will demonstrate using simulation that treatment failures in BBECT are lower (alternatively, efficient allocation proportion for BBECT is superior). A grid search for r_0 for each possible value of Δ (with $\beta = 0.9$ employed) yields value of N suited to BBECT using UCB1-MPA. These values are compared below in Table 1 for BBECT versus static clinical trials, where the maximum possible value of N over varied p_1, p_2 pairs is recorded, s.t. $p_1 - p_2 = \Delta$. We use the formula for hypothesis testing applicable to settings of dichotomous outcomes and 2 independent samples, as given in (Sullivan).

Table 1: Minimum sample size required for power $\beta = 0.9$

Δ	N for BBECT ($\beta = 0.9$)	N in static trial ($\beta = 0.9$, $1-\alpha = 0.99$)
0.1	3222	1302
0.15	1290	578
0.2	668	326
0.25	400	208
0.3	262	145
0.35	184	107
0.4	134	82
0.45	102	65
0.5	80	53

In practice, BBECT using UCB1-MPA will also be more ethical due to the larger number of patients allocated to ND, i.e. the arm with efficacy p_1. There will, however, be a marginal increase in the duration and cost associated with the clinical trial. It is useful to reiterate what Theorem 1 implies: suppose that the first L_0 subjects are allocated to arm 1, followed by which one subject is allocated to arm 2. Then, the probability of even one more subject from the remaining $N-(L_0+1)$ subjects being allocated to arm 1 is less than β if BBECT with UCB1-MPA is used.

A note about calculation of N for static trial, (Sullivan), is also required here for the sake of completeness. The minimum sample size N is calculated based on a basic quantity named 'effect size' $E(p_1, p_2)$:

$$\begin{aligned} E(p_1,p_2) &= \frac{p_2 - p_1}{\sqrt{p(1-p)}} \text{ where,} \\ p &= \frac{p_1+p_2}{2} \\ N &= \left\lceil 4\left(\frac{z_{1-\alpha}+z_\beta}{E(p_1,p_2)}\right)^2\right\rceil \end{aligned}$$

The maximum N over a fine grid of possible (p_1, p_2), for each Δ, has been calculated and placed in the third column of Table 1 above. The N calculated in (Rosenberger et al. 2001) for the comparisons below (e.g. Tables 4, 5) appear to be higher but the authors do not point there to any formula to infer N.

3 BBECT COMPARED TO STATIC AND ETHICO-OPTIMAL CLINICAL TRIALS

We implemented a simulation with $100,000$ trials where Bernoulli parameters p_1, p_2 were sampled uniformly from $(0,1)$ and $p_1 - p_2 = \Delta$, with Δ set to 0.2. The percentile values for number of patients in each trial, from a total of 668, that were allotted to treatment represented by p_1 were captured in the simulations. These are given in Table 2. This indicates that in 99% of the simulated BBECT runs, 73% or more of the patients were allocated to ND.

Similarly, in 99% of the simulated BBECT runs where the effective Δ is such that $\Delta \in [0.2, 0.3]$, 77% or more of the patients were allocated to ND. This experiment models situations where Δ is not known accurately, but for inferring N it is sufficient to know a d such that $\Delta > d$

Note also how the ethical outcome is achieved: consider the $10-$th percentile mark when $\Delta = 0.2$, implying 90% of simulations have higher allocations to ND. We choose this level since the power of the test as designed according to Theorem 1 above is also pegged at 90%. For this particular level, reading off the table, note that $668 - 529 = 139$ is less than $0.5 \times 326 = 163$, where $N = 326$ is the maximum static trial size N for $\Delta = 0.2$ from Table 1. It may similarly be useful to compare the difference between N and the $10-$th percentile level for each Δ, with $\frac{N}{2}$ of a static trial, to verify the efficacy of BBECT with UCB1-MPA as a technique. This is done in Table 3, where notice the advantage for BBECT using UCB1-MPA for all $\Delta > 0.1$.

Table 2: Percentile threshold allotted to ND under BBECT using UCB1-MPA

Percentile	$p_1 - p_2 = 0.2$, $p_1, p_2 \in (0,1)$	$p_1 - p_2 = \Delta$, s.t. $\Delta \in [0.2, 0.3]$
1	493	518
5	517	542
10	529	553
50	564	587
90	593	612
95	601	618
99	615	629

We next tried a simulation with $1,000,000$ trials to estimate the level of significance α, often called the $p-$value threshold of the test, by turning $\Delta = 0.0$ and modifying the algorithm slightly. We set $N = 668$ using Table 1, but set the criterion that ND would be declared as the better therapy - i.e. null hypothesis rejected - only if patients allotted to it exceed 60%. In a strict sense, the sample size N would change for such a rule (where threshold is 60% and not merely MPA) but we continue use of N recommended by Table 1. It is important to note here that situations where $\Delta = 0$ were handled differently. The p_2 used in

Table 3: Allocation to SOC under BBECT using UCB1-MPA

Δ	$10-$th percentile of allocations to SOC (using BBECT)	$\frac{N}{2}$ from Table 1
0.10	658	651
0.15	267	289
0.20	139	163
0.25	84	104
0.30	55	73
0.35	39	54
0.40	28	41
0.45	22	33
0.50	17	27

the simulation was $p_1 + 0.05$ (with probab. 0.5), or alternatively $p_1 := p_2 - 0.05$ (with probab. 0.5). In none of the experiments was the expected response of either SOC or ND required, neither was information about variability.

However, the BBECT algorithm is distribution-dependent in the sense that information about approximate Δ is still required. Our simulations compare BBECT using UCB1-MPA within the setting of (Rosenberger et al. 2001) which proposes an optimal adaptive rule for binary response trials. Taking $\Delta = p_1 - p_2$, the situations compared are those where significance of trial (when $\Delta = 0$) as well as power of trial ($\Delta > 0$) is observed over 1 million simulations. Notice in Table 4 below that the 'error rate' when $\Delta > 0$ is better than the optimal adaptive rule, signifying more power than the 90% for which the optimal adaptive rule was designed. It is also the case that significance calculated empirically from BBECT simulations when $\Delta = 0$ is such that Type-1 error stays below 5%. Note that the N in these experiments - where power observed is more than 95% - happens to be even lower than N calculated analytically for 90% power in Table 1.

Further, the BBECT algorithm also has lower 'treatment failure', i.e. a lesser number of subjects who did not recover irrespective of which arm they were allocated. Table 5 presents the number of treatment failures with standard deviation (SD). However, the SD metric is observed as being higher in some pairs when BBECT with UCB1-MPA is employed. In addition, even in cases where the SD metric is lower, an advantageous disjoint interval of confidence for the treatment failure metric isn't seen.

Note the extensive simulation in (Smith and Villar 2018) where 'Type-1 error inflation' for bandit-based methods occurs, incl. UCB1. This inflation is similar to the outcome we would obtain if we employed the original MPA rule of declaring arm as winner if more than 50% of pulls correspond to it. Notice also that C_α there has been tuned with simulations to suit the bandit algorithm, just as our mark of 60% is obtained here using simulations. For example, the standard C_α would be 1.645, but is adjusted to 2.068, 1.867, 1.701 for the algorithms UCB1, KL-UCB1 and Thompson Sampling, respectively, cf. (Smith and Villar 2018, Table 1). Notice also the power values of our algorithm in the lower half of Table 4 (all above 95%) whilst power values in (Smith and Villar 2018, Table 1), even for the bandit methods, are less than 90%. Of these methods, KL-UCB has the best balance of statistical test's power and efficient allocation proportion (77% and 82%, respectively) both of which are unfavourable compared to our figures.

We also present the comparison with work in (Villar et al. 2015) where Bayesian bandit clinical trial algorithms are introduced for the Bernoulli case, like ours. There are 3 algorithms introduced there, using Gittins Index (GI), Whittle Index (WI), and a Randomized Gittins index (RGI), all having the advantage of being 'non-myopic', i.e. sensitive to the horizon left for the trial. The work compares p_2 of 0.3 and p_1 of 0.5 after deriving $N = 148$ for the static fixed allocation clinical trial. The BBECT figure for expected number of successes (ENS) was 65.95 at this N, compared to the approximately 70 (ratio $\frac{70}{148} = 0.473$) that GI and WI were able to achieve. However, note that statistical power of the GI, WI methods was very low, at $0.3 - 0.4$ compared to values greater than 90% for BBECT. The work in (Villar et al. 2015) has however demonstrated higher statistical power empirically for a variant of one these methods, viz.

Controlled Gittins index. If using $N = 588$ in this setting for BBECT, where N is calculated from Theorem 1 for power setting $1-\beta = 0.8$, we have an ENS of 274.21 which at 0.491 exceeds the ratio that GI and WI achieve. Type-1 error evaluated using our method of perturbing p_1 or p_2 by 0.05 yields 0.056 for $N = 148$, whilst it is a low 0.01 for $N = 558$ (well within limit for design criterion of 5%).

Table 4: Error-rate and power in BBECT vs Optimal Adaptive rule

p_2	p_1	N	BBECT	Optimal Adaptive rule
0.10	0.10	200	0.00	0.04
0.30	0.30	200	0.04	0.05
0.50	0.50	200	0.05	0.04
0.70	0.70	200	0.04	0.04
0.90	0.90	200	0.00	0.04
0.10	0.20	526	0.96	0.89
0.10	0.30	162	0.98	0.89
0.10	0.40	82	0.99	0.89
0.40	0.60	254	0.99	0.89
0.60	0.90	82	0.98	0.90
0.70	0.90	162	0.98	0.91
0.80	0.90	526	0.96	0.90

Table 5: Treatment failures in BBECT vs. Optimal Adaptive rule

p_2	p_1	N	BBECT	Optimal Adaptive rule
0.10	0.20	526	436.4 (9.6)	443 (8.5)
0.10	0.30	162	122.0 (6.2)	126.2 (5.4)
0.10	0.40	82	55.2 (4.8)	58.5 (4.2)
0.40	0.60	254	113.3 (8.6)	124.4 (7.8)
0.60	0.90	82	14.1 (3.0)	19.3 (3.7)
0.70	0.90	162	24.7 (4.2)	31.5 (4.8)
0.80	0.90	526	68.1 (7.4)	78.3 (8.1)

4 FUTURE DIRECTIONS

The N proposed by Theorem 1 is higher when compared to a static clinical trial, and this requires registering more volunteers which is a challenge if the condition is rare. However, the Hoeffding bound used above produces an upper limit and alternative bounds exist whereby if $p_2 - p_1 \geq \Delta$ and $p_2 - p_1 \leq d$ are both known, then the bound is tighter. A closed form expression for the Type-1 error, even under the assumption of 'region of indifference' viz. a minor difference between p_1 and p_2, would also be welcome for practitioners.

ACKNOWLEDGMENTS

We thank Shivaram Kalyanakrishnan of IIT Bombay for reading an early draft of this paper and giving valuable inputs. We also thank an anonymous referee for pointing out several typos in the text, and also in helping to improve the explanation at many points.

REFERENCES

Auer, P., N. Cesa-Bianchi, and P. Fischer. 2002. "Finite-time Analysis of the Multiarmed Bandit Problem". *Machine Learning* (47):235–256.

Biswas, A., and R. Bhattacharya. 2011. "Optimal response-adaptive allocation designs in Phase III clinical trials: Incorporating ethics in optimality". *Statistics and Probability Letters* (81(8)):1155–1160.

Bubeck, S., R. Munos, and G. Stoltz. 2011. "Pure exploration in finitely-armed and continuous-armed bandits". *Theoretical Computer Science* (412(19)):1832–1852.

Kaibel, C., and T. Biemann. 2021. "Rethinking the Gold Standard With Multi-armed Bandits: Machine Learning Allocation Algorithms for Experiments". *Organizational Research Methods* (24(1)):78–103.

Lattimore, T. 2015. "Optimally Confident UCB: Improved Regret for Finite-Armed Bandits". *arXiv preprint* (arXiv:1507.07880).

Lattimore, T. 2016. "Regret Analysis of the Finite-Horizon Gittins Index Strategy forMulti-Armed Bandits". *Annual Conference on Learning Theory* (49):1–32.

Press, W. H. 2009. "Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research". *Proceedings of the National Academy of Sciences* (106(52)):22387–22392.

Rosenberger, W. F., N. Stallard, A. Ivanova, C. N. Harper, and M. L. Ricks. 2001. "Optimal Adaptive Designs for Binary Response Trials". *Biometrics* (57):909–913.

Smith, A. L., and S. S. Villar. 2018. "Bayesian adaptive bandit-based designs using the Gittins index for multi-armed trials with normally distributed endpoints". *Journal of Applied Statistics* (45(6)):1052–1076.

Lisa Sullivan. "Power and Sample Size Determination". https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Power/BS704_Power_print.html.

Villar, S. S., J. Bowden, and J. Wason. 2015. "Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges". *Statistical Science* (30(2)):199–215.

Williamson, S. F., and S. S. Villar. 2020. "A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes". *Biometrics: Journal of the International Biometric Society* (76(1)):197–209.

AUTHOR BIOGRAPHIES

MOHAMMED SHAHID ABDULLA is an associate professor in the Information Systems Area at Indian Institute of Management, Kozhikode, Kerala, INDIA. He holds a PhD in Computer Science and Automation from Indian Institute of Science. His research interests are in machine learning, simulation, and optimization, with some applications in the area of finance. His email address is shahid@iimk.ac.in.

L RAMPRASATH is an associate professor in the Finance Area at Indian Institute of Management,Kozhikode, Kerala, INDIA. He holds a PhD in Statistics from Rutgers University. His research interests are in financial derivatives and methods of simulation. His email address is lrprasath@iimk.ac.in.