Novel bandit algorithms for early phase efficacy trials

Sofía S. Villar

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

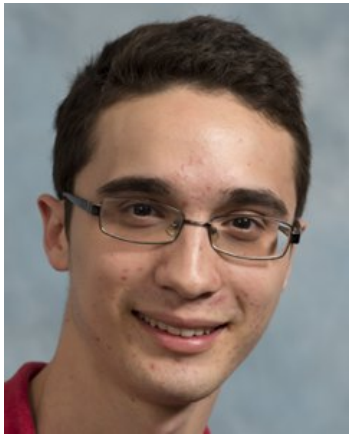
Journees MAS 2022 Rouen

30th August 2022

Acknowledgments

With thanks to

Helen Barnett



David Robertson

Nina Deliu

and many other colleagues

Outline

Learning-Earning in Clinical Trials

Sequential (long) process. Decisions are sequential but learning is compartmentalised

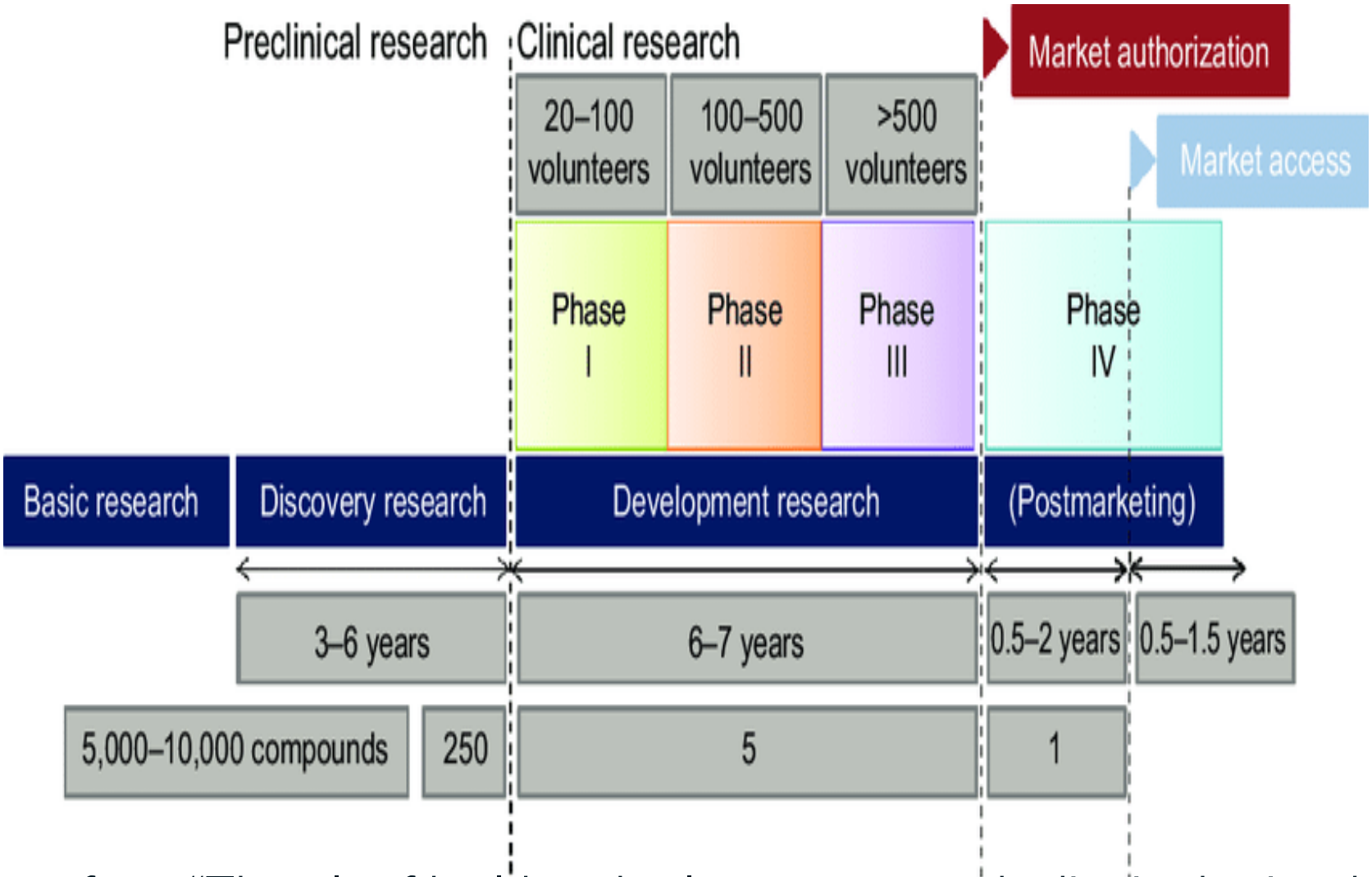


Figure: from "The role of health technology assessment bodies in shaping drug development November 2014Drug Design, Development and Therapy

Learning-Earning in Clinical Trials

Current trend, increased uptake (but still slow) in Adaptive Designs - see Pallmann et al (2018).

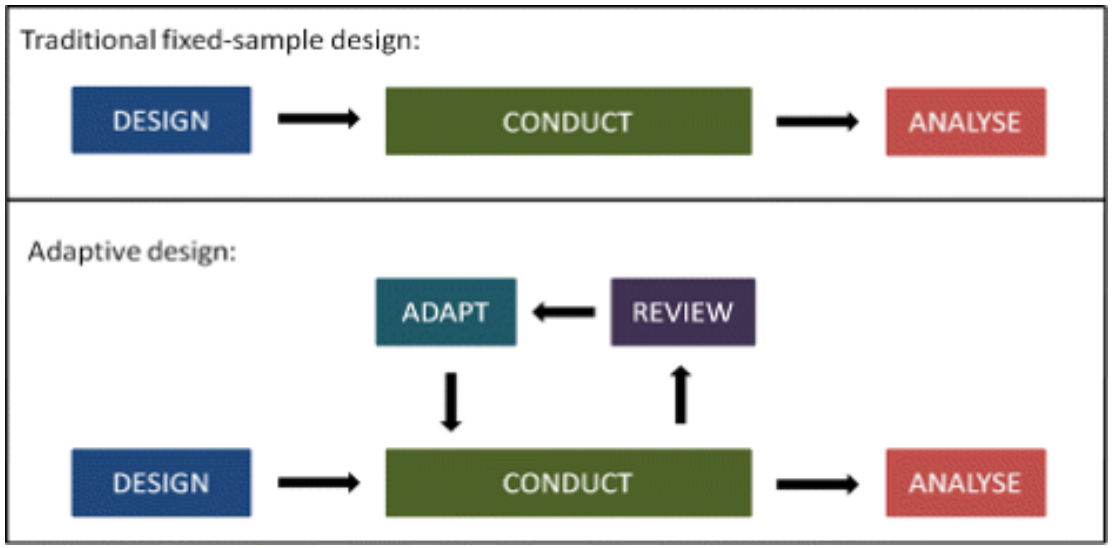


Figure: Figure from Pallmann et al (2018)

probability of treatment by the two methods of $f_{(P)}$ and $1 - f_{(P)}$, respectively. If such a discipline were adopted, even though it were not the best possible, it seems apparent that a considerable saving of individuals otherwise sacrificed to the inferior treatment might be effected. This would be important in cases where either the rate of accumulation of data is slow or the individuals treated are valuable, or both.

Bandit algorithms are a form of Adaptive designs (known in the biostatistics literature as Response-Adaptive designs). Quote from Thompson 1933 paper

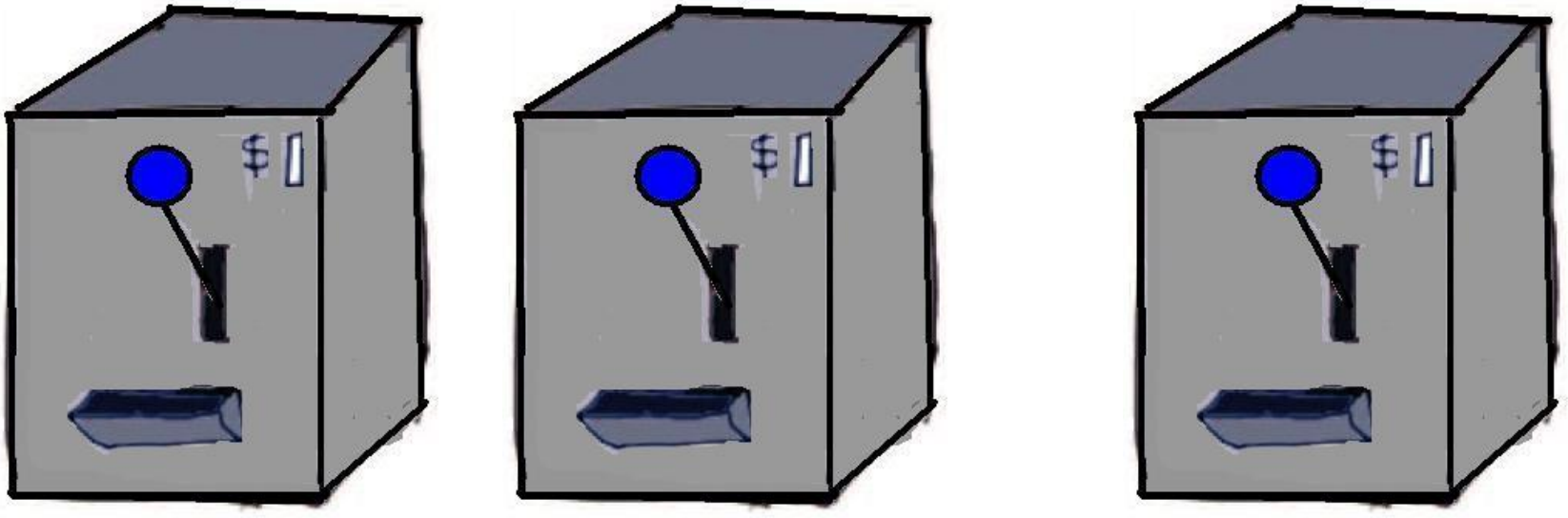
Outline

What is a multi-armed bandit problem?

"In fact, the problem represents in a simplified way the general question of how we learn - or should learn - from past experience." - Peter Whittle

The K-armed Bandit Problem

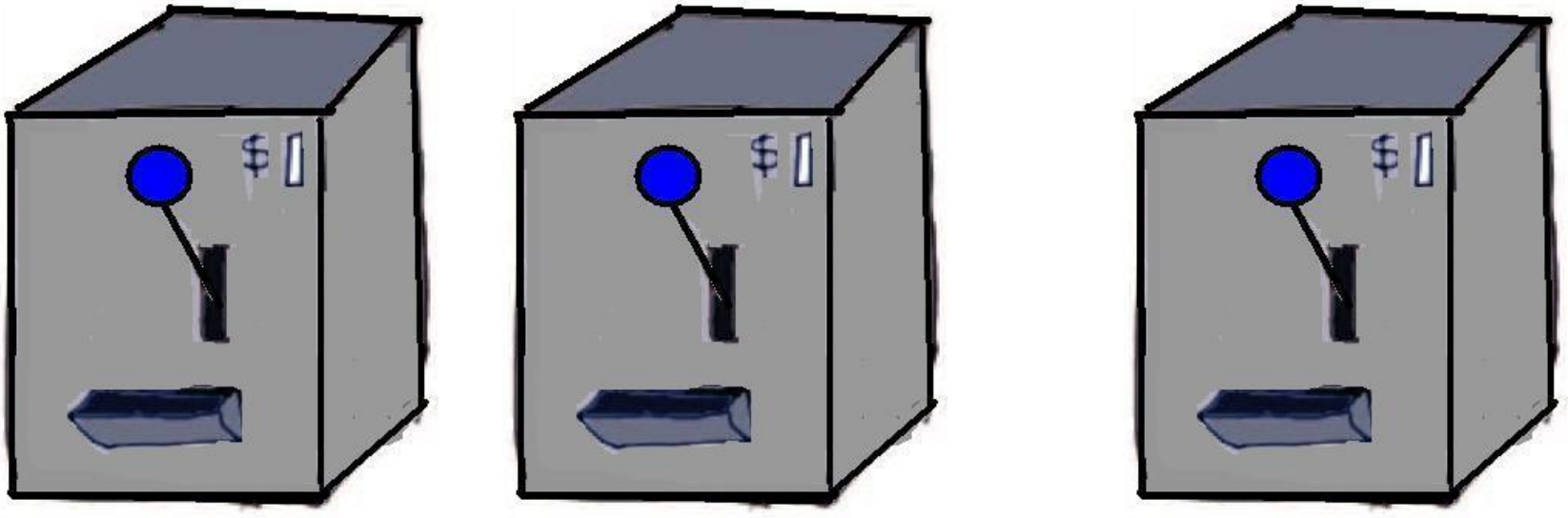
Bandit1 Bandit 2 ... Bandit K



How to optimally play this game? Depends on how we formulate it

The K-armed Bandit Problem

Bandit1 Bandit 2 ... Bandit K



How to optimally play this game? **Depends on how we formulate it**

The "Classic" Bayesian Bernouilli MABPs

A popular formulation is for a binary outcome $Y_{k,t} \sim \text{Bernoulli}(p_k)$:

- **Temporal space:** Discrete decision times $t = \{0, 1, \ldots, t\}$
- **State variable:** Information state. $X_k(t) \triangleq \{S_{k,t}, F_{k,t}\}$
- **Action space:** Decision variables. $a_{k,t} = \{0, 1\}$
- **Restrictions**: Play only 1 bandit at a time $\sum_{k=0}^{K} a_{k,t} \leq 1 : \ \forall t$
- **One-period Dynamics:** Markovian transition probabilities $p_k^a(x, y) \triangleq \mathbb{P}(x \to y)$ for every $x, y \in \mathbb{X}_k$ and $a \in A_k$ (Bayes Update)
- **Objective function:** one-period reward function $R_k(x_k(t), a_k(t)) = a_{k,t}\mathrm{E}(Y_{k,t}|x_{k,t})$
- **Time Horizon:** $T < \infty$ (Dynamic Programming) or $T = \infty$ (Gittins' Index) for $0 \leq d < 1$

This is one specific bandit formulation, there are many others out there!

The "Classic" Bayesian Bernouilli MABPs

A popular formulation is for a binary outcome $Y_{k,t} \sim \text{Bernoulli}(p_k)$:

- **Temporal space:** Discrete decision times $t = \{0, 1, \ldots, t\}$
- **State variable:** Information state. $X_k(t) \triangleq \{S_{k,t}, F_{k,t}\}$
- **Action space:** Decision variables. $a_{k,t} = \{0, 1\}$
- **Restrictions**: Play only 1 bandit at a time $\sum_{k=0}^{K} a_{k,t} \leq 1 : \ \forall t$
- **One-period Dynamics:** Markovian transition probabilities $p_k^a(x, y) \triangleq \mathbb{P}(x \to y)$ for every $x, y \in \mathbb{X}_k$ and $a \in A_k$ (Bayes Update)
- **Objective function:** one-period reward function $R_k(x_k(t), a_k(t)) = a_{k,t}\mathrm{E}(Y_{k,t}|x_{k,t})$
- **Time Horizon:** $T < \infty$ (Dynamic Programming) or $\underline{T = \infty}$ (Gittins' Index) for $0 \leq d < 1$

This is one specific bandit formulation, there are many others out there!

MABPS and the learning and earning objectives

Why this MABP formulation for clinical trials?

- Let π be a "playing strategy" (or sampling strategy)

  Objective: play/sample to maximise the expected total β-discounted rewards, given the initial information $\boldsymbol{I} = (s_{k,0}, f_{k,0})_{k=1}^{K}$:

$$\text{ENS} := \mathbb{E}_{\boldsymbol{I}}^{\pi}\left[\sum_{t=0}^{T}\sum_{k=1}^{K} a_{k,t}\, Y_{k,t}\beta^{t}\right] \tag{1}$$

  Q: How to make treatment allocations so as to treat as many of these T patients effectively? $\longrightarrow$ π^* maximise (1)

Theory for discounted Markov Control Process ensures existence of optimal solution π^* for $T \to \infty$ (Bellman's principle of optimality).

The Classic MABPs: from Formulations to Solutions

How can we optimally solve a classic MABP formulation?

(1) For $T < \infty$, one can solve it by Dynamic Programming.

Note, brut force needs to explore $2^{T\,K}$ possible states while a DP algorithm performs $\frac{(T-1)!}{(2\,K)!(T-2\,K-1)!}$ operations (Villar et al, 2015). $K = 2$ Julia 1.0.1 RAM 31 GB T=300 1.6sec T_{max} 1440 Jacko (2019)

(2) For $T = \infty$ with $0 \leq d < 1$, the Gittins Index (GI) rule is optimal. Computation time grows quadratically on T for a given state. Gittins and Jones (1979)

(3) Strategies to nearly optimal solve (1) for large T. Heuristics: going from a computationally feasible solution method to a satisfactory performance of the objective. Examples are Thompson (1933); Villar et al (2015)

The challenge of balancing rewards and inference

- Set $T = 148$ and $K = 2$. Consider the following strategies in two settings: a null and an alternative

	$H_0 : p_0 = p_1 = 0.3$			$H_1 : p_0 = 0.3\,,\ p_1 = 0.5$		
	α	p^* (s.d.)	ENS (s.d.)	$1-\beta$	p^* (s.d.)	ENS (s.d.)
ER	0.052	0.500 (0.04)	44.3 (5.6)	**0.809**	0.501 (0.04)	59.2 (6.0)
DP	0.028	0.500 (0.35)	44.4 (5.6)	0.078	0.888 (0.17)	**70.7** (8.0)
TS	0.066	0.499 (0.10)	44.4 (5.6)	0.795	0.685 (0.09)	64.8 (6.6)
GI	*0.053*	0.501 (0.26)	44.4 (5.6)	0.364	0.862 (0.11)	70.2 (7.1)

Design Analysis. Comparison of the OCs of different two-arm trial designs of size $T = 148$. α: type I error; $1-\beta$: power; p^*: expected number of patients assigned to best treatment; ENS: expected number of patient successes;

UB: upper bound. In Villar et al (2015); Villar and Jacko (2022)

ML Why ER? Typically the goal is optimise performance.

CT ER, sufficient power, good statistical properties (frequentist), simple tests. Maximise Learning. Simplicity of design!

The challenge of balancing rewards and inference

- Set $T = 148$ and $K = 2$. Consider the following strategies in two settings: a null and an alternative

	$H_0 : p_0 = p_1 = 0.3$			$H_1 : p_0 = 0.3\,,\ p_1 = 0.5$		
	α	p^* (s.d.)	ENS (s.d.)	$1-\beta$	p^* (s.d.)	ENS (s.d.)
ER	0.052	0.500 (0.04)	44.3 (5.6)	**0.809**	0.501 (0.04)	59.2 (6.0)
DP	0.028	0.500 (0.35)	44.4 (5.6)	0.078	0.888 (0.17)	**70.7** (8.0)
TS	0.066	0.499 (0.10)	44.4 (5.6)	0.795	0.685 (0.09)	64.8 (6.6)
GI	*0.053*	0.501 (0.26)	44.4 (5.6)	0.364	0.862 (0.11)	70.2 (7.1)

Design Analysis. Comparison of the OCs of different two-arm trial designs of size $T = 148$. α: type I error; $1-\beta$: power; p^*: expected number of patients assigned to best treatment; ENS: expected number of patient successes;

UB: upper bound. In Villar et al (2015); Villar and Jacko (2022)

ML Why ER? Typically the goal is optimise performance.

CT ER, sufficient power, good statistical properties (frequentist), simple tests. Maximise Learning. Simplicity of design!

The challenge of balancing rewards and inference

- Set $T = 148$ and $K = 2$. Consider the following strategies in two settings: a null and an alternative

	$H_0 : p_0 = p_1 = 0.3$			$H_1 : p_0 = 0.3\,,\ p_1 = 0.5$		
	α	p^* (s.d.)	ENS (s.d.)	$1-\beta$	p^* (s.d.)	ENS (s.d.)
ER	0.052	0.500 (0.04)	44.3 (5.6)	**0.809**	0.501 (0.04)	59.2 (6.0)
DP	0.028	0.500 (0.35)	44.4 (5.6)	0.078	0.888 (0.17)	**70.7** (8.0)
TS	0.066	0.499 (0.10)	44.4 (5.6)	0.795	0.685 (0.09)	64.8 (6.6)
GI	*0.053*	0.501 (0.26)	44.4 (5.6)	0.364	0.862 (0.11)	70.2 (7.1)

Design Analysis. Comparison of the OCs of different two-arm trial designs of size $T = 148$. α: type I error; $1-\beta$: power; p^*: expected number of patients assigned to best treatment; ENS: expected number of patient successes;

UB: upper bound. In Villar et al (2015); Villar and Jacko (2022)

ML Why ER? Typically the goal is optimise performance.

CT ER, sufficient power, good statistical properties (frequentist), simple tests. Maximise Learning. Simplicity of design!

The challenge of balancing rewards and inference

- Set $T = 148$ and $K = 2$. Consider the following strategies in two settings: a null and an alternative

	$H_0 : p_0 = p_1 = 0.3$			$H_1 : p_0 = 0.3\,,\ p_1 = 0.5$		
	α	p^* (s.d.)	ENS (s.d.)	$1-\beta$	p^* (s.d.)	ENS (s.d.)
ER	0.052	0.500 (0.04)	44.3 (5.6)	**0.809**	0.501 (0.04)	59.2 (6.0)
DP	0.028	0.500 (0.35)	44.4 (5.6)	0.078	0.888 (0.17)	**70.7** (8.0)
TS	0.066	0.499 (0.10)	44.4 (5.6)	0.795	0.685 (0.09)	64.8 (6.6)
GI	*0.053*	0.501 (0.26)	44.4 (5.6)	0.364	0.862 (0.11)	70.2 (7.1)

Design Analysis. Comparison of the OCs of different two-arm trial designs of size $T = 148$. α: type I error; $1-\beta$: power; p^*: expected number of patients assigned to best treatment; ENS: expected number of patient successes;

UB: upper bound. In Villar et al (2015); Villar and Jacko (2022)

ML Why ER? Typically the goal is optimise performance.

CT ER, sufficient power, good statistical properties (frequentist), simple tests. Maximise Learning. Simplicity of design!

Outline

MABP and the Gittins index for a clinical trial

Limitations

" Their chief practical motivation comes from clinical trials..." - John Gittins ('79 Biometrika).

Yet, bandit index rules are still to be applied to a real clinical trial.

Many important **barriers** to its practical implementation remain.

(1) Patients' Outcomes $Y_{k,t}$ must be immediately available

(2) Optimal decisions are not randomized (potential for treatment allocation bias)

(3) Insufficient statistical power to detect a relevant difference, lack of type I error control

(4) Appropriate, valid and efficient tests in finite samples.

(5) Others (not obvious): bias in estimation of treatment effect (typically over-estimation), temporal trends and error control, how to deal with prognostic covariates and missing data, etc.

MABP and the Gittins index for a clinical trial

Limitations

" Their chief practical motivation comes from clinical trials..." - John Gittins ('79 Biometrika).

Yet, bandit index rules are still to be applied to a real clinical trial.

Many important **barriers** to its practical implementation remain.

(1) Patients' Outcomes $Y_{k,t}$ must be immediately available

(2) Optimal decisions are not randomized (potential for treatment allocation bias)

(3) Insufficient statistical power to detect a relevant difference, lack of type I error control

(4) Appropriate, valid and efficient tests in finite samples.

(5) Others (not obvious): bias in estimation of treatment effect (typically over-estimation), temporal trends and error control, how to deal with prognostic covariates and missing data, etc.

Outline

1. Introducing Randomization to the Gittins Index Rule

The Forward Looking Gittins Index

Assume T patients are enrolled sequentially in groups of size b over J stages, so that $J \times b = T$ is the trial size. In Villar et al (2015) we defined group allocation probabilities based on the GI as follows:

Simplest example: $b = 2$. Priors: control $(s_{(0,0)}, f_{(0,0)}) = (1, 2)$ and experimental $(s_{(1,0)}, f_{(1,0)}) = (1, 1)$

$$j = 1, \quad \mathcal{G}_1(1,1) = 0.8699, \quad \mathcal{G}_0(1,2) = 0.7005$$

$\frac{1}{2}$ / $Y_{1,0} = 1$

$$j = 2, \quad \mathcal{G}_1(2,1) = 0.9102, \quad \mathcal{G}_0(1,2) = 0.7005$$

$\frac{1}{2}$ / $Y_{1,0} = 0$

$$j = 2, \quad \mathcal{G}_1(1,2) = 0.7005, \quad \mathcal{G}_0(1,2) = 0.7005$$

What is the (patient-average) probability of each arm being allocated in the next block using the GI (and given the priors)?

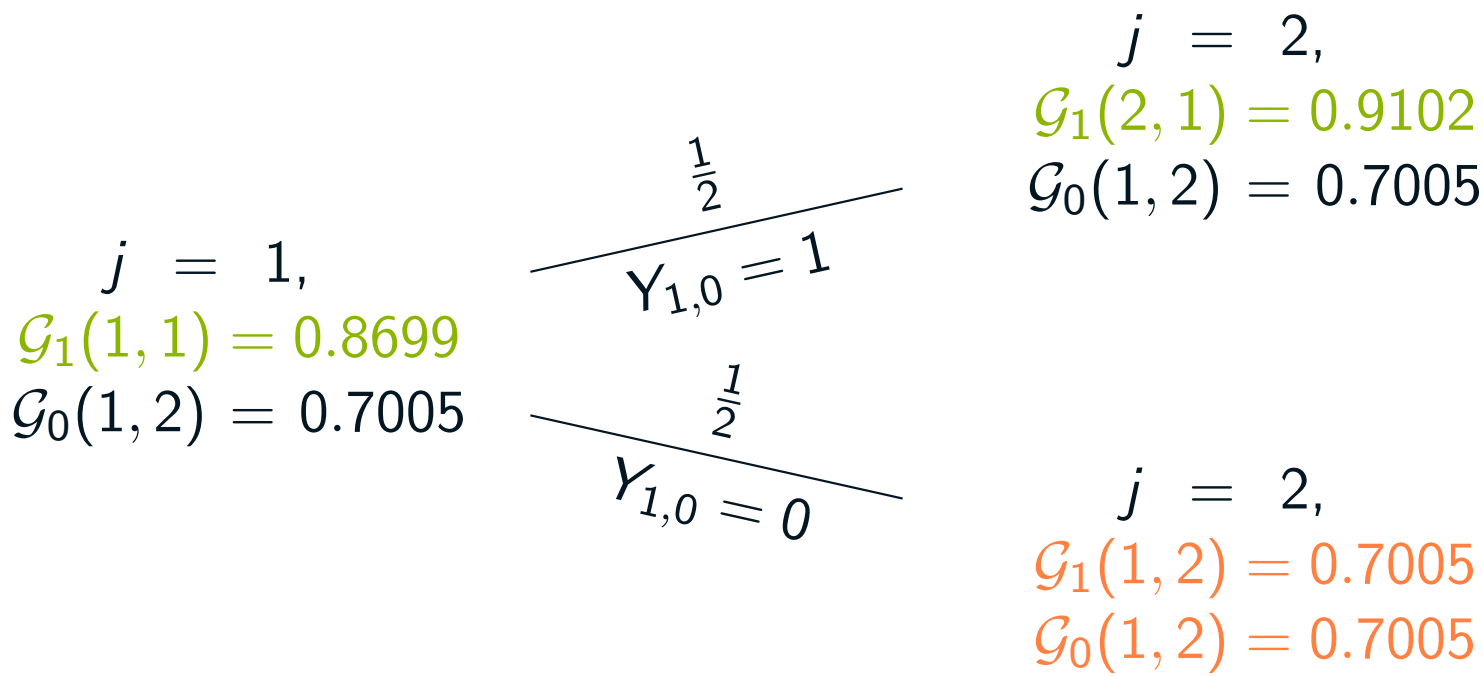
$$\pi_{1,0} = \frac{(0 \times 1) + (0 \times 1/2 + 1/2 \times 1/2)}{2} = 1/8 \ , \ \pi_{1,1} = \frac{(1 \times 1) + (1 \times 1/2 + 1/2 \times 1/2)}{2} = 7/8.$$

1. Introducing Randomization to the Gittins Index Rule

The Forward Looking Gittins Index

Assume T patients are enrolled sequentially in groups of size b over J stages, so that $J \times b = T$ is the trial size. In Villar et al (2015) we defined group allocation probabilities based on the GI as follows:

Simplest example: $b = 2$. Priors: control $(s_{(0,0)}, f_{(0,0)}) = (1, 2)$ and experimental $(s_{(1,0)}, f_{(1,0)}) = (1, 1)$



What is the (patient-average) probability of each arm being allocated in the next block using the GI (and given the priors)?

$$\pi_{1,0} = \frac{(0 \times 1) + (0 \times 1/2 + 1/2 \times 1/2)}{2} = 1/8 \ , \ \pi_{1,1} = \frac{(1 \times 1) + (1 \times 1/2 + 1/2 \times 1/2)}{2} = 7/8.$$

FLGI Probabilities: Computation & Properties

A Non-myopic Group Randomised Procedure

C FLGI probabilities **can be computed exactly** but they are computationally infeasible. Just as for the MABP, the computational cost explodes with the number of arms (K) and b (block size).

In practice, and in my papers computation done via Monte Carlo simulation. Example: $\boldsymbol{I} = [1\ 1\ ;\ 2\ 1\ ;\ 1\ 2\ ;\ 2\ 2]$ ($K = 4$) and block $b = 9$ then $\boldsymbol{\pi} \approx [0.2646\ ;\ 0.5901\ ;\ 0.0246\ ;\ 0.1208]$ after $5 * 10^2$ replicas of 9 patients and 4 arms assigned via GI rule.

P1 For **equal priors** the algorithm defines **equal allocation probabilities** or balanced sampling.
As the block size tends to grow (in the limit it equals the trial size T), the design tends to a **balanced design** (given initial equipoise).

If the block is of only 1 patient (i.e. there is an interim after every patient), the FLGI rule recovers the **GI rule**.

P2 The choice of b controls the level of randomisation (impacts power).

2. Incorporating Covariate Information to the Gittins Index

Increasing Patient Benefit by Personalising Treatment

MABP with covariates: let patient outcome $Y_{k,t} \sim Bernoulli(p_k(z_t))$ where $Z_t \sim Bernoulli(q)$ (with q known).
E.g., $p_k(z_n) = Expit(\alpha_k + \beta_k z_t)\ \forall t$, where $\text{Expit}(u) = \frac{exp(u)}{1+exp(u)}$.

For patient t, we observe their covariate value z_t then we treat them.

P: What if patients in the two subgroups respond differently to treatment? (treatment-covariate interaction)

- Solving the associated MABP with DP: computational complexity even larger than in the classic case. (Deterministic)

Q: Can we define a simple randomised index rule in this case? Some work in the literature: Clayton '89; Woodroofe '79

- In Villar and Rosenberger (2017): a heuristic (extended) FLGI rule for a binary endpoint with a discrete covariate with C levels. Reformulated MABP: for every treatment-covariate combination there exists a combination arm kz. E.g., the arm "00" corresponds to the control arm

The CARA FLGI in Practice

Simulation Results

3-arm trial 300 patients $p_{k0} = (0.22; 0.34; \mathbf{0.49})$, $p_{k,1} = (0.47; \mathbf{0.71}; 0.37)$. Treatment-covariate interaction: **best** arm for covariate **negative** patients is **arm 2** while for covariate **positive** patients is **arm 1**.

	Power		Patient Benefit		
	$(1-\beta_0)$	$(1-\beta_1)$	p_0^* (s.d)	p_1^* (s.d)	ENS (s.d)
ER $_{(b=300)}$	0.82	0.63	0.33 (0.04)	0.33 (0.04)	**130.71 (9.3)**
CARA CFLGI $_{(b=10)}$	**0.85**	**0.79**	**0.55 (0.16)**	**0.62 (0.06)**	**148.36 (9.6)**
CARA FLGI $_{(b=10)}$	0.13	0.03	0.75 (0.22)	0.86 (0.16)	166.73 (11.2)
CARA GI $_{(b=1)}$	0.11	0.03	0.78 (0.24)	0.88 (0.18)	169.39 (11.4)

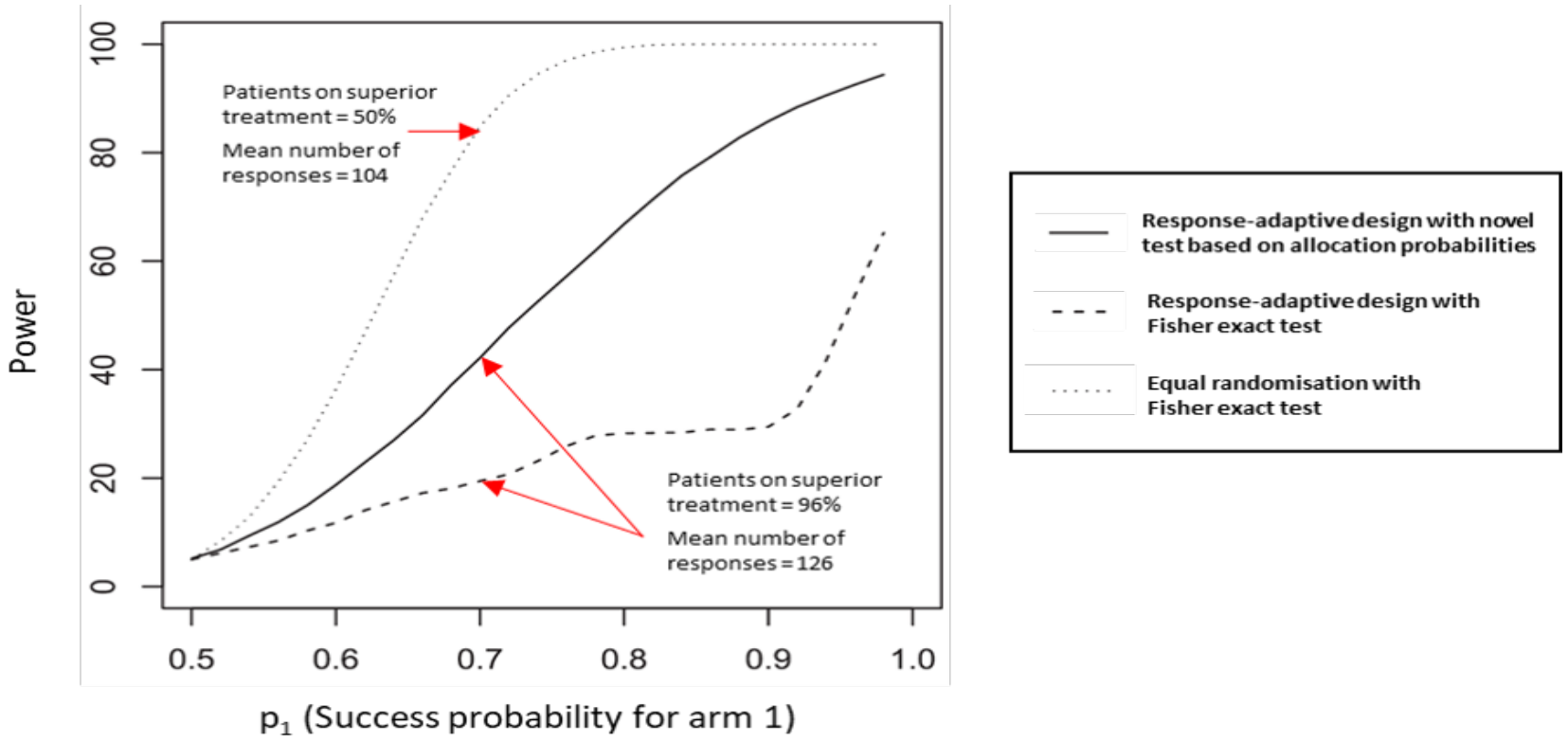
CARA FLGI probabilities (Monte Carlo simulation), $T = 300$, $p_z = 0.5$ and 5000 runs.

- **Treatment-covariate interactions are detected** by the CARA (Covariate-Adjusted Response Adaptive) FLGI procedure but its statistical power is naturally very low.
- In a multi-armed case the CARA CFLGI addresses the power limitation (though in a two-arm setting power may be insufficient).

Outline

An allocation test for increased power for the FLGI algorithms

In Barnett et al (2021), we present results for a new test, called Q^{FLGI} in that paper, specifically designed for the GI designs in Villar et al (2015); Villar and Rosenberger (2017)



- Effects on power: Q-test considerable increased power/ (no) reduction in ENS (patient benefit). Still not ideal, further work

Power for N=160 & B=2

Case study: 2-arm CT with 160 patients $p_0 = 0.5$. FLGI versus ER for variuos p_1 when covariate's cardinality C increases

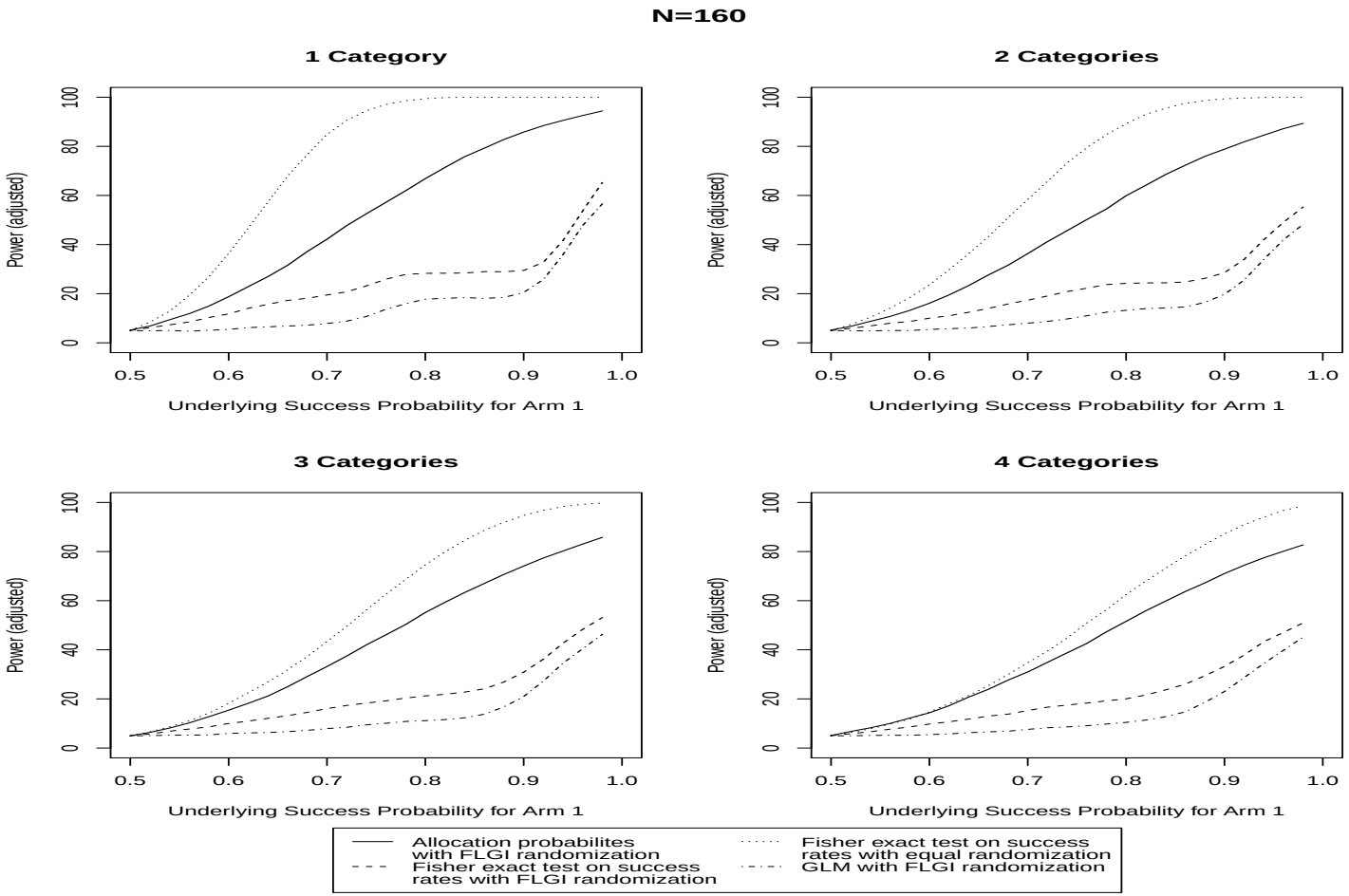


Figure: Comparison of power for $N = 160$ & $B = 2$; rejection criteria adjusted for type I error rate.

Outline

Recent developments & Ongoing Debate

Response-adaptive randomization in clinical trials: from myths to practical considerations

David S. Robertson[1], Kim May Lee[1], Boryana C. López-Kolkovska[1], and Sofía S. Villar *[1]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Opinion

VIEWPOINT

Optimizing the Trade-off Between Learning and Doing in a Pandemic

Derek C. Angus, MD, MPH
University of Pittsburgh and UPMC Health System, Pittsburgh, Pennsylvania; and Associate Editor, *JAMA*.

The world is united regarding the goal of ending the coronavirus disease 2019 (COVID-19) pandemic but not the strategy to achieve that goal. One stark example is the debate over whether to prescribe available therapies, such as quinine-based antimalarial drugs (eg, chloroquine or hydroxychloroquine), or test these drugs in randomized clinical trials (RCTs). At the heart of the problem is one of the oldest dilemmas in human organizations: the "exploitation-exploration" trade-off.[1] Exploitation refers to acting on current knowledge, habits, or beliefs despite uncertainty. This is the "just do

Three Major Challenges to Learning While Doing

The chief tool in the learning toolkit is the RCT, primarily because randomization is such a powerful mechanism for inferring causal effects. It is not perfect, and there are alternatives, but in the absence of a miracle drug that dramatically eradicates the disease, randomization will be crucial to determine what therapies work. There are, however, 3 major challenges.

Randomization is profoundly uncomfortable. Kalil has suggested that a clinician who wishes to administer chloroquine (rather than defer to randomized assignment)

Clin Infect Dis. 2020 Dec 31;71(11):3002-3004. doi: 10.1093/cid/ciaa334.

Resist the Temptation of Response-Adaptive Randomization

Michael Proschan [1], Scott Evans [2]

Affiliations
PMID: 32222766 DOI: 10.1093/cid/ciaa334

Abstract

Response-adaptive randomization (RAR) has recently gained popularity in clinical trials. The intent is noble: minimize the number of participants randomized to inferior treatments and increase the amount of information about better treatments. Unfortunately, RAR causes many problems, including

The Temptation of Overgeneralizing Response-adaptive Randomization

Sofía S Villar ✉, David S Robertson, William F Rosenberger

Clinical Infectious Diseases, ciaa1027,
https://doi.org/10.1093/cid/ciaa1027
Published: 22 July 2020 **Article history** ▾

TO THE EDITOR—We read with interest the recent article by Proschan and Evans [1] on the use of response-adaptive randomization (RAR) and its potential problems; however, these problems are neither new nor applicable in general to all

Concluding Remarks

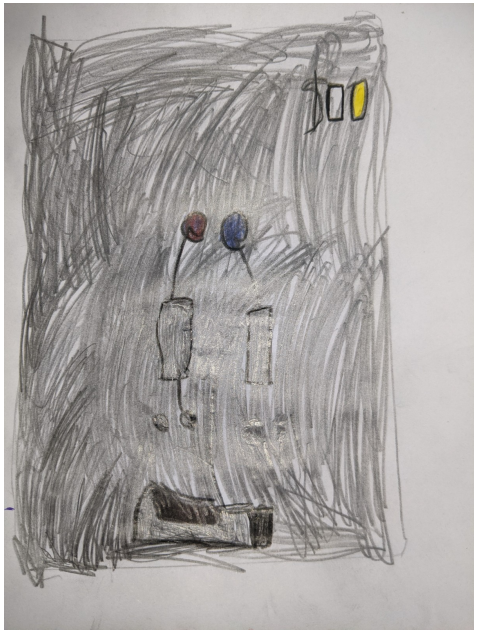
- Bandit algorithms are a form of Adaptive designs. These can offer ethical and efficiency advantages but will add complexity to analysis to ensure validity.
- In many contexts, experiments have tight constraints on their size (pilot studies) and/or need to"pick up" a signal early.
- Bandit algorithms (like the ones I described based on indices or DPs ideas) result in a good solution to the above problem but ...
- substantial work is still needed to ensure validity and efficiency of adaptive methods in accordance to the trials context and more importantly to drive change in a very traditional community!

Questions?

Thank you for listening! sofia.villar@mrc-bsu.cam.ac.uk

Questions?

Bandit1 Bandit 2 ... Bandit K



With thanks to Bruno (my 5 yo) for these :)

Some References I

Hadad, Vitor and Hirshberg, David A and Zhan, Ruohan and Wager, Stefan and Athey, Susan *Confidence intervals for policy evaluation in adaptive experiments*. Proceedings of the National Academy of Sciences 118.15 (2021)

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

Villar, S., Wason, J. and Bowden, J. Response-adaptive Randomization for Multi-arm Clinical Trials using the Forward Looking Gittins Index rule *Biometrics*. 71(4): 969-978, 2015

Williamson, S Faye and Jacko, Peter and Villar, Sofía S and Jaki, Thomas *A Bayesian adaptive design for clinical trials in rare diseases*. Computational statistics & data analysis 113 136–153 Elsevier

John C Gittins and David M Jones. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, 66(3):561–565, 1979.

Villar, S. S. and Jacko, P. Bandit Procedures for Designing Patient-Centric Clinical Trials *"The Elements of Joint Learning and Optimization in Operations Management"*Estimated pub. date Aug. 2022

Jacko, P. *The finite-horizon two-armed bandit problem with binary responses: A multidisciplinary survey of the history, state of the art, and myths.* arXiv preprint arXiv:1906.10173 (2019).

Deliu, N., Williams, J. and Villar, S. Efficient Inference Without Trading-off Regret in Bandits: An Allocation Probability Test for Thompson Samplings, *(Work in Progress - Submitted)*

Some References II

Hu, Feifang and Rosenberger, William F *The theory of response-adaptive randomization in clinical trials*. John Wiley & Sons

Pallmann P, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med.*: 16(1):29. 2018 doi:10.1186/s12916-018-1017-7

Simon, Richard and Simon, Noah Robin *Using randomization tests to preserve type I error with response adaptive and covariate adaptive randomization*, Statistics & probability letters 81(7)767–772,

Barnett, Helen Yvette and Villar, Sofia S and Geys, Helena and Jaki, Thomas *A Novel Statistical Test for Treatment Differences in Clinical Trials using a Response Adaptive Forward Looking Gittins Index Rule* arXiv preprint arXiv:2011.03270

Thompson, W. R. *On the likelihood that one unknown probability exceeds another in view of the evidence of two samples*. Biometrika 25(3/4) 285–294, 1933.

Villar, S. and Rosenberger, W. Covariate-Adjusted Response-Adaptive Randomization for Multi-Arm Clinical Trials Using a Modified Forward Looking Gittins Index Rule *Biometrics, 74(1), 49-57.*

Smith A. and Villar, S. (2017) A Bayesian bandit design using the Gittins indices for trials with normally distributed endpoints. *Journal of Applied Statistics* 1–25 Taylor & Francis