Real-time Adaptive Randomization of Clinical Trials

Supplementary Online Content

Online Appendix

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix A. Notation and simulation procedures

eA.1. Notation and technical details

eA.2.1. Optimality (step IV.4.1 described below)

RTARs balance earning, providing the (likely) best treatment on day d, with learning, trying (currently) inferior arms to learn about endpoint rates so that better decisions can be made about all future patients, including those after the trial. For many reasons, positive endpoints and knowledge is better now than in the future, so we slightly discount future endpoints by a factor, δ, where $\delta \in (0, 1)$.

We represent knowledge about each arm by a (Bayesian) posterior distribution, a Beta distribution, with parameters α_{ad} and β_{ad}, for arm a based on endpoints observed up to day d. (For ease of exposition, we switch notation from the text ($\alpha_a(d)$) to this eAppendix (α_{ad}).) Gittins [e4] proved that the optimal balance between earning and learning is to compute a "Gittins index" for each arm and assign patients to the arm with the lowest index. (We are minimizing mortality, so lower is better.) We compute the Gittins index by comparing the "rewards" from a (currently) uncertain arm to the rewards from assigning patients to an arm where the endpoint rate is the Gittins index, $G(\alpha_{ad}, \beta_{ad})$. For the trials we analyze, "rewards" is mortality (GUSTO-1 and EUROPA) or another negative event (EUROPA). Fewer mortalities/negative-events are better. We write this value as G_{ad} for short. G_{ad} represents the comparative daily rewards for an arm in which the anticipated endpoint is known with certainty.

Let $R(\alpha_{ad}, \beta_{ad})$ be the expected discounted rewards for acting optimally on day d and all future days. We compare an uncertain arm to a certain arm to compute the Gittins index. In this comparison, the rewards for day d are related to G_{ad} and the rewards for day $d + 1$ by the Bellman equation. See derivation in [e2,e3].

$$R(\alpha_{ad}, \beta_{ad}) = min\left\{\frac{G_{ad}}{1-\delta}, \frac{\alpha_{ad}}{\alpha_{ad} + \beta_{ad}}[1 + aR(\alpha_{ad} + 1, \beta_{ad})] + \frac{\beta_{ad}}{\alpha_{ad} + \beta_{ad}}\delta R(\alpha_{ad}, \beta_{ad} + 1)\right\}$$

There is no analytical solution to this Bellman equation, but the $G(\alpha_{ad}, \beta_{ad})$'s are easy to compute numerically. We do the numerical calculations and store a table of Gittins indices by α and β.

The Gittins solution is provably optimal if one patient is assigned on day d and if the endpoint for patients assigned on day d is observed before assignments on day $d + 1$. The Gittins solution is an heuristic algorithm, which we hope will reduce mortality relative to an RCT or the block-based FLGI, when more than one patient is assigned on day d and there are delays in observed endpoints. This is an empirical question. The main paper tests whether the approximate optimal solution provides benefits relative to an RCT, the previously-proposed block-based FLGI solution, and an η-variant.

Future research might improve assignments with the solution to a fully optimal Bellman equation that accounts for multiple patients on day d and delays in endpoints. Thus, all results for the (hopefully) approximately-optimal RTAR in this paper are conservative relative to the solution to such a Bellman equation. Note that one way of handling delays is to change the discount rate, δ, to reflect a 30-day lag in updating. Fortunately, for our data, the Gittins solution appears to be robust to changes in the discount rate suggesting that the loss of optimality due to delays may not be severe. Another heuristic might be to modify the RTAR to use the FLGI within each day d for which multiple patients are assigned. Such a variant, and many other variants, are readily explored with our resampling simulation code. We did not explore all variations to avoid overfitting the empirical data.

eA.1.2. Learning (step IV.4.3)

The learning step is based on updating the Beta priors, α_{ad} and β_{ad}, with observations of the endpoints at day d. (The Beta priors are updated at the end of day d, patients assignments at the beginning of day d are based on all data up to, but not including, endpoints observed on day d.) Assuming the endpoints are observations from a Bernoulli process with stationary endpoint rates, the updating is simple and quick. When one endpoint is observed per day:

$$\alpha_{a,d+1} = \alpha_{ad} + 1, \beta_{a,d+1} = \beta_{ad} \text{ if the endpoint is a mortality}$$

$$\alpha_{a,d+1} = \alpha_{ad}, \beta_{a,d+1} = \beta_{ad} + 1 \text{ if the endpoint is survival}$$

If more than one endpoint is observed at the end of day d, say n_{md} mortalities and n_{sd} survivals, then we update using n_{md} and n_{sd}.

eA.2. Endpoints

In GUSTO-1, the endpoint is death or survival at 30-days since randomization. In EUROPA, the endpoint is a composite of cardiovascular mortality, non-fatal MI, and resuscitated cardiac arrest at any point in the trial.

eA.3. Simulation procedures

The RTAR multi-arm bandit (MAB) simulation can be run for any number of replicates. The number of replicates is set in the file "Parameters.R." All reported results in the manuscript are results averaged across 200 replicates. The RTAR MAB code is in the file "MAB.R." The η-variant follows the same procedure except that, with probability η, patients are assigned as in an RCT (equally likely) until the arm reaches a pre-defined minimum number of patients. With probability $1 - \eta k_d$, patients are assigned with the RTAR MAB. k_d is the number of arms that have not yet reached the minimum number of patients at the start of day d. The block-based forward-looking Gittins index (FLGI) algorithm is described in [e1]. In the block-based FLGI algorithm, patients are randomized in blocks. The code is available from the authors. We provide here the conceptual steps in the RTAR resampling simulations..

Step I. Load parameters and set seed.

When the parameter file indicates that a single replicate is to be run, the system uses a fixed seed. When more than one replicate is to be run, the system uses different random seeds in each replicate. Results are averaged across replicates. In the case of confidence intervals, we note the values where 2.5% are below (lower) or 2.5% are above (upper) the confidence limits.

Step II. Load support functions.

Step III. Load data.

Step IV. Loop over all replicates (this the main part of the code)

Step IV.1. Build the pools of patients for this replicate.

The original RCTs (GUSTO-1 and EUROPA) assigned one set of patients to each arm (treatment). We refer to each of these sets of patients as a "pool of patients" for that arm. These pools will be used in step IV.4.2, when the RTAR MAB algorithm draws (with replacement) from these pools when making its assignments.

Step IV.2. Load priors for α_a(d) and $\beta_a(d)$ for this replicate.

These priors are set in the file "Parameters.R."

Step IV.3. Initialize intermediate data structures

See the file "Data dictionary.txt" for details on the intermediate variables.

Step IV.4. For each day d *of the trial in this replicate, perform the following:*

Step IV.4.1 DECIDE: select arm a_d^* *to use on day* d*. This is the optimality step.*

Select the optimal treatment arm, a_d^*, given current $\alpha_a(d)$ and $\beta_a(d)$ parameters of the Beta distribution over treatment-arm endpoint rates. The parameters are based on all endpoints observed at the start of day d. The RTAR algorithm chooses the arm with the largest Gittins index, G_{ad}. G_{ad} is a pre-computed tabled function $\alpha_a(d)$ and $\beta_a(d)$. The Gittins index optimally balances, on a daily basis, the amount of learning and earning the system does [e2, e3]. For details on the optimality step, please refer to §e3.2.

Step IV.4.2. RESAMPLING PATIENTS: The RTAR MAB draws patients from the pool of arm a_d^**.*

Compute the number of patients N_d that were randomized by the RCT on day d.

Assign N_d patients to the optimal treatment arm, a_d^*, by drawing with replacement N_d patients from the pool of patients that were randomized by the original RCT to the treatment a_d^*.

Step IV.4.3. LEARN

For each treatment arm, learn from the endpoints observed for all the patients that had been assigned to that treatment arm and for whom endpoints have been observed by the start of day d. Update $\alpha_a(d)$ and $\beta_a(d)$ as described in §e3.2.

Step IV.5. Summarize results of the original RCT.

Step IV.6. Summarize the results of this RTAR replicate.

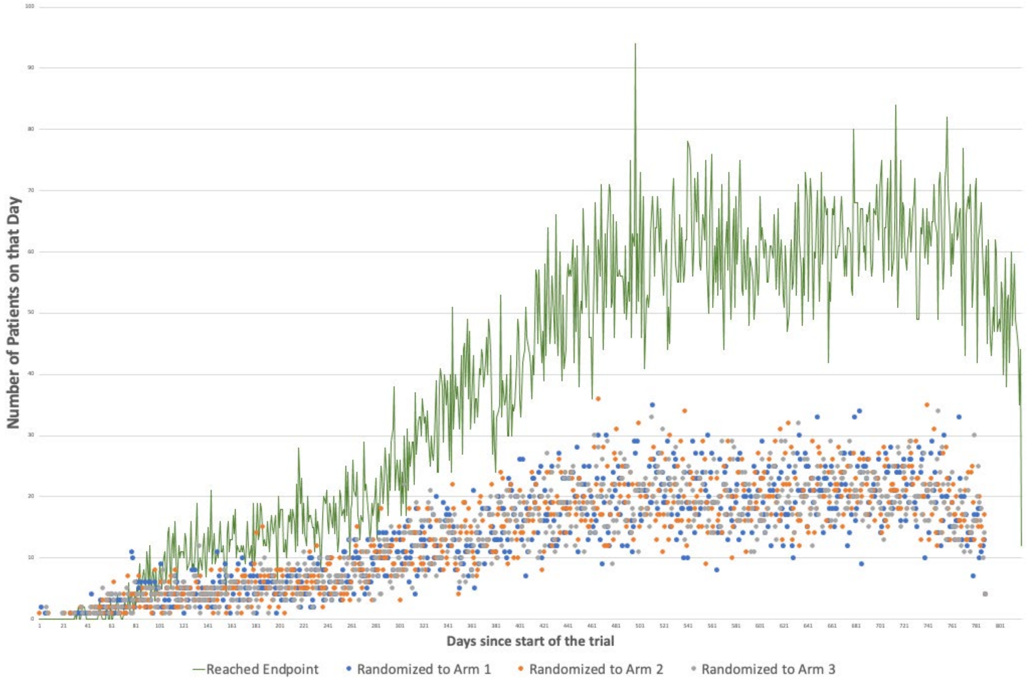
Step IV.7. Save all outputs to .csv files.

eAppendix B. Empirical details on the original RCTs and the RTAR simulations

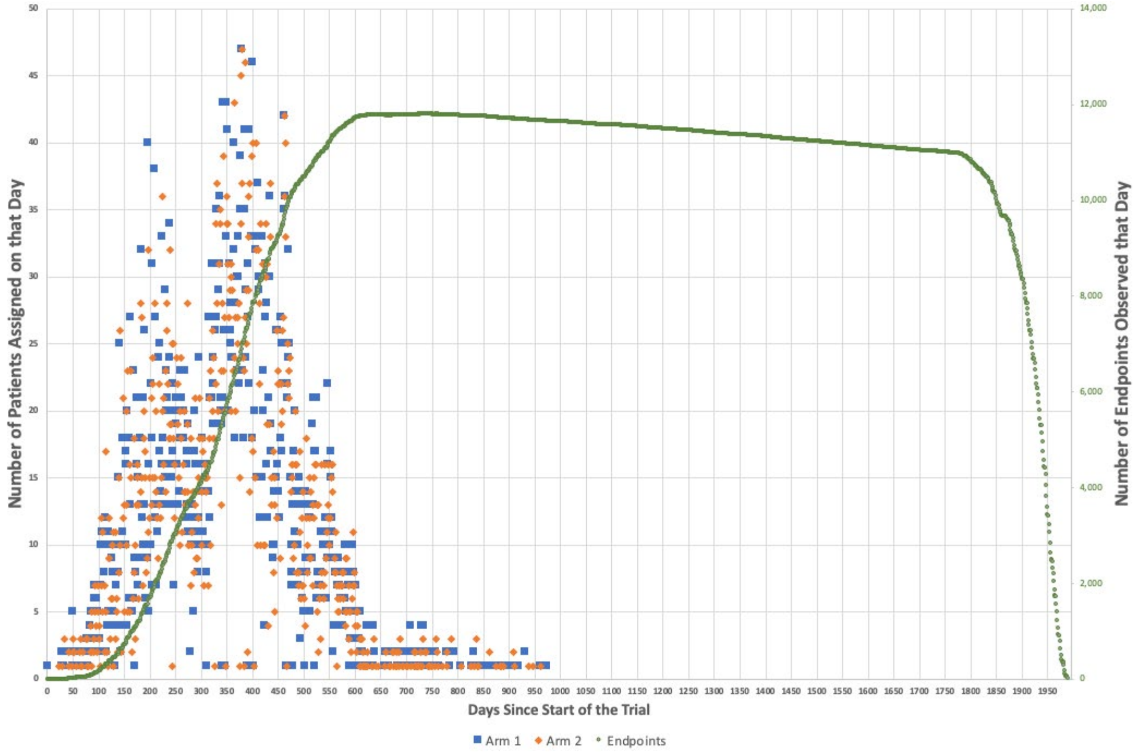
eB.1. Randomization and endpoints for GUSTO-1 and EUROPA RCTs

Figure 1 (for GUSTO-1) and Figure 2 (for EUROPA) present the randomization of patients and the endpoints observed in both studies.

The dots at the bottom of eFigure 1 and to the left of eFigure 2 correspond to the number of patients that were randomized by the RCT in each day of the GUSTO-1 trial (Figure 1) and EUROPA trial (Figure 2). This information is shown separately per treatment, using a color code. For GUSTO-1, blue corresponds to RCT randomizations to t-PA+Heparin (arm 1). Red corresponds to RCT randomizations to SK+Heparin (arm 2). Gray corresponds to RCT randomizations to t-PA+SK+Heparin (arm 3). For EUROPA, blue corresponds to RCT randomizations to Perindopril and orange corresponds to RCT randomizations to placebo. eFigures 1 and 2 also present, in the green solid line on the top of the figures, the number of endpoints that were observed in each day of the trial. This is the total number of daily endpoints summed over all arms in each study (GUSTO-1 had three arms; EUROPA had two arms).



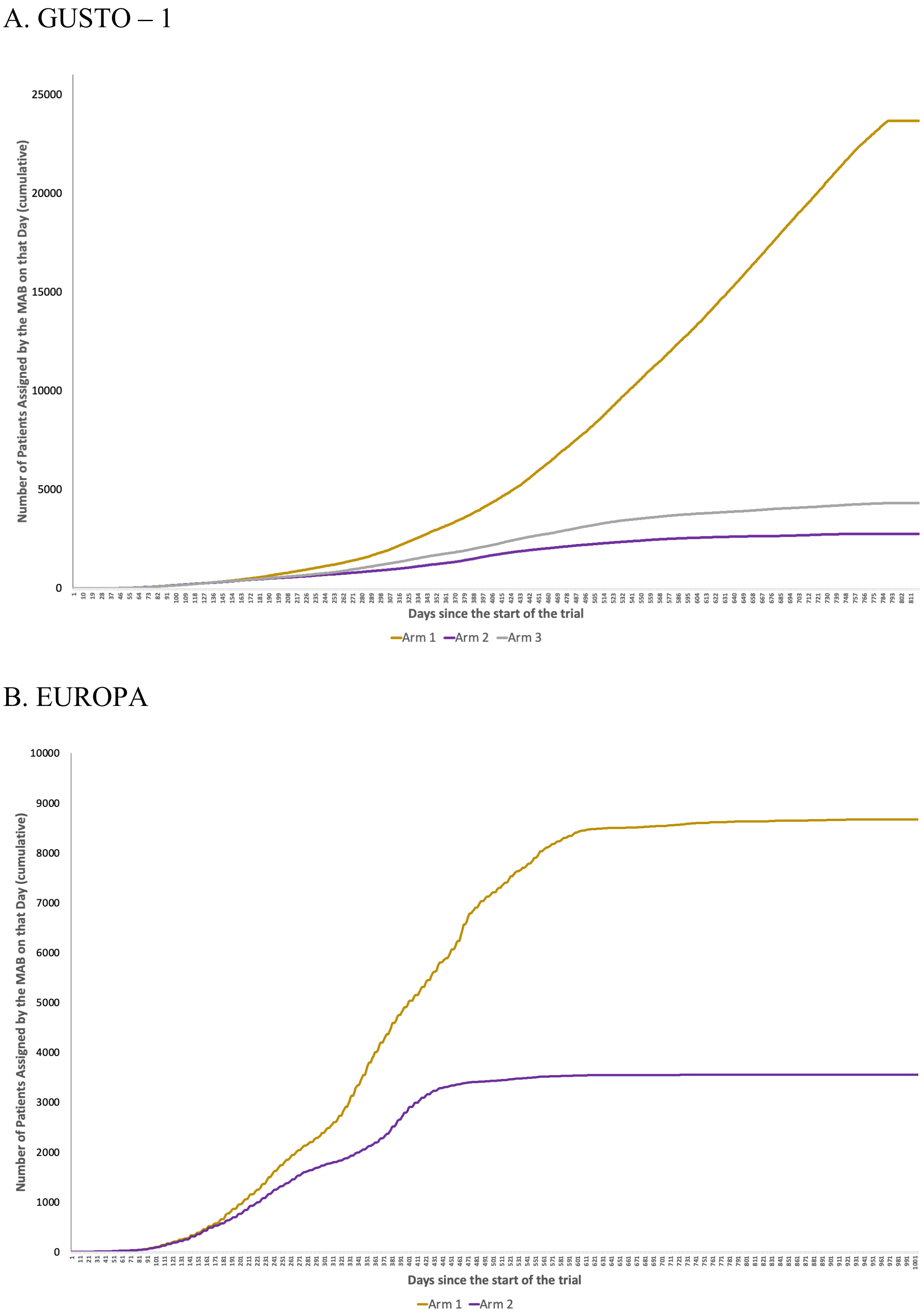
eFig. 1: RCT Randomizations (in blue, orange and gray) and endpoints (in green) in GUSTO-1



eFig. 2: RCT Randomizations (in blue and orange) and follow-up study's endpoints (in green) in EUROPA

eB.2. Evolution of RTAR arm assignments for GUSTO-1 and EUROPA

eFigure 3summarizes arm assignments for (a) GUSTO-1 and (b) EUROPA. The purple, gold, and gray lines (GUSTO-1) or purple and gold lines (EUROPA) and the left vertical axis present the cumulative number of assignments over the duration of the trials. The horizontal axis represents the days of the trial. The RTAR adapts as data on patient endpoints become available. As the trial progresses, the RTAR automatically assigns more patients to the superior arm (gold line). By the roughly the 500th day of the 819-day GUSTO-1 trial, and the 500th day of the 1,989-day EUROPA trial, the RTAR begins to assign almost all patients to the superior arm (gold line). In theory, a particularly adverse, but random, run of negative endpoints might lead an MAB to explore inferior arms after stabilization, but that probability is low. Future research might explore optimal stopping rules which could save even more lives than the RTAR studied in this paper.


A. GUSTO – 1

B. EUROPA

eFig. 3. Assignments to arms using the day-to-day RTAR

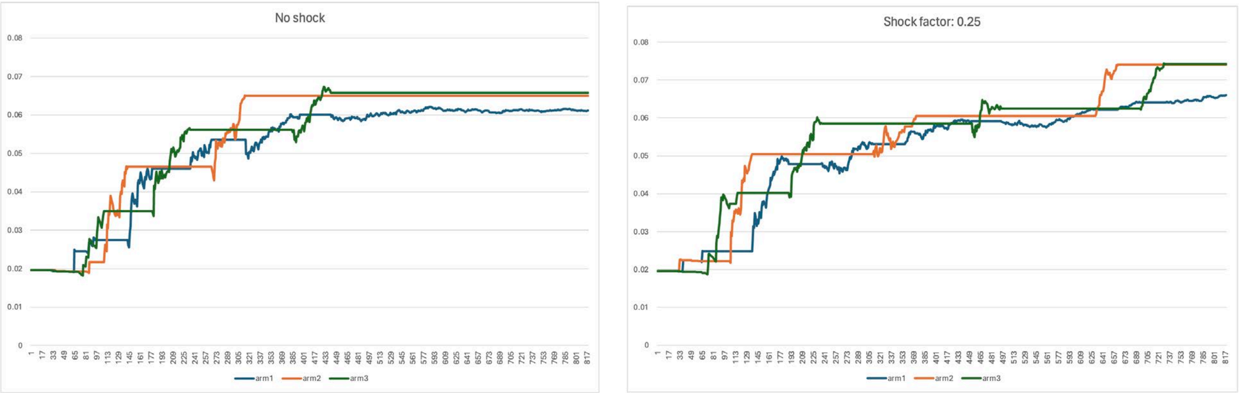
***e*Appendix C.** Temporal Changes in Endpoint Rates

To explore the impact of temporal changes in the endpoint rates on the performance of an RTAR, we use sample enrichment to simulate the effect of a change in endpoint rates midway through the trial. Patient enrichment is an accepted way to model temporal changes [11] and will provide equivalent implications to changing endpoint rates in simulations. With patient enrichment, we add sufficiently many patients to the pool for each arm such that the endpoint rate in each arm is the endpoint rate we seek to simulate.

For example, the RCT endpoint rate (mortality) for arm 1 in GUSTO-1 is 0.0615 and we wish to simulate a shock of 5%. For arm a in the first period, we draw from GUSTO-1 patients for the days corresponding to the first $N_{a,total}/2$ patients in GUSTO-1 arm a. We modify the pool of patients from which we draw patients for the days corresponding to the second $N_{a,total}/2$ patients in the GUSTO-1 trial. A 5% increase in 0.0615 is 0.0646, an increase of 0.0031 in the mortality rate. To maintain consistency with the literature, we increase all arms by 0.0031 resulting in a vector of endpoint rates of [0.0646, 0.0753, 0.0717] for arms 1, 2, and 3, respectively.

eC1. Does an RTAR detect non-stationarity in end-outcome rates (e.g., shocks)?

RTAR assignments are based on the smallest Gittins index at any point in time. eFigure 4 illustrates how the Gittins index changes during the trial for two separate GUSTO-1 simulations. A change in the Gittins index indicates that the arm is being used. This is so because when an arm is used, the Gittins index is updated with the end-outcome. For example, the left pane of eFigure 4 shows that *when there are no shocks*, the RTAR algorithm for this replicate stabilizes slightly before the 500th day of the trial, i.e., the RTAR stops assigning the two worst arms (arms 2 and 3). After the 500th day, the Gittins indices for arms 2 and 3 do not change as represented by the flat green and red lines. Note also that the ranking of the three arms also does not change.



eFig. 4. Gittins indices indicating assignments to arms using the day-to-day RTAR for one replicate of the GUSTO-1 data in the absence (left) and presence (right) of a 25% shock.

Separately, in another replicate, we introduced a 25% shock after 50% of the patients to be assigned (day 530). The results, shown in the right pane of eFigure 4, indicates that the Gittins index driving RTAR assignments had stabilized for this replicate around the 450th day (represented by the flat red and green lines). Around day 560 (when the first 30-day after-shock GUSTO endpoint-outcomes were observed), the index detected the changes in mortality rates and the RTAR returns to using all arms. This is represented by changes in the green, blue and red lines. By day 730, the RTAR algorithm has sufficiently explored and learned about the best arms given the new mortality rates (i.e., post shock). Assignments again stabilize, and RTAR assigns patients only to the arm it has automatically determined is the best arm (arm 1, blue line). The green and red lines for arms 2 and 3 are flat, representing no new patients being assigned to those arms. The data for several shock levels (from 0% to 25%) are provided in a separate spreadsheet file.

Non-stationarity (e.g., drift and shock) is an active area of research for multi-arm bandits [e6, e7, e8, e9]. We are hopeful that researchers will, in the future, develop real-time adaptive algorithms that are optimal in the presence of delays and various forms of non-stationarity.

***e*References** (repeated here for convenience)

[e1] Villar SS, Wason J, Bowden J. Response-adaptive randomization for multi-arm clinical trials using the forward-looking Gittins Index rule. 2015;Biometrics 71:969-978.

[e2] Gittins J, Glazebrook K and Weber R. Multi-armed bandit allocation indices. London: Wiley. 2011.

[e3] Hauser JR, Urban G, Liberali G and Braun M. Website Morphing. Marketing Science 2019;28(2):202-224.

[e4] Gittins JC 1979. Bandit processes and dynamic allocation indices. Journal of the Royal Statistical. Society.1979; (Ser B)41(2):148–177, plus commentary.

[e5] Villar SS, Bowden J and Wason J. Response-adaptive designs for binary responses: how to offer patient benefit while being robust to time trends? 2018;Pharm Stat; 17: 182–19

[e6] Simchi-Levi, David and Wang, Chonghuan and Zheng, Zeyu, Non-stationary Experimental Design under Structured Trends (July 18, 2023). Available at SSRN: https://ssrn.com/abstract=4514568 or http://dx.doi.org/10.2139/ssrn.4514568

[e7] Chen Q, Golrezaei N, Bouneffouf D. Non-stationary bandits with auto-regressive temporal dependency. 37th Conference on Neural Information Processing Systems. 2023:1-22.

[e8] Liu, Y., Kuang, X., & Van Roy, B. (2023). Non-stationary bandit learning via predictive sampling [arXiv:2205.01970]. arXiv preprint arXiv:2205.01970.

[e9] Liu, Y., Kuang, X., & Van Roy, B. (2023). *A definition of non-stationary bandits* [arXiv:2302.12202]. arXiv preprint arXiv:2302.12202.