The Trade-off Between Bias and Efficiency in Borrowing RWD in Hybrid Designed Trials

Ying Lu, PhD, Stanford University, USA
Ruben van Eijk, MD, PhD, UMC Utrecht, the Netherlands
Jiapeng Xu, MS, Stanford University, USA
Tianyu Pan, PhD, Stanford University
Lu Tian, Sc.D., Stanford University, USA

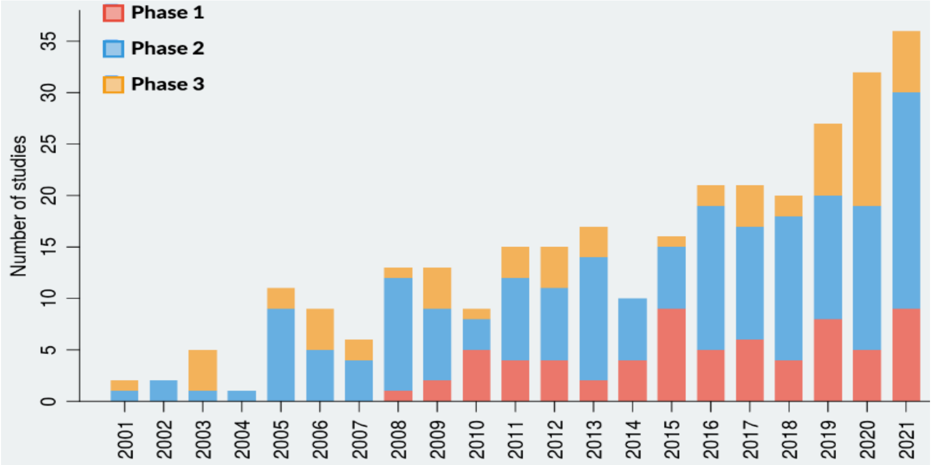
Department of Biomedical Data Science AT STANFORD MEDICINE

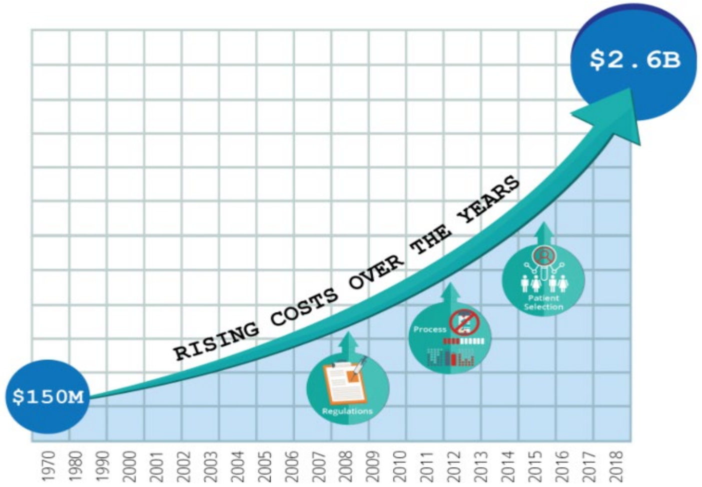
Stanford MEDICINE | Center for Innovative Study Design (CISD)

http://med.stanford.edu/cisd.html

Real world data (RWD) and real-world evidence (RWE)

- Real world data (RWD) is increasingly available to support trial design, safety monitoring, and efficacy evaluation in clinical trials for drug development.
- RWD has been used to augment RCT of rare diseases
  - Amyotrophic Lateral Sclerosis (ALS) is an example of rare diseases

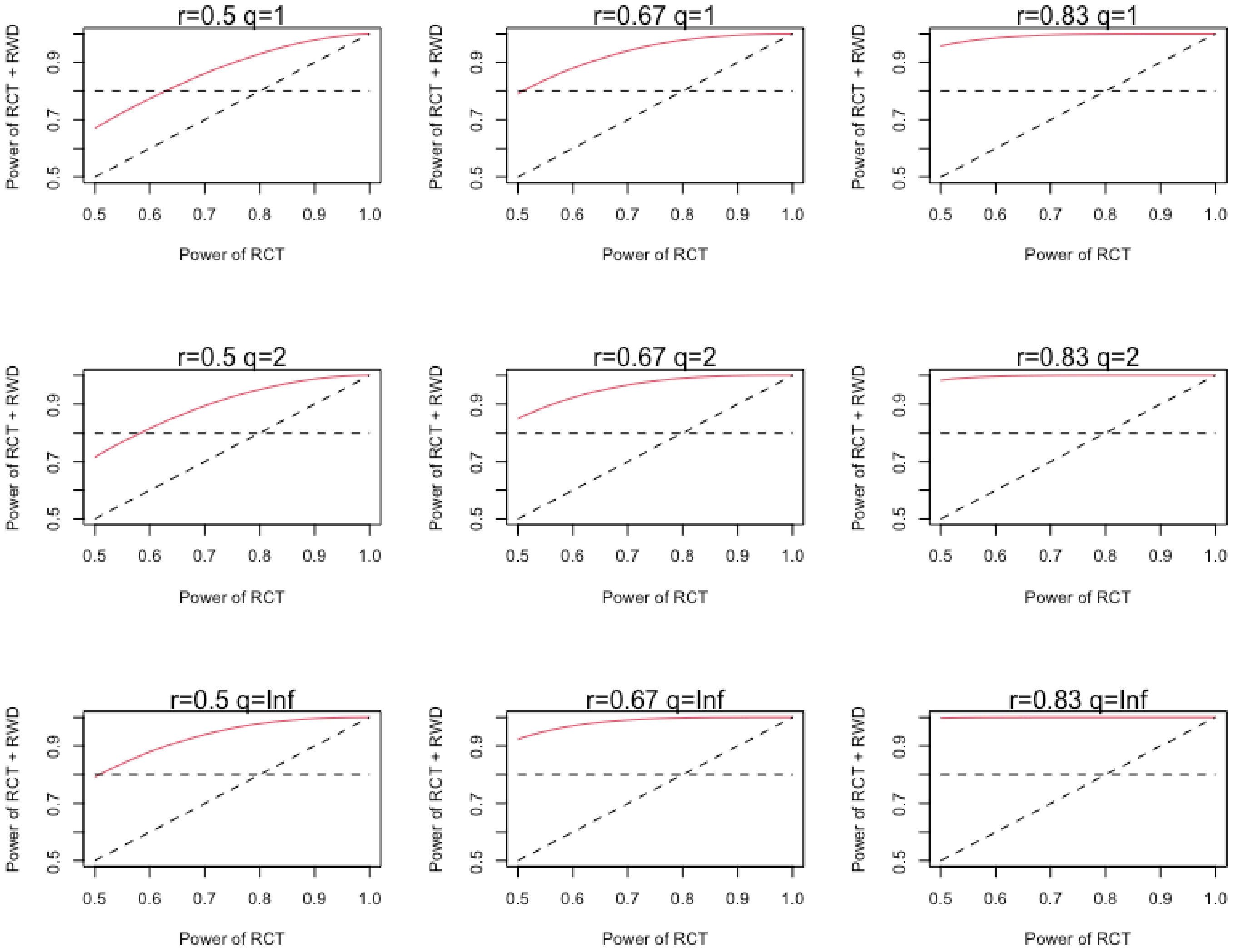




Power improvement by pooling RWD with RCT

- **r**: is the proportion of participants to treatment arm in a two arm RCT

- **q**: is the ratio of the number of perfect matched RWD patients with RCT sample size



Power improvement by pooling RWD with RCT

- Percentage increase in relative sample size when borrowing the RWD for normal and equal variance for power of 60% and 2-sided α of 5%

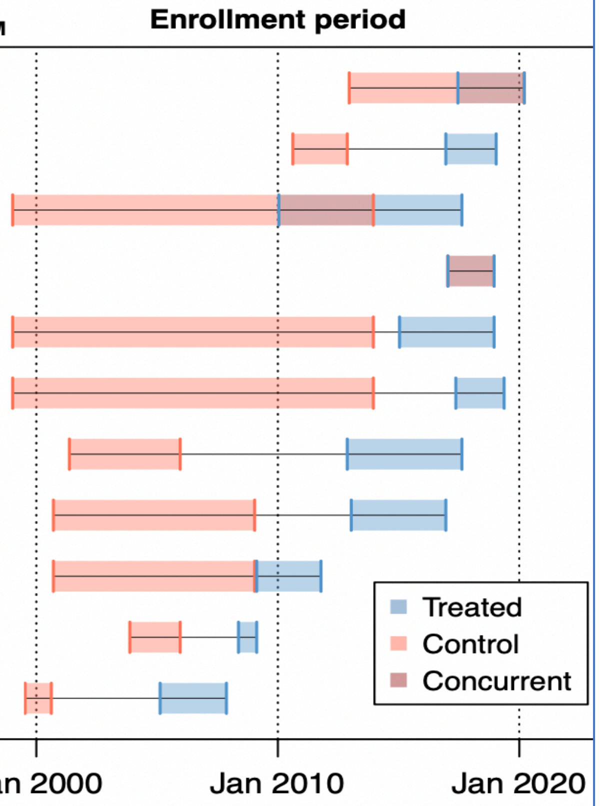
q: ratio between trial and RWD sample size	r:(1-r) randomization ratio for treatment to control arm			
	1:1 (r=0.5)	2:1 (r=0.67)	4:1 (r=0.8)	6:1 (r=0.86)
1	23%	34%	41%	43%
2	28%	38%	42%	43%
∞	34%	41%	43%	43%

- Examples:
  - q=1 and r=0.5, trial power is 80%, RWD borrowing increase power to 93%.
  - q=1 and r=0.86, trial power is 50%, RWD borrowing increase power to 98%.

Historical controls in ALS trials

- Historical controls have been used in ALS trials

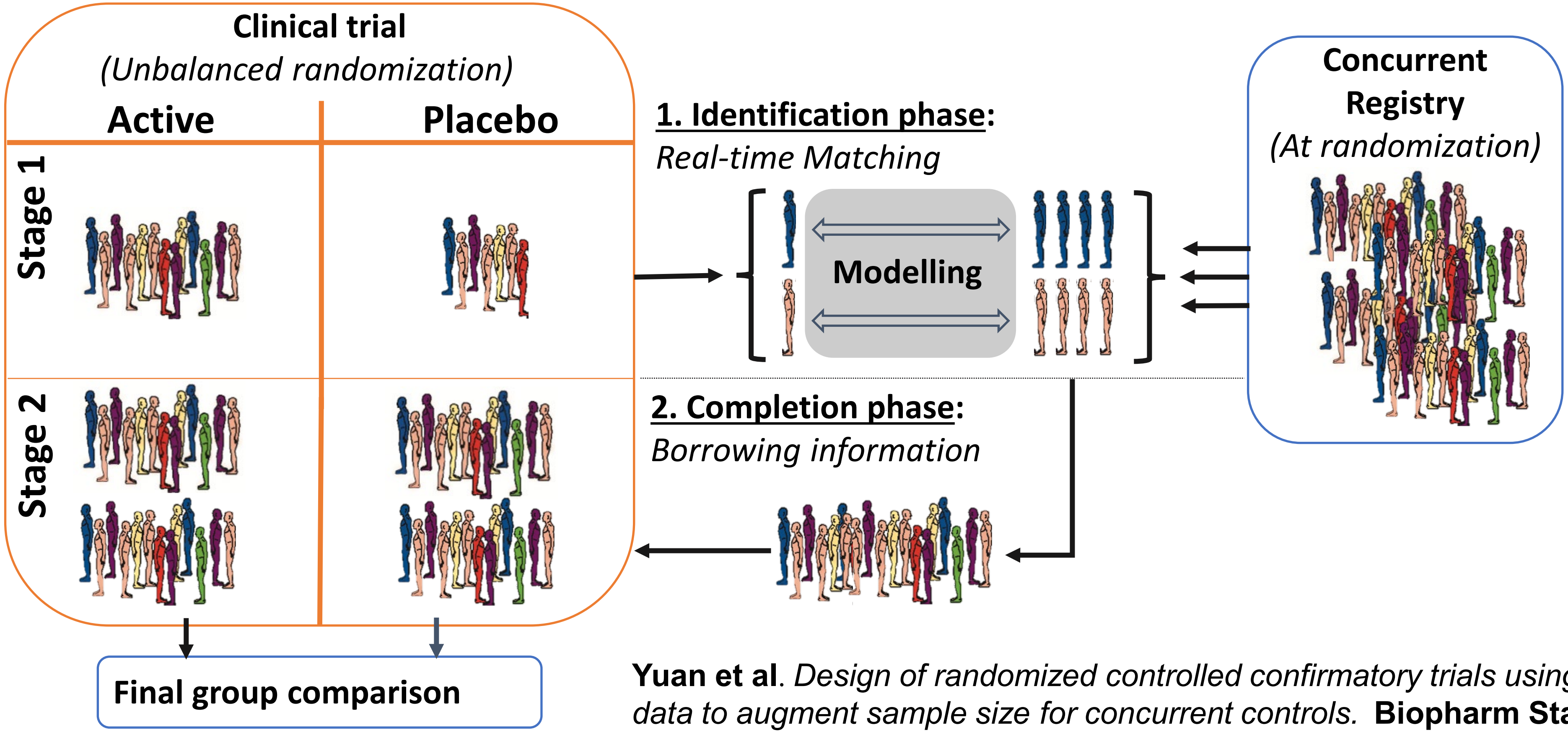
Author, year	Drug	N_T	N_C	Source	Site	Eli.	N_M
Witzel, 2022	Edaravone	130	130	Cohort	Yes	Yes	4
Dalla Bella, 2021*	Guanabenz	151	200	RCT	Yes	No	0
Johannesen, 2021	Filgrastim	36	708	PRO–ACT	No	No	3
Vu, 2020	Edaravone	223	669	Registry	Yes	No	7
Barczewska, 2020	Stem cells	67	67	PRO–ACT	No	No	6
Lunetta, 2020	Edaravone	197	290	PRO–ACT	No	Yes	0
Statland, 2018*	Rasagiline	60	177	RCT	No	Yes	0
Levine, 2017*	L–serine	20	430	RCT	No	No	0
Macchi, 2015	Rasagiline	30	478	RCT	Yes	No	0
Miller, 2011	Lithium	107	249	RCT	No	No	0
Stommel, 2009	Thalidomide	23	279	RCT	No	No	0



Potential bias:
- Time
- Country
- Selection
- …

RWD may not be exchangeable with **RCT**

Methods: a two-step hybrid design



Yuan et al. *Design of randomized controlled confirmatory trials using historical control data to augment sample size for concurrent controls.* **Biopharm Stat. 2019;29(3):558-573 doi: 10.1080/10543406.2018.1559853**

Methods: retrospective application to an RCT

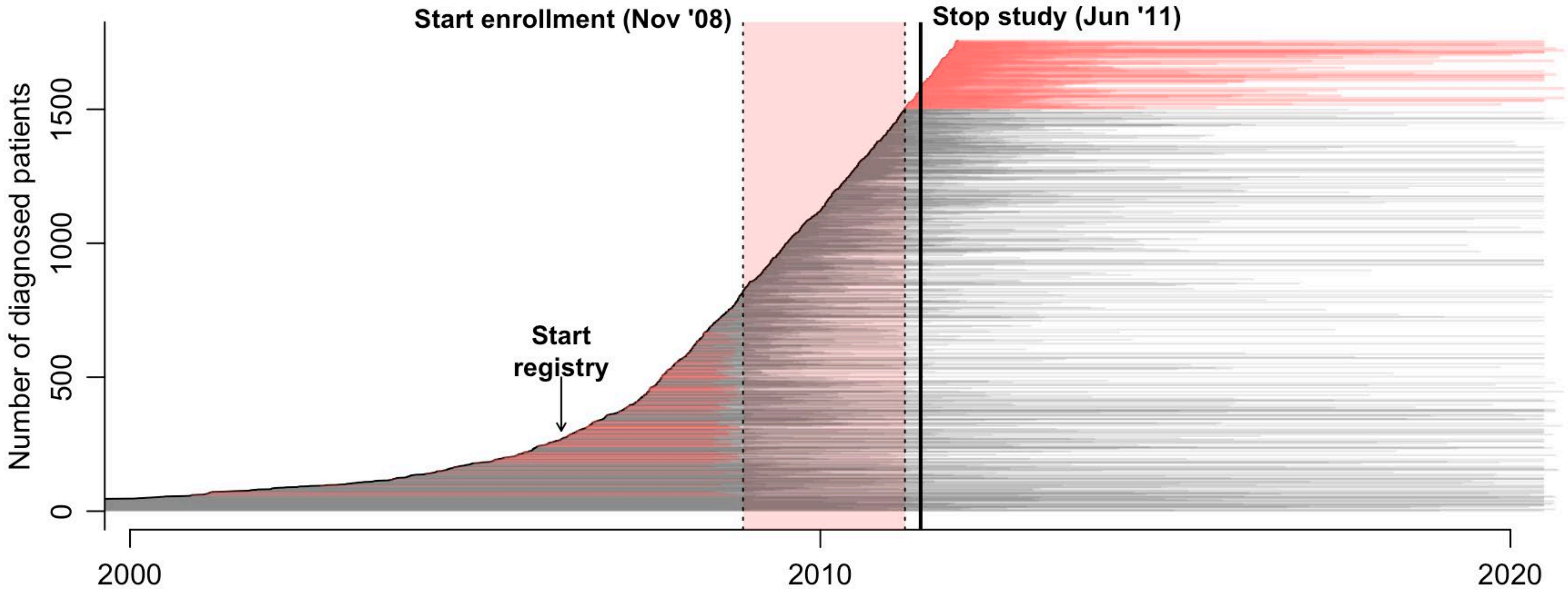
RESEARCH PAPER

Lithium lacks effect on survival in amyotrophic lateral sclerosis: a phase IIb randomised sequential trial

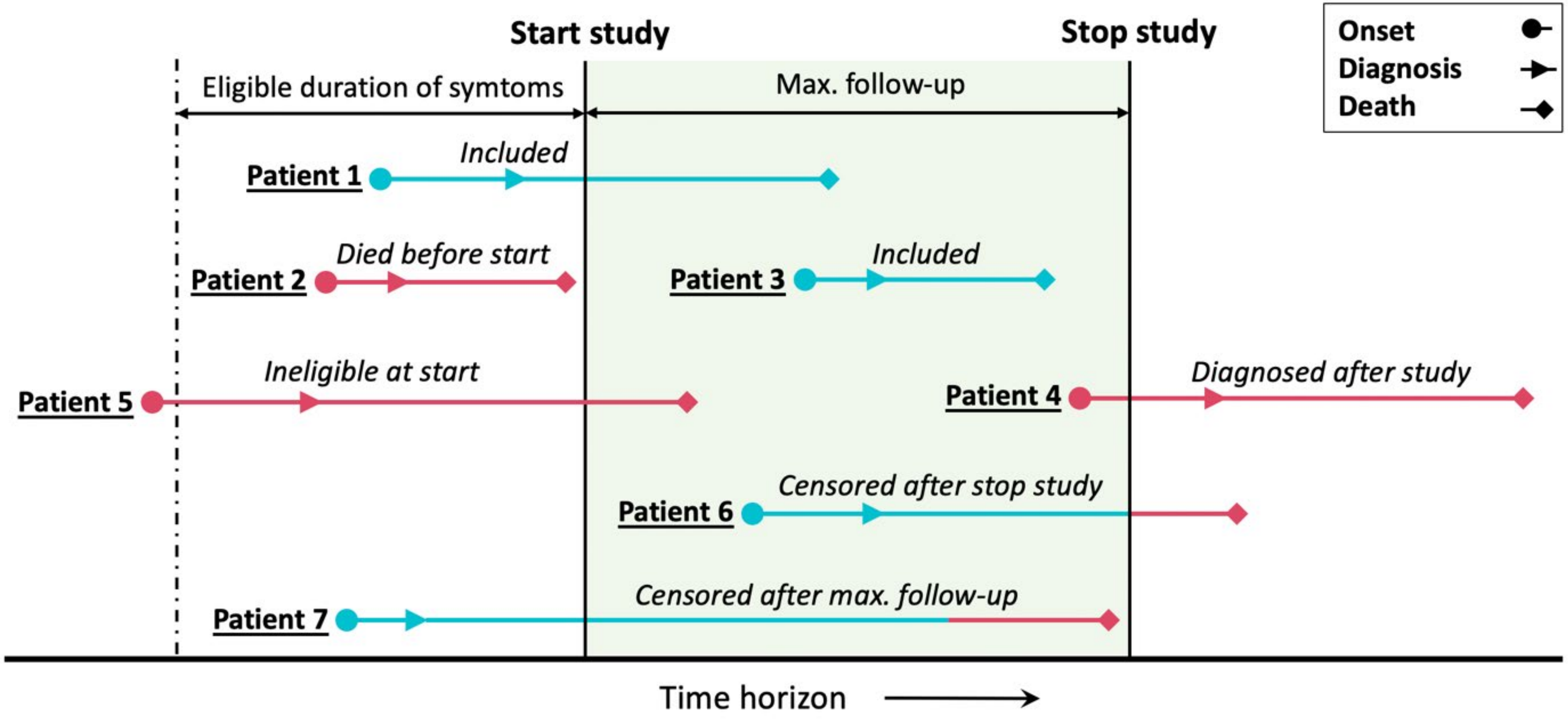
	Lithium group (n=66)	Placebo group (n=67)
Male gender (n (%))	37 (56)	43 (64)
Age (years) (median (IQR))	59.5 (52.3—66.8)	59.0 (51.0—66.5)
Limb onset (n (%))	47 (71)	48 (72)
ALSFRS-R score (median (IQR))	42 (40—44)	41 (37—43)
Forced vital capacity (n (%))		
<85%	18 (27)	20 (30)
≥85%	48 (73)	47 (70)
Location (n (%))		
Utrecht	31 (47)	34 (51)
Amsterdam	19 (29)	17 (25)
Nijmegen	16 (24)	16 (24)

Verstraete et al.
Lithium lacks effect on survival in ALS: a phase IIb randomised sequential trial. JNNP 2012

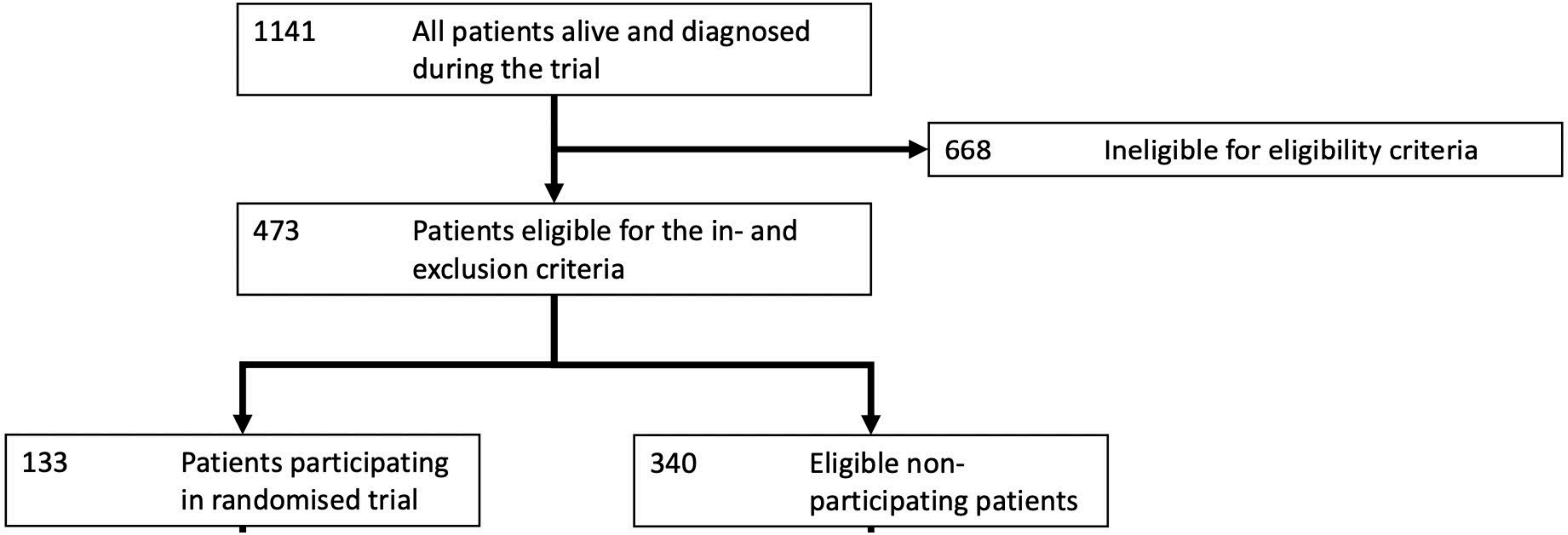
Methods: retrospective application to registry



Methods: Reconstructing the eligible population



Reconstructed eligible population

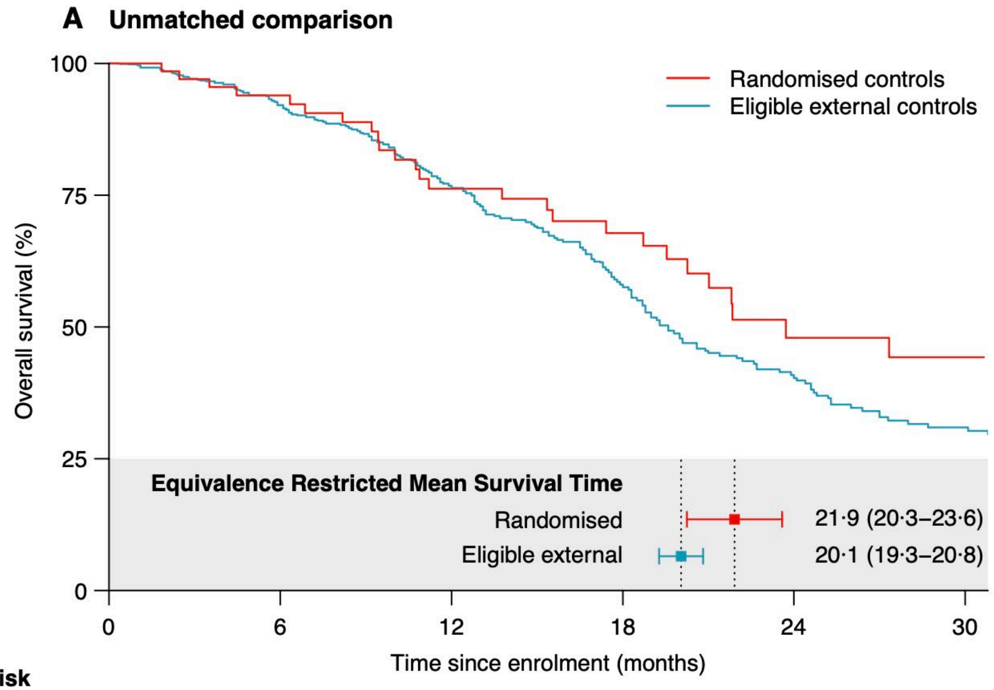


Results: the eligible population

Characteristic	All patients alive during trial	Eligible for in- and exclusion criteria		
	(N = 1,141)	*Non-participants (N = 340)*	*Participants (N = 133)*	*P-value*
Age, *years*	62 (12)	64 (11)	58 (12)	< 0.001
Sex, *male*	677 (59%)	201 (59%)	80 (60%)	0.92
Onset, *spinal*	806 (71%)	238 (70%)	95 (71%)	0.84
El Escorial, *clinically definite*	221 (19%)	71 (21%)	34 (26%)	0.20
Symptom duration, *months**	11.6 (11.4)	10.9 (6.5)	12.4 (10.0)	0.016
FVC, *%predicted*	89 (21)	93 (15)	95 (17)	0.175
ALSFRS-R total score	40 (6)	40 (5)	41 (5)	0.33
ΔFRS, *points per month**	-0.58 (0.73)	-0.59 (0.60)	-0.50 (0.51)	0.044
Body mass index, *kg/m²*	24.8 (4.0)	24.6 (3.8)	25.4 (4.5)	0.081

van Eijk et al. *Hybrid controlled clinical trials using concurrent registries in ALS: A feasibility study.* **Clin Pharmacol Ther. 2023 Jul 8. doi: 10.1002/cpt.2994**

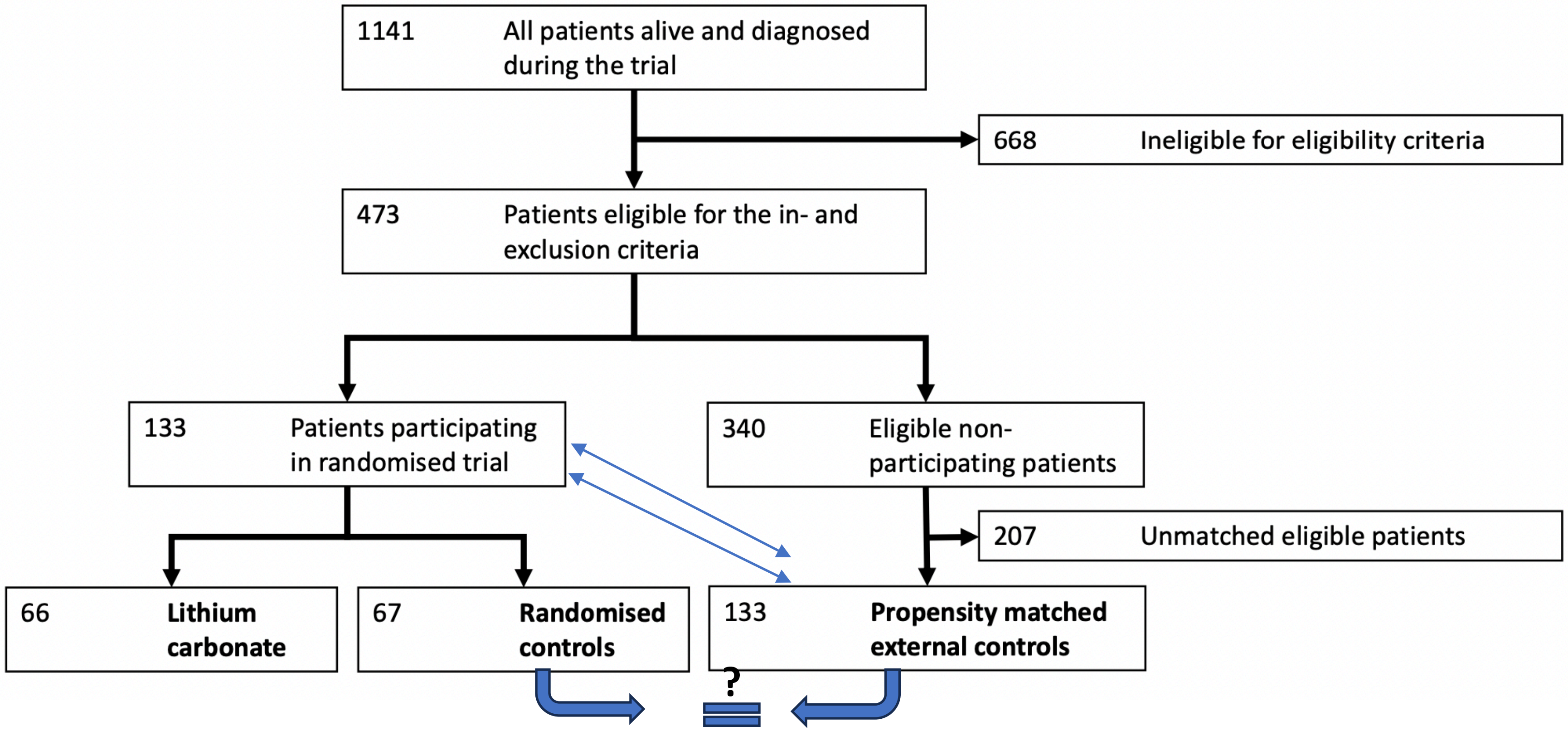
Results: not equivalent in survival



Number at risk (number censored)						
Randomised	67 (0)	56 (7)	41 (12)	29 (20)	14 (28)	2 (39)
Eligible external	340 (0)	277 (37)	186 (85)	117 (112)	70 (126)	47 (134)

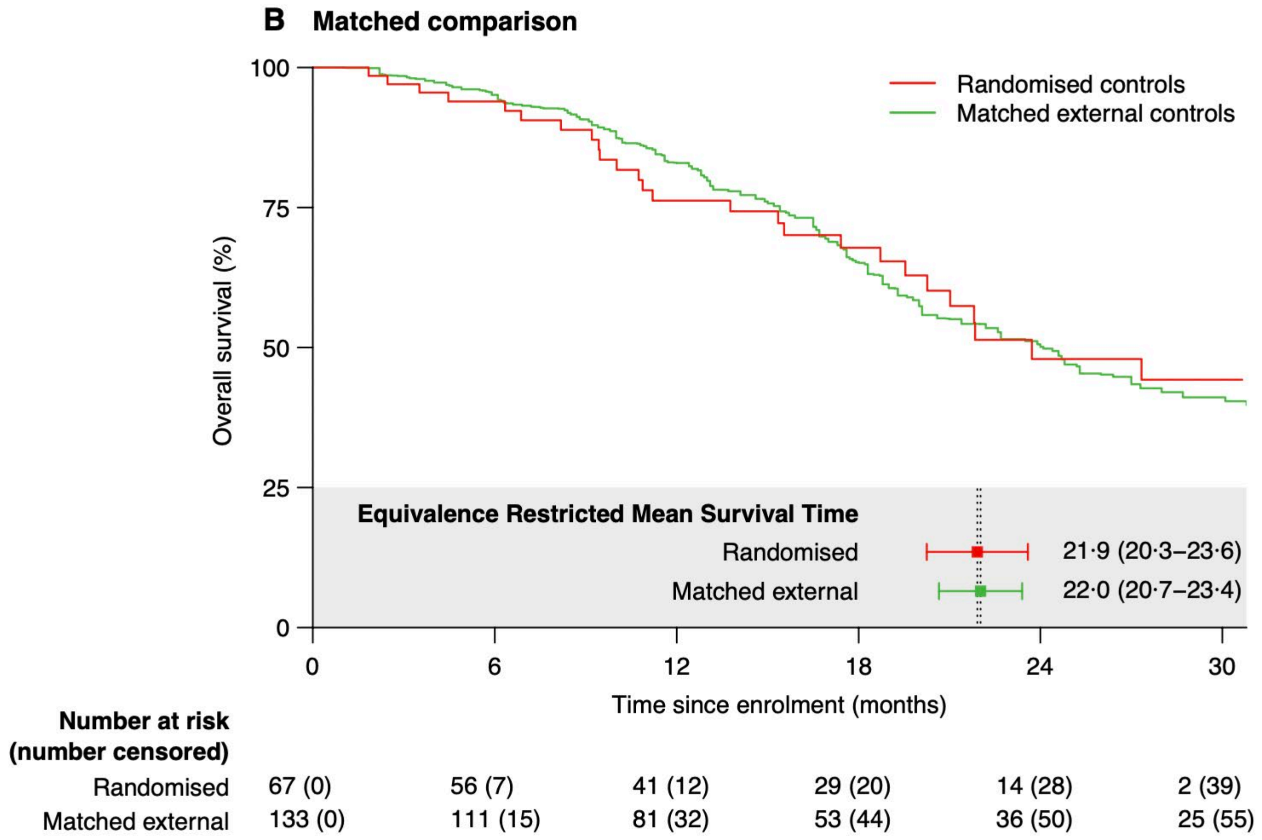
van Eijk et al. *Hybrid controlled clinical trials using concurrent registries in ALS: A feasibility study.* **Clin Pharmacol Ther. 2023 Jul 8. doi: 10.1002/cpt.2994**

Results: identifying external matched controls



van Eijk et al. *Hybrid controlled clinical trials using concurrent registries in ALS: A feasibility study.* **Clin Pharmacol Ther. 2023 Jul 8. doi: 10.1002/cpt.2994**

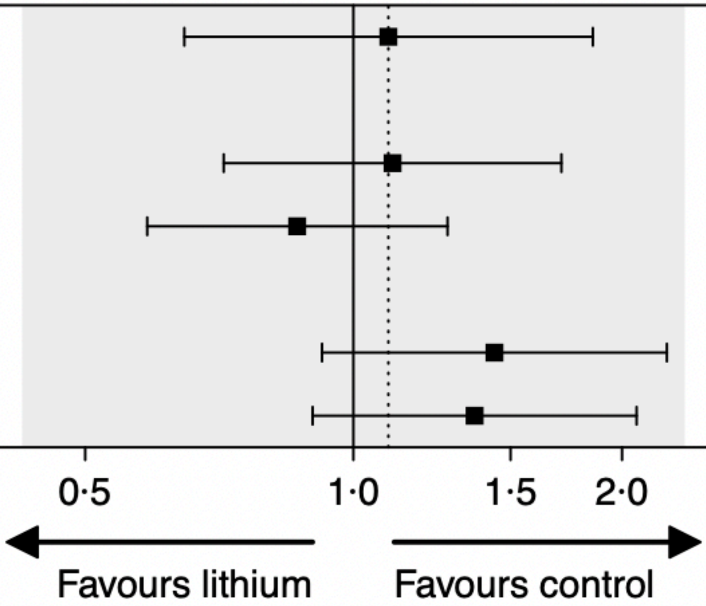
Results: identifying external matched controls



van Eijk et al. *Hybrid controlled clinical trials using concurrent registries in ALS: A feasibility study.* **Clin Pharmacol Ther. 2023 Jul 8. doi: 10.1002/cpt.2994**

Results: augmented comparisons

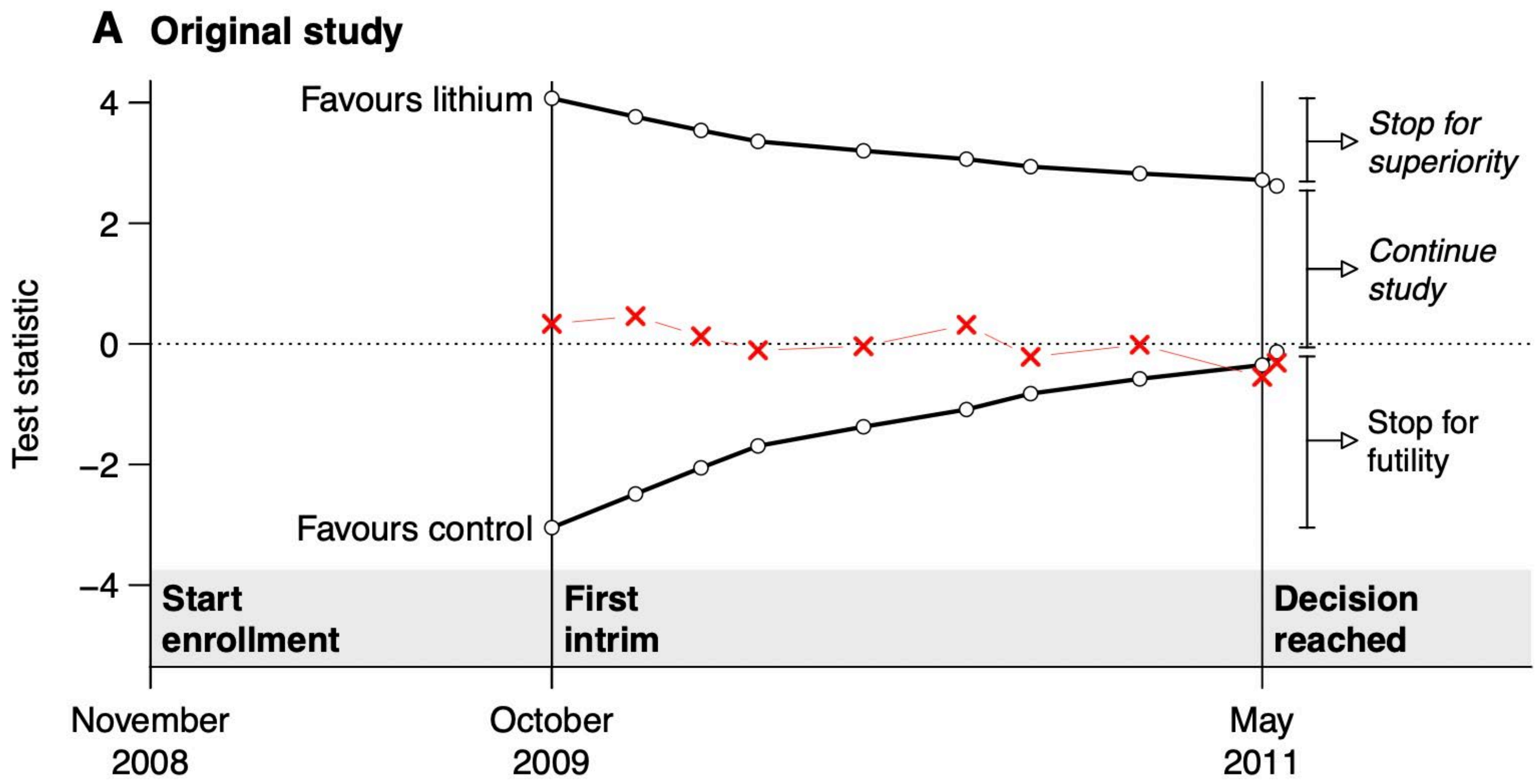
	Lithium carbonate	Randomised control	External control	Equivalence RMST	Hazard ratio (95% CI)
Primary analysis	66	67	0	–	1·09 (0·65–1·85)
Concurrent					
Matched	66	67	133	Yes	1·11 (0·72–1·71)
All eligible	66	67	340	No	0·86 (0·59–1·27)
Non–concurrent					
Matched	66	67	133	No	1·44 (0·92–2·25)
All eligible	66	67	212	No	1·37 (0·90–2·08)



→ **An increase in statistical power from 58.1% to 77.3%**

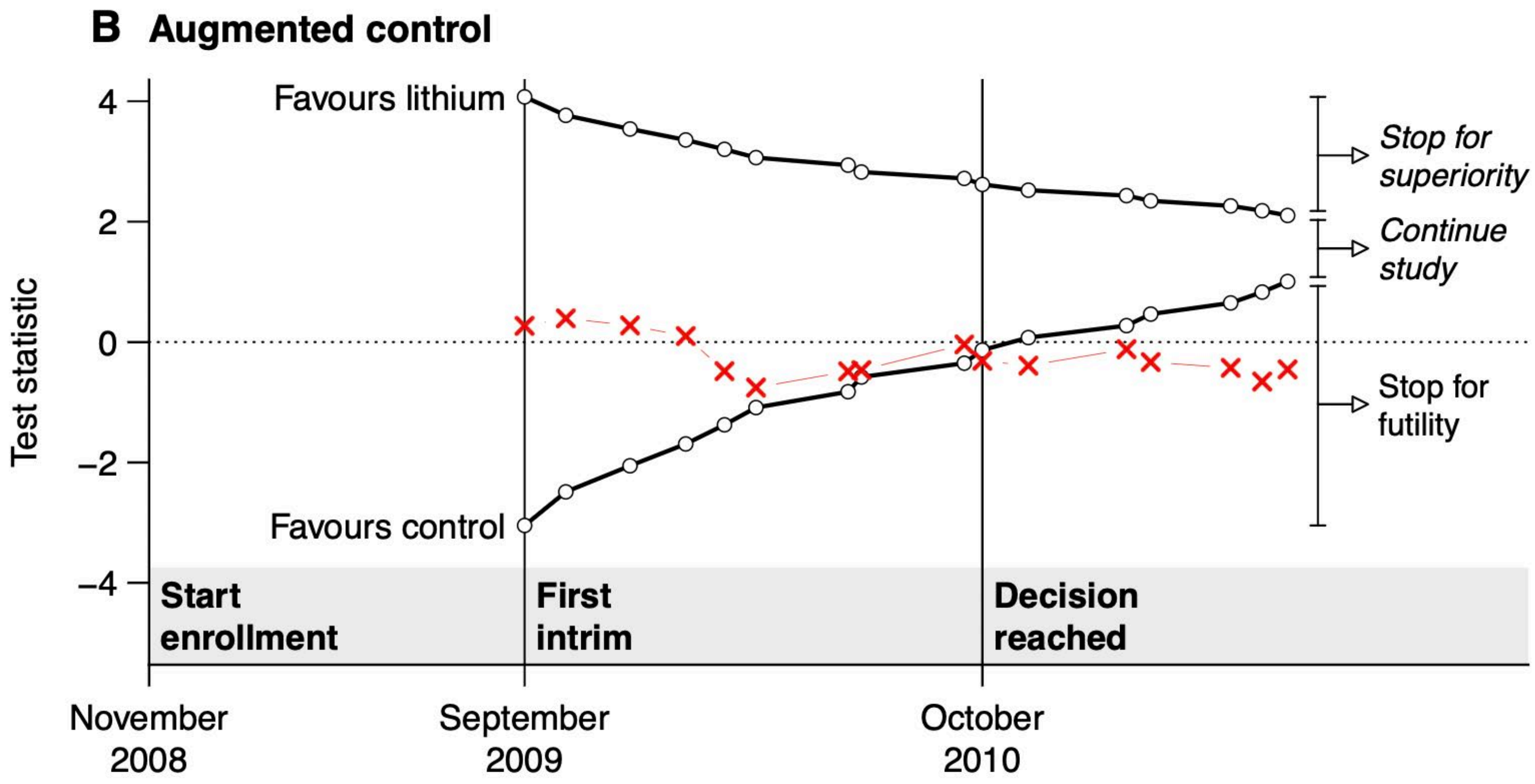
van Eijk et al. *Hybrid controlled clinical trials using concurrent registries in ALS: A feasibility study.* **Clin Pharmacol Ther. 2023 Jul 8. doi: 10.1002/cpt.2994**

Results: impact on decision-making



van Eijk et al. *Hybrid controlled clinical trials using concurrent registries in ALS: A feasibility study.* **Clin Pharmacol Ther. 2023 Jul 8. doi: 10.1002/cpt.2994**

Results: impact on decision-making



van Eijk et al. *Hybrid controlled clinical trials using concurrent registries in ALS: A feasibility study.* **Clin Pharmacol Ther. 2023 Jul 8. doi: 10.1002/cpt.2994**

Mathematical framework of the 2-step hybrid design

- Data from RCT:
  - Y be the statistics of interest for the treatment group, $Y \sim N(\mu_T, \sigma_T^2)$
  - X_1 be the statistics of interest for the control group, $X_1 \sim N(\mu_C, \sigma_C^2)$
  - We are interested in treatment difference:
$$Z_1 = Y - X_1 \sim N(\mu_T - \mu_C, \sigma_T^2 + \sigma_C^2) = N\left(\Delta, \sigma_{Z_1}^2\right)$$
  - Hypotheses to be tested:
$$H_0\text{: } \Delta = 0 \text{ versus } H_1\text{: } \Delta \neq 0$$

Mathematical framework of the 2-step hybrid design

- Data from RWD:
  - X_2 be the statistics of interest from the RWD, $X_2 \sim N(\mu_R, \sigma_R^2)$
  - The difference between the RWD and the control arm

$$Z_2 = X_2 - X_1 \sim N(\mu_R - \mu_C, \sigma_R^2 + \sigma_C^2) = N\left(\delta, \sigma_{Z_2}^2\right)$$

  - To borrow the RWD in the comparison to the treatment arm, we have

$$Z_3 = Y - wX_2 - (1-w)X_1 = Z_1 - wZ_2$$

$$Z_3 \sim N(\Delta - w\delta, \sigma_T^2 + w^2\sigma_R^2 + (1-w)^2\sigma_C^2) = N\left(\Delta - w\delta, \sigma_{Z_3}^2\right)$$

  - The optimal weight w is the regression coefficient, i.e.,

$$w = \sigma_C^2 / \sigma_{Z_2}^2 \text{ that minimizes } \sigma_{Z_3}^2$$

  - Notice that $E(Z_3) = \Delta$ iff $\delta = 0$, i.e., no difference between the control arm and the RWD.

Two-step hybrid clinical trial design by Yuan et al. (2019)

- Step 1: Perform equivalence test between RWD and control arm
  - Let α_2 be the type I error for the equivalence test, δ_M (>0) be the equivalence margin, the equivalence test is

    $$H_{0L}: \delta \le -\delta_M \text{ or } H_{0U}: \delta \ge \delta_M \quad \text{versus} \quad H_1: -\delta_M < \delta < \delta_M$$

  - The null hypothesis will be rejected if $|Z_2| < \delta_M - Z_{1-\alpha_2}\sigma_{Z_2} = \theta$
  - For example, if X_2 is sample mean of 200 RWD patients and X_1 is sample mean of 100 controlled participants, each of from a normal distribution with standard deviation of 1, for δ_M=0.3 and α_2 =10%,

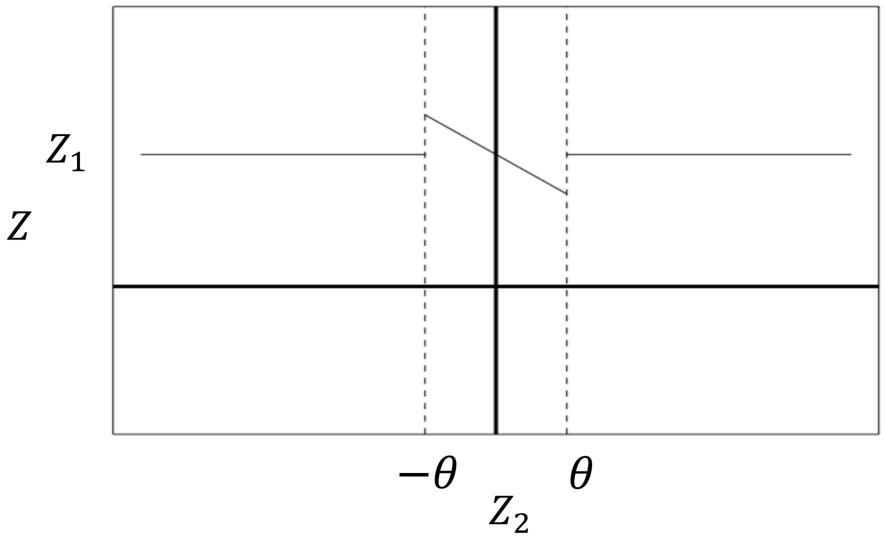
    $$\theta = 0.3 - 1.28 \times \sqrt{\frac{1}{200} + \frac{1}{100}} = 0.14$$

Yuan et al. *Design of randomized controlled confirmatory trials using historical control data to augment sample size for concurrent controls.* **Biopharm Stat. 2019;29(3):558-573 doi: 10.1080/10543406.2018.1559853**

Two-step hybrid clinical trial design by Yuan et al. (2019)

- Step 2: Test the null hypothesis of $\Delta = 0$ conditioning on Step 1
  - Let Z be the final test statistics
  - If $|Z_2| \geq \theta, Z = Z_1$; Reject H_0: $\Delta = 0$ if $|Z_1/\sigma_{Z_1}| > z_{1-\alpha/2}$
  - If $|Z_2| < \theta, Z = Z_3 = Z_1 - wZ_2$; Reject H_0: $\Delta = 0$ if $|Z_3/\sigma_{Z_3}| > z_{1-\alpha/2}$
- Type I error was not adjusted because Step 1 didn't test H_0: $\Delta = 0$.
- However, the critical value for rejecting H_0 needs to account for the adaptation.

- $\boldsymbol{Z = Z_1 - wZ_2 1_{|Z_2|<\theta}}$



Yuan et al. *Design of randomized controlled confirmatory trials using historical control data to augment sample size for concurrent controls.* **Biopharm Stat. 2019;29(3):558-573 doi: 10.1080/10543406.2018.1559853**

Yuan et al. (2019) inflated type I error even when $\Delta = \delta = 0$

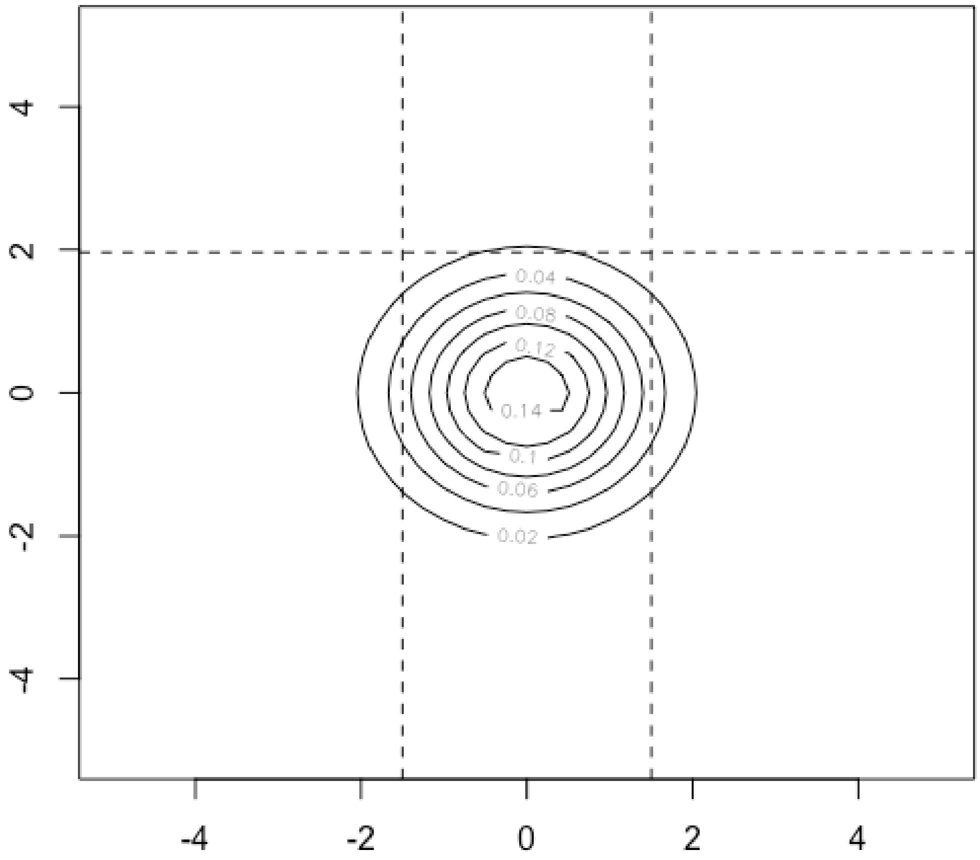
- The type I error

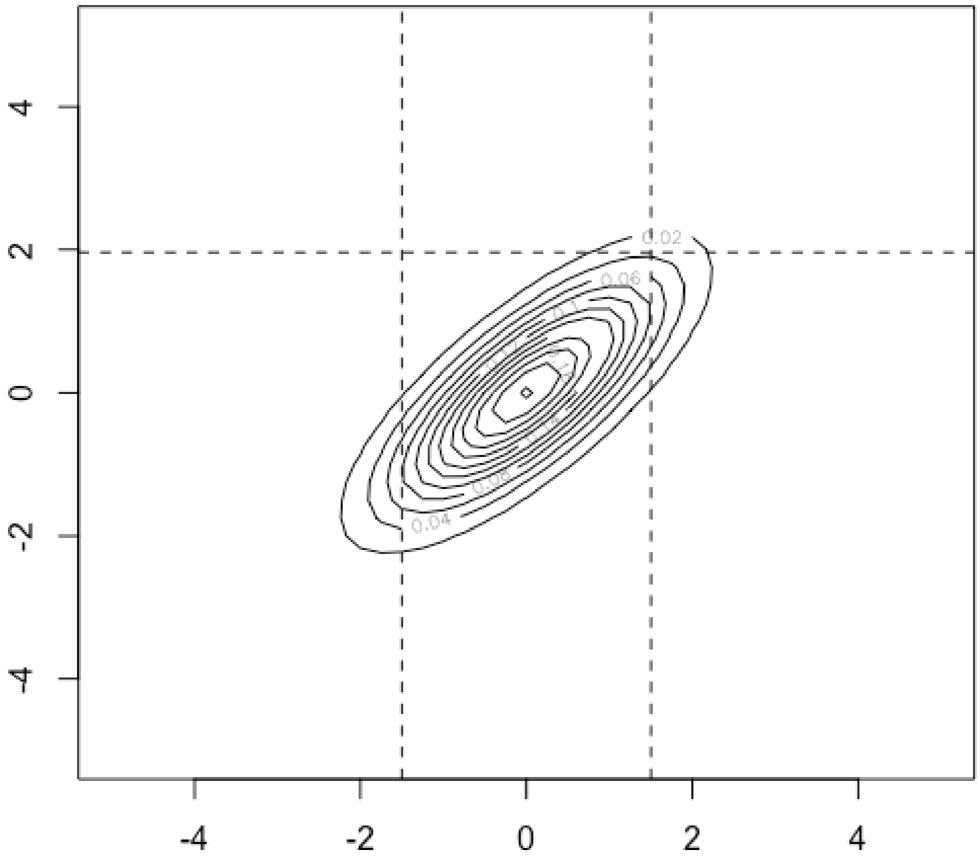
$$
\begin{aligned}
& P\left(Z > z_{1-\alpha/2}\left(\sigma_{Z_1} 1_{1_{|Z_2| \geq \theta}} + \sigma_{Z_3} 1_{|Z_2| < \theta}\right) \mid \Delta = 0, \delta = 0\right) \\
& = P\left(Z_1/\sigma_{Z_1} > z_{1-\alpha/2}, \left|Z_2/\sigma_{Z_2}\right| \geq \theta/\sigma_{Z_2} \mid \Delta = 0, \delta = 0\right) \\
& \quad + P\left(Z_3/\sigma_{Z_3} > z_{1-\alpha/2}, \left|Z_2/\sigma_{Z_2}\right| < \theta/\sigma_{Z_2} \mid \Delta = 0, \delta = 0\right) \\
& = \alpha/2 \\
& \quad + \left\{P\left(Z_3/\sigma_{Z_3} > z_{1-\alpha/2}, \left|Z_2/\sigma_{Z_2}\right| < \theta/\sigma_{Z_2} \mid \Delta = 0, \delta = 0\right)\right. \\
& \quad \left. - P\left(Z_1/\sigma_{Z_1} > z_{1-\alpha/2}, \left|Z_2/\sigma_{Z_2}\right| < \theta/\sigma_{Z_2} \mid \Delta = 0, \delta = 0\right)\right\}
\end{aligned}
$$

- The difference is small but not zero.
- Numerical examples showed inflated type I error.

Yuan et al. *Design of randomized controlled confirmatory trials using historical control data to augment sample size for concurrent controls.* **Biopharm Stat. 2019;29(3):558-573 doi: 10.1080/10543406.2018.1559853**

Yuan et al. (2019) inflated type I error even when $\Delta = \delta = 0$





Methods for type I error control

Approach 1: Normal Approximation

- We can derive the variance of Z and then estimate them from the data.

$$V(Z) = V(Z_1) + w^2 V\left(Z_2 1_{|Z_2|<\theta}\right) - 2\mathrm{w}Cov\left(Z_1, Z_2 1_{|Z_2|<\theta}\right)$$

- We can then reject the null based on normal approximation of Z:

$$|Z| \geq z_{1-\frac{\alpha}{2}}\sqrt{Var(Z)}$$

- $Var(Z)$ has a closed from expression
- Approach 1 can bound bias for both effect size estimation and type I error rate.

Methods for type I error control

Approach 2: Splitting type I error (when $\alpha < 1 - \beta_2$)

- Instead of pre-specified global type I error for Z_1 and Z_3, we can split p-values conditioning on whether $|Z_2| \geq \theta$.
- Let v $(0 < v < 1)$ be the proportion of type I error when $|Z_2| \geq \theta$, we can select the critical value z_1^* based on
$$P\left(Z_1/\sigma_{Z_1} > z_1^* \middle| |Z_2| \geq \theta, \Delta = 0, \delta = 0\right) = v\alpha/(2\beta_2)$$
- then $(1 - v)$ will be given when $|Z_2| < \theta$. The critical value z_3^* can be derived as
$$P\left(Z_3/\sigma_{Z_3} > z_3^* \middle| |Z_2| < \theta, \Delta = 0, \delta = 0\right) = (1 - v)\alpha/\left(2(1 - \beta_2)\right)$$
- Total type I error is
$$P\left(Z > \sigma_{Z_1} z_1^* 1_{|Z_2| \geq \theta} + \sigma_{Z_1} z_3^* 1_{|Z_2| < \theta} \mid \Delta = 0, \delta = 0\right) = \alpha/2$$
- When we select $v = \beta_2$, we will have an equal conditional type I error for two conditions.

Methods for type I error control

Approach 3: A common exact threshold (Li et al. 2020)

- This approach searches a common exact threshold via numerical search for the solution:

$$P\left(Z > z^*\left(\sigma_{Z_1} 1_{1_{|Z_2|\geq\theta}} + \sigma_{Z_3} 1_{|Z_2|<\theta}\right) \mid \Delta = 0, \delta = 0\right)$$
$$= P\left(Z_1/\sigma_{Z_1} > z^*, \left|Z_2/\sigma_{Z_2}\right| \geq \theta/\sigma_{Z_2} \mid \Delta = 0, \delta = 0\right)$$
$$+ P\left(Z_3/\sigma_{Z_3} > z^*, \left|Z_2/\sigma_{Z_2}\right| < \theta/\sigma_{Z_2} \mid \Delta = 0, \delta = 0\right) = \alpha/2$$

Li et al. *Revisit of test-then-poll methods and some practical considerations.* **Pharm Stat. 2020;19(5):498-517.**

Methods for type I error control

Approach 4: Bayesian power prior (Ibrahim et al., SIM 2015)

- Bayesian approach that doesn't require a two-step procedure for the hybrid design
- Instead, it places power prior on the likelihood of RWD
- The power α was chosen to maximized the following marginal likelihood function

$$m(\alpha) = \int L\left(\mu \middle| X_1, \sigma_C^2\right) L\left(\mu \middle| X_2, \sigma_R^2\right)^{\alpha} \pi(\mu) d\mu \Big/ \int L\left(\mu \middle| X_2, \sigma_R^2\right)^{\alpha} \pi(\mu) d\mu$$

- The inference uses the Z-test based on the posterior MLEs.

Ibrahim et al. *The Power Prior: Theory and Application.* **Stat Med. 2015 Dec 10; 34(28): 3724–3749.**

Bias for the estimated treatment effect Δ

- Z_1 is an unbiased estimator of Δ, $E(Z_1) = \Delta$
- Z_3 is an unbiased estimator of Δ only when $\delta = 0$: $E(Z_3) = \Delta - w\delta = \Delta$
- Z can be a biased estimator due to $\delta \neq 0$ because of the type I error of the equivalence test:

$$E(Z) = \Delta - wP(|Z_2| < \theta)E\left(Z_2 \middle| |Z_2| < \theta\right)$$

- When $Z_2 \middle| |Z_2| < \theta$ is symmetric around 0, $E(Z)$ is an unbiased estimator.
- Under normal models,

$$E(Z) = \Delta - w\delta\left[\Phi\left(\frac{\theta - \delta}{\sigma_{Z_2}}\right) - \Phi\left(\frac{-\theta - \delta}{\sigma_{Z_2}}\right)\right] + w\sigma_{Z_2}\left(\phi\left(\frac{\theta - \delta}{\sigma_{Z_2}}\right) - \phi\left(\frac{-\theta - \delta}{\sigma_{Z_2}}\right)\right)$$

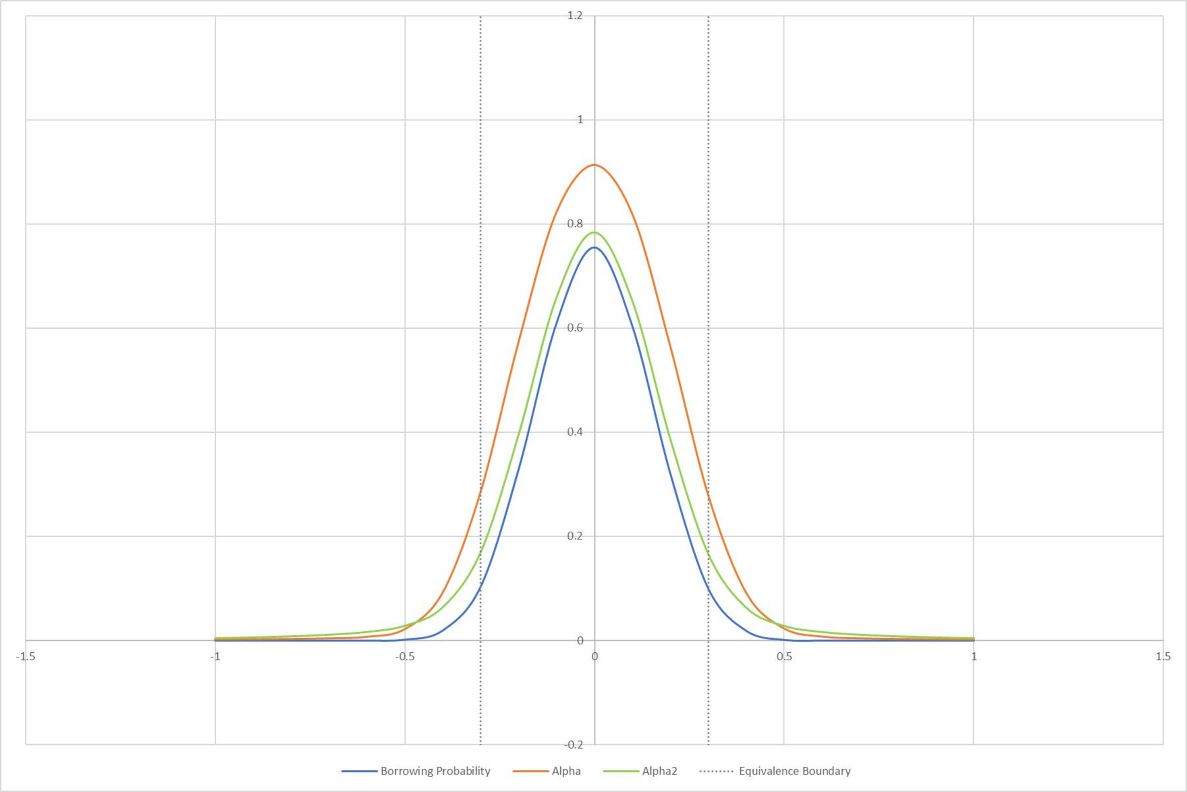
A simulation study

Simulation conditions:

- Clinical trial randomization ratio is 1:1
- 100 participants in treatment and control arms, respectively
- 200 patients are matched from RWD
- $Y_i \sim N(\mu_T, 1)$; $X_{1,j} \sim N(\mu_C, 1)$; and $X_{2,k} \sim N(\mu_R, 1)$
- Equivalence margin is 0.3 and type I error for the equivalent trial is 10%
- Each simulation condition repeated 10,000 times.

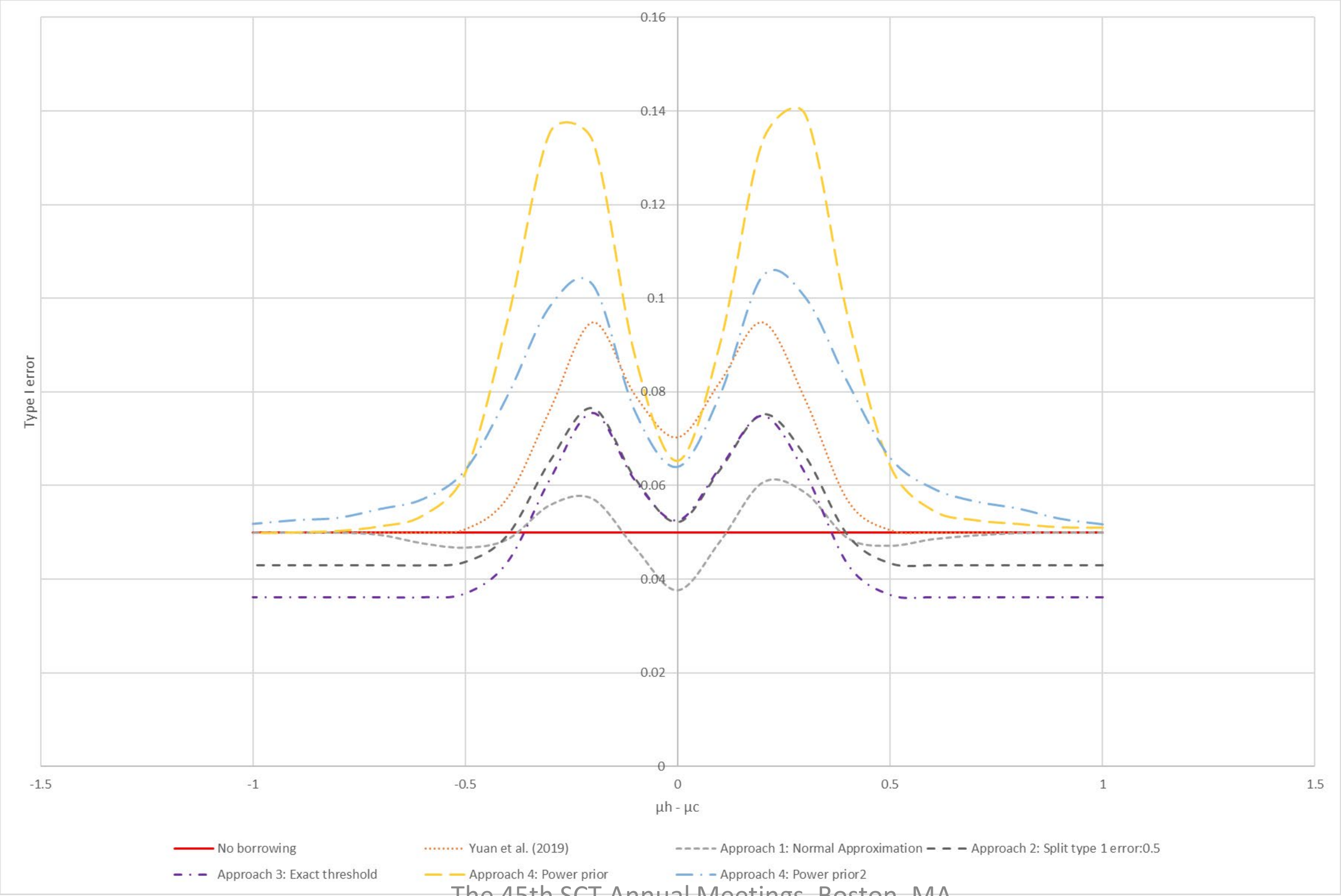
A simulation study

Probability of Borrowing and Bayesian Power prior (alpha)



A simulation study

Two-sided type I error rate when $\Delta = 0$

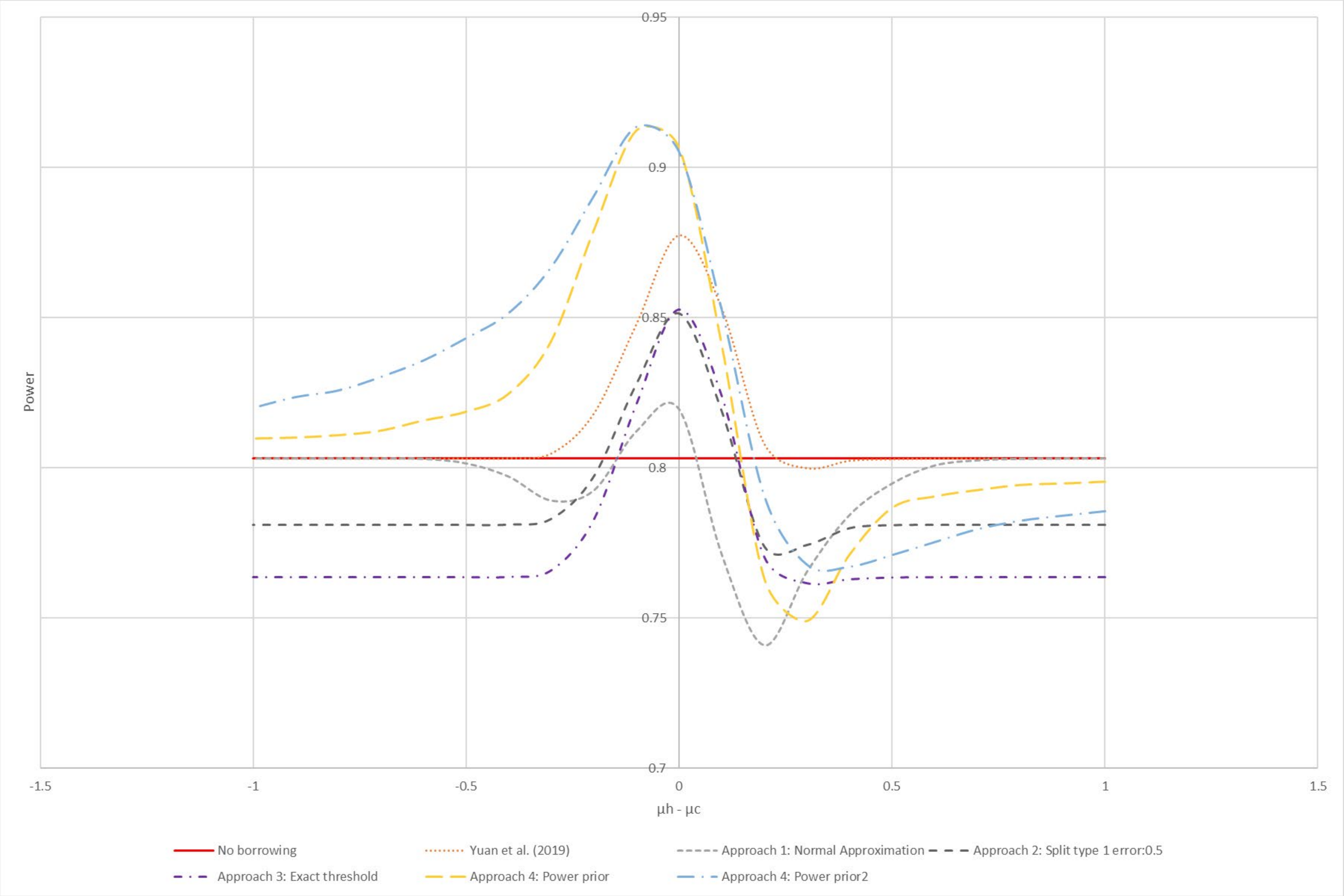


A simulation study

- Yuan (2019) has inflated type I error even when $\delta = 0$.
- Three frequentist approaches maintain type I error $\leq$ 0.05 when $\delta = 0$.
- The type I inflated when δ is in or near the equivalent margin for the proposed frequentist approaches
- Bayesian power prior has inflated type I error even when $\delta = 0$
- Bayesian power prior has the largest absolution type I error inflation and has the longest range of inflation.

A simulation study

Power when Δ=0.4



A simulation study

- Statistical power is highest near $\delta = 0$, except Bayesian approaches.
- The power gain is relatively small under the simulation condition for approach 1.
- Frequentist approaches may lose power when $\delta \neq 0$ due to lower power when no borrowing.
- Among frequentist approaches, splitting approach has the highest power.
- Bayesian prior has the highest power gain but also inflated type I error rate.

Conclusion and Discussion

- Using RWD in 2-step hybrid trials is feasible but need to aware potential bias in both patient characteristics and out exchangeability.
- Even we establish statistical exchangeability, there is still a chance for bias due to the type I error of equivalence test.
- New statistical approaches are proposed to reduce and control for the potential bias. However, such bias cannot be completely eliminated.

Acknowledgement

- This work is funded by the UCB-Stanford Digital Health Research Collaborative.

  UCB Pharma is the primary funder.

Stanford University	UMC Utrecht	ucb
Ying Lu	Ruben van Eijk	Maria Sarno
Tze Lai	Leonard van den Berg	Alicia Ellis
Lu Tian	Kit Roes	Marc van Dijk
Lorene Nelson		Anna Scowcroft
Jiapeng Xu		Jesus Garcia-Segovia

Thank you!

Questions?

Nature-Inspired Meta-heuristics for Designing Novel Drug Studies

Weng Kee Wong
Department of Biostatistics
Fielding School of Public Health
University of California at Los Angeles

SCT 45th Annual Meeting, Boston, USA
May 19 – 22 2024

May 21, 2024

Outline

- 1 Motivation
- 2 Nature-inspired Meta-heurstic Algorithms
- 3 Sample Applications for Novel Drug Studies:
  - 3.1 Dose-finding Adaptive Optimal Designs for the Continuation-Ratio (CR) Models;
  - 3.2 Extending Simon's Two-stage Response-Adaptive Designs for Phase II Trials to Multiple Stages
- 4 Summary

1.1 A Motivating Example: Locally D-optimal Approximate Designs for the Logistic Model on $X = [-1, 1]$ (Ford, 1976)

$$log\ \frac{\pi(x)}{1-\pi(x)} = \theta_1 + \theta_2 x, \quad \theta^T = (\theta_1, \theta_2), \quad \theta_1 > 0 \ \ \& \ \ \theta_2 > 0.$$

1.1 A Motivating Example: Locally D-optimal Approximate Designs for the Logistic Model on $X = [-1, 1]$ (Ford, 1976)

$$log \ \frac{\pi(x)}{1-\pi(x)} = \theta_1 + \theta_2 x, \quad \theta^T = (\theta_1, \theta_2), \ \ \theta_1 > 0 \ \ \& \ \ \theta_2 > 0.$$

- Let a solve $exp(z) = (z+1)/(z-1)$ and let u^* solve
$$exp(\theta_1 + \theta_2 u) = \frac{2 + (u+1)\theta_2}{-2 + (u+1)\theta_2}.$$

1.1 A Motivating Example: Locally D-optimal Approximate Designs for the Logistic Model on $X = [-1, 1]$ (Ford, 1976)

$$log \frac{\pi(x)}{1-\pi(x)} = \theta_1 + \theta_2 x, \quad \theta^T = (\theta_1, \theta_2), \quad \theta_1 > 0 \;\; \& \;\; \theta_2 > 0.$$

- Let a solve $exp(z) = (z+1)/(z-1)$ and let u^* solve

$$exp(\theta_1 + \theta_2 u) = \frac{2 + (u+1)\theta_2}{-2 + (u+1)\theta_2}.$$

-

condition	locally D-optimal design
$\{\theta : \theta_2 - \theta_1 \geq a\}$	$\{\frac{a-\theta_1}{\theta_2}, \frac{-a-\theta_1}{\theta_2}; \frac{1}{2}, \frac{1}{2}\}$
$\{\theta : \theta_2 - \theta_1 < a, exp(\theta_1 + \theta_2) \leq \frac{\theta_2+1}{\theta_2-1}\}$	$\{-1, u^*; \frac{1}{2}, \frac{1}{2}\}$
$\{\theta : exp(\theta_1 + \theta_2) > \frac{\theta_2+1}{\theta_2-1}\}$	$\{-1, 1; \frac{1}{2}, \frac{1}{2}\}$

1.1 A Motivating Example: Locally D-optimal Approximate Designs for the Logistic Model on $X = [-1, 1]$ (Ford, 1976)

$$log \ \frac{\pi(x)}{1-\pi(x)} = \theta_1 + \theta_2 x, \quad \theta^T = (\theta_1, \theta_2), \ \ \theta_1 > 0 \ \ \& \ \ \theta_2 > 0.$$

- Let a solve $exp(z) = (z+1)/(z-1)$ and let u^* solve

$$exp(\theta_1 + \theta_2 u) = \frac{2 + (u+1)\theta_2}{-2 + (u+1)\theta_2}.$$

-

condition	locally D-optimal design
$\{\theta : \theta_2 - \theta_1 \geq a\}$	$\{\frac{a-\theta_1}{\theta_2}, \frac{-a-\theta_1}{\theta_2}; \frac{1}{2}, \frac{1}{2}\}$
$\{\theta : \theta_2 - \theta_1 < a, exp(\theta_1 + \theta_2) \leq \frac{\theta_2+1}{\theta_2-1}\}$	$\{-1, u^*; \frac{1}{2}, \frac{1}{2}\}$
$\{\theta : exp(\theta_1 + \theta_2) > \frac{\theta_2+1}{\theta_2-1}\}$	$\{-1, 1; \frac{1}{2}, \frac{1}{2}\}$

- Corrected results in Sebastiani and Settimi (JSPI, 1997)

1.2 Need for Efficient Algorithms

- Derivation of optimal designs for nonlinear models is usually tedious, difficult and method for one model invariably does not generalize to another;
- Formulae for optimal designs rarely exist and if they do, they are complicated and frequently unhelpful to the practitioners;
- Algorithms are very helpful - available only for finding some types of optimal designs;
- Criteria of good algorithms: Proof of convergence, speed, ease of use and availability of software/codes;
- Is there an easy-to-use and efficient method for finding optimal designs for various types of optimal designs for different types of models including those with multiple interacting factors?

1.2 Need for Efficient Algorithms

- Derivation of optimal designs for nonlinear models is usually tedious, difficult and method for one model invariably does not generalize to another;
- Formulae for optimal designs rarely exist and if they do, they are complicated and frequently unhelpful to the practitioners;
- Algorithms are very helpful - available only for finding some types of optimal designs;
- Criteria of good algorithms: Proof of convergence, speed, ease of use and availability of software/codes;
- Is there an easy-to-use and efficient method for finding optimal designs for various types of optimal designs for different types of models including those with multiple interacting factors?
- Are there effective general purpose optimization tools for solving any type of optimization problems without requiring technical assumptions???

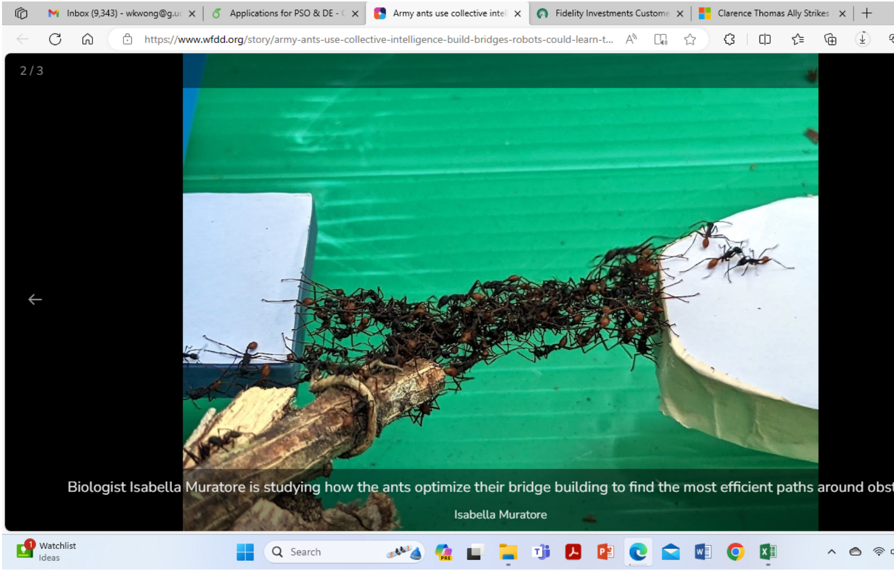
From Wikipedia, the free encyclopedia: Meta-heuristic

In computer science, meta-heuristic designates a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. Meta-heuristics make few or no assumptions about the problem being optimized and can search very large spaces of candidate solutions. However, meta-heuristics do not guarantee an optimal solution is ever found. Many have stochastic components in them (to get the algorithm out of a local optimum) and they have tuning parameters (that user may have to input);

- Perhaps Simulated Annealing and Genetic Algorithms are most familiar to statisticians, but there are many modern ones;
- They are widely used in computer science and engineering, AI but seem under-utilized in statistical research.

2.1 Our interest is nature-inspired meta-heuristic algorithms

2.2 Exploiting Animal Instincts - Collective Power



2.3 Some Applications of PSO proposed by Eberhard and Kennedy (IEEE, 1995)

- artificial neural network training
- K-means cluster analysis mathematical finance
- social networks
- data mining
- foraging techniques
- intrusion detection
- resources allocation problems
- course+exam scheduling in real time
- designing ideotypes for sustainable product systems in genetics
- prediction of stock market indices using hybrid genetic algorithm and PSO with a perturbed term
- bioinformatics (Cui, Li, Song & Wong, Bioinformatics, 2022)
- reactive power and voltage control in electric power systems
- COVID19 prevention and control, monitoring and prediction

2.4 Main Features of meta-heuristics:

- Random generation of an initial population
- Each particle has a fitness value (design criterion value or objective function value);
- The population moves or reproduces itself based on their fitness values; the former is swarm based and the latter is evolutionary;
- An exemplary swarm based algorithm is Particle Swarm Optimization (PSO) and an exemplary evolutionary algorithm is Differential Evolution (DE). Both are very widely used.
- If user-specified requirements are met, stop; otherwise each particle updates its fitness value and iteratively searches for the optimum;
- They are general purpose optimization algorithms,virtually assumptions free, and they search by exploring and exploiting the domain based on animal instincts or behavior.

2.5 Basic Equations and Tuning Parameters in PSO

$$\mathrm{v_{i+1}} = \omega_\mathrm{i}\mathrm{v_i} + \mathrm{c_1}\beta_1(\mathrm{p_i} - \mathrm{x_i}) + \mathrm{c_2}\beta_2(\mathrm{p_g} - \mathrm{x_i}), \tag{1}$$

$$\mathrm{x_{i+1}} = \mathrm{x_i} + \mathrm{v_i}. \tag{2}$$

- x_i and v_i: position and velocity for the i^{th} particle β_1 and β_2: random vectors
- ω_i: inertia weight that modulates the influence of the former velocity
- c_1 and c_2: cognitive learning parameter and social learning parameter
- p_i and p_g: Best position for the i^{th} particle (local optimal) and for all particles (global optimal)

2.6 Locally c-optimal Approximate Designs for a Nonlinear Model

Given a nonlinear model with mean function $f(x, \theta)$, we want to find a design to estimate a nonlinear function $g(\theta)$.
$\theta^T = (\theta_0, \theta_1, \theta_2)$, the sought design ξ^* minimizes

$$c^T(\theta) M(\xi, \theta)^{-1} c(\theta)$$

over all designs ξ on the dose interval X, where

$$c(\theta) = \nabla g(\theta) = \left(\frac{\partial g(\theta)}{\partial \theta_0}, \frac{\partial g(\theta)}{\partial \theta_1}, \frac{\partial g(\theta)}{\partial \theta_2}\right)^T$$

and $M(\xi, \theta)$ is the information matrix from design ξ. Assume nominal values for θ are available. Then, an approximate design ξ^* is c-optimal if and only if

$$\{f^T(x, \theta) M(\xi^*, \theta)^{-1} c(\theta)\}^2 - c(\theta)^T M(\xi^*, \theta)^{-1} c(\theta) \leq \ 0 \ \ \forall x \in X,$$

with equality at the optimal doses of ξ^* (Berger and Wong, Intro. to Optimal Designs, 2016). We apply this result to find Biological Optimal Dose (BOD) or Most Successful Dose (MSD).

2.7 Recent Applications of Meta-heuristics in statistics

1 Xu, Tan & Wong (2019). Finding High-Dimensional D-Optimal Designs for Logistic Models via Differential Evolution. IEEE Access.
2 Liu, Zhang, Yue & Wong (2021). G-optimal Designs for Hierarchical Linear Models: an Equivalence Theorem and a Nature-inspired Meta-heuristic Algorithm. Soft Computing
3 Kim & Wong (2021). Meta-heuristics for Pharmacometrics. Pharmacometrics and Systems Pharmacology.
4 Chen, Chen & Wong (2023). Particle Swarm Optimization for Finding Efficient Longitudinal Optima Exact Designs for Nonlinear Models. NEJSDS
5 Stokes, Wong & Xu (2023). Meta-heuristic Solutions to Order-of-Addition Design Problems. JCGS.
6 Schepps & Wong (2024). Optimizing Patient Enrollment in Global Clinical Trials by Meta-heuristics. Stat. in Biopharmaceutical Research.
7 Cui, Li, Song & Wong (2024). Applications of Nature-Inspired Meta-heuristic Algorithms for Tackling Optimization Problems Across Disciplines. Scientific Reports. (Concerns non-design problems and finds M-estimates for Cox regression model, estimates for Rasch model, matrix completion, variables selection in ecology, etc.)

3 Sample Applications of PSO to find Novel Drug Trials

- 1. Estimate the Biologically Optimal Dose (BOD) for a Continuation Ratio (CR) model
  - Fan and Chaloner (J. Stat. Plan. Inference, 2004)
  - Qiu and Wong (New Eng. J. of Stat. & Data. Sc., 2023).
- 2. Extend Simon Two-Stage Designs for a Phase II trial to Multiple Stages (Simon, Controlled Clinical Trials, 1989)

3.1.1 Find Biological Optimal Dose (BOD) for an Early Phase Clinical Trial

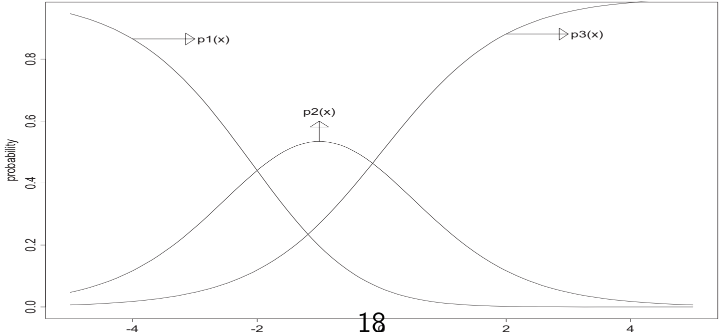
The Continuation Ratio Model relates probabilities of no response (p_1),efficacy and no severe toxicity (p_2) and severe toxicity (p_3) by:

$$\ln[p_3(\theta, x)/(1 - p_3(\theta, x))] = a_1 + b_1 x, \quad b_1 > 0 \quad (3)$$

and

$$\ln[p_2(\theta, x)/p_1(\theta, x)] = a_2 + b_2 x, \quad b_2 > 0, \quad (4)$$

where $\theta^T = (a_1, b_1, a_2, b_2)$.



3.1.2 Calculus (Fan & Chaloner, JSPI, 2003)

The biologically optimal dose x_{BOD} depends on $\theta^T = (a_1, b_1, a_2, b_2)$ and solves

$$g(x, \theta) = b_2(1 + e^{-a_1 - b_1 x}) - b_1(1 + e^{a_2 + b_2 x}) = 0.$$

3.1.2 Calculus (Fan & Chaloner, JSPI, 2003)

The biologically optimal dose x_{BOD} depends on $\theta^T = (a_1, b_1, a_2, b_2)$ and solves

$$g(x, \theta) = b_2(1 + e^{-a_1 - b_1 x}) - b_1(1 + e^{a_2 + b_2 x}) = 0.$$

- By the implicit function theorem, the gradient of the solution to the above equation is

$$\left[\frac{\partial g(x_{BOD}(\theta), \theta)}{\partial x}\right]^{-1} \frac{\partial g(x_{BOD}(\theta), \theta)}{\partial \theta}$$

$$= \begin{pmatrix} e^{-a_1 - b_1 x_{BOD}} / [b_1(e^{-a_1 - b_1 x_{BOD}} + e^{a_2 + b_2 x_{BOD}})] \\ x_{BOD} e^{-a_1 - b_1 x_{BOD}} / [b_1(e^{-a_1 - b_1 x_{BOD}} + e^{a_2 + b_2 x_{BOD}})] \\ e^{a_2 + b_2 x_{BOD}} / [b_2(e^{-a_1 - b_1 x_{BOD}} + e^{a_2 + b_2 x_{BOD}})] \\ x_{BOD} e^{a_2 + b_2 x_{BOD}} / [b_2(e^{-a_1 - b_1 x_{BOD}} + e^{a_2 + b_2 x_{BOD}})] \end{pmatrix}.$$

3.1.2 Calculus (Fan & Chaloner, JSPI, 2003)

The biologically optimal dose x_{BOD} depends on $\theta^T = (a_1, b_1, a_2, b_2)$ and solves

$$g(x, \theta) = b_2(1 + e^{-a_1 - b_1 x}) - b_1(1 + e^{a_2 + b_2 x}) = 0.$$

- By the implicit function theorem, the gradient of the solution to the above equation is

$$\left[\frac{\partial g(x_{BOD}(\theta), \theta)}{\partial x}\right]^{-1} \frac{\partial g(x_{BOD}(\theta), \theta)}{\partial \theta}$$
$$= \begin{pmatrix} e^{-a_1 - b_1 x_{BOD}}/[b_1(e^{-a_1 - b_1 x_{BOD}} + e^{a_2 + b_2 x_{BOD}})] \\ x_{BOD} e^{-a_1 - b_1 x_{BOD}}/[b_1(e^{-a_1 - b_1 x_{BOD}} + e^{a_2 + b_2 x_{BOD}})] \\ e^{a_2 + b_2 x_{BOD}}/[b_2(e^{-a_1 - b_1 x_{BOD}} + e^{a_2 + b_2 x_{BOD}})] \\ x_{BOD} e^{a_2 + b_2 x_{BOD}}/[b_2(e^{-a_1 - b_1 x_{BOD}} + e^{a_2 + b_2 x_{BOD}})] \end{pmatrix}.$$

- Use standard algorithm to generate a locally c-optimal design

3.1.3 BOD- & D-optimal designs and BOD-efficiencies

dose	weight	(a_1, b_1, a_2, b_2)	dose	weight	BOD-efficiency
-5.67	0.001	$(-3.3, 0.5, 3.4, 1)$	-4.63	0.292	56%
-0.64	0.800		-1.32	0.416	
4.84	0.199		4.19	0.056	
			8.64	0.236	
-1.26	0.632	$(-1, 0.5, 2, 1)$	-3.54	0.366	67%
4.11	0.368		-0.59	0.403	
			4.80	0.231	
-1.30	0.549	$(-1.04, 0.81.2, 1)$	-2.67	0.370	77%
2.37	0.451		0.00	0.398	
			2.88	0.232	
-14.00	0.100	$(0.4, 0.2, 2, 1)$	-13.00	0.070	62%
-1.14	0.628		-4.11	0.400	
9.99	0.272		-0.77	0.372	
			9.08	0.158	

3.1.4 Optimal Adaptive Designs for Finding BOD

- Above design strategy is not response adaptive;
- Optimal design ideas and research in adaptive designs are helpful since many latter designs do not incorporate optimal design techniques;
- For example, adaptive design ideas from the below paper for the CR model can be integrated into a new group sequential response-adaptive strategy for finding a BOD via meta-heuristics (Qiu & Wong, NEJSDS, 2023).
- Reference: Alam, et al. (Stat. in Medicine, 2019) Combined criteria for dose optimisation in early phase clinical trials using a CR model.

3.1.4 Optimal Adaptive Designs for Finding BOD

- Above design strategy is not response adaptive;
- Optimal design ideas and research in adaptive designs are helpful since many latter designs do not incorporate optimal design techniques;
- For example, adaptive design ideas from the below paper for the CR model can be integrated into a new group sequential response-adaptive strategy for finding a BOD via meta-heuristics (Qiu & Wong, NEJSDS, 2023).
- Reference: Alam, et al. (Stat. in Medicine, 2019) Combined criteria for dose optimisation in early phase clinical trials using a CR model.
- Current work is underway with Ryeznik (Uppsala University) and Sverdlov (Novartis) to implement an adaptive group sequential design to estimate BOD that also controls for the maximum toxicity level.

3.2.1 The Simon 2-Stage design for Phase II Trials

In Simon 2-Stage design for Phase II trials, user first selects two response rates of interest, p_0 and p_1 with $p_0 < p_1$.

- The hypothesis: $H_O : p \leq p_0$ versus $H_1 : p > p_1$
- Determine 4 positive integers subject to type 1 and type 2 error constraints:

3.2.1 The Simon 2-Stage design for Phase II Trials

In Simon 2-Stage design for Phase II trials, user first selects two response rates of interest, p_0 and p_1 with $p_0 < p_1$.

- The hypothesis: $H_O : p \leq p_0$ versus $H_1 : p > p_1$
- Determine 4 positive integers subject to type 1 and type 2 error constraints:
  - number of patients in Stage 1
  - number of responders in Stage 1
  - number of (additional) patients required in Stage 2
  - number of (additional) responders in Stage 2

In Simon 2-Stage design for Phase II trials, user first selects two response rates of interest, p_0 and p_1 with $p_0 < p_1$.

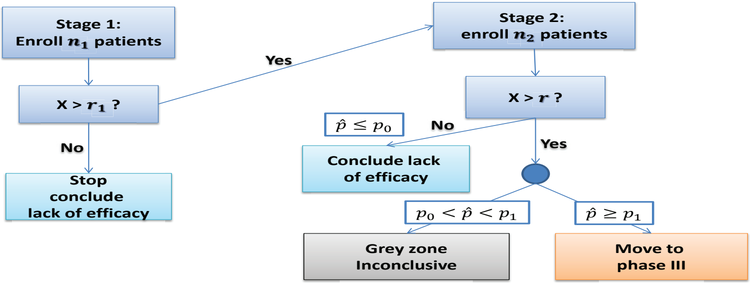
- The hypothesis: $H_O : p \leq p_0$ versus $H_1 : p > p_1$
- Determine 4 positive integers subject to type 1 and type 2 error constraints:
  - number of patients in Stage 1
  - number of responders in Stage 1
  - number of (additional) patients required in Stage 2
  - number of (additional) responders in Stage 2
- Problems: (Simon, Controlled Clinical Trials, 1989): Given Type I and II error constraints, the objective is to
  (i) minimize the expected sample size under H_0, or
  (ii) minimize the maximum sample size required for the trial.

3.2.2 Simon 2-stage Designs for Phase 2 Trials

▶

Simon's Two-Stage Designs

- X: the number of responders

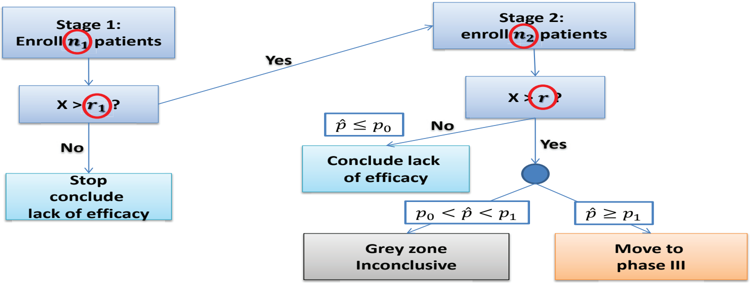


3.2.3 Simon's 2-stage Adaptive Designs for Phase 2 Trials

Review of Simon's Design (Controlled Clinical Trials, 1989)

Simon's Two-Stage Designs

- X: the number of responders



3.2.4 Various adaptive 2-stage optimal designs with 1 target response when $\alpha = 0.05$ and $\beta = 0.20$.

p_0	p_1	Optimal criteria	Method	s_1/n_1	s/n	$1-\alpha$	β	$E(N\|p_0)$	$E(N\|p_1)$	CPU time (mins)
0.05	0.20	C1	GS	0/10	3/29	0.953	0.199	17.624	26.960	0.09
			G-DPSO	1/15	4/40	0.968	0.197	19.274	35.822	2.51
		C2	GS	0/11	3/28	0.956	0.199	18.330	26.540	0.1
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540	2.32
0.20	0.35	C1	GS	5/22	19/72	0.951	0.200	35.368	63.855	15.15
			G-DPSO	5/22	19/72	0.951	0.200	35.368	63.855	2.87
		C2	GS	3/21	15/53	0.950	0.200	41.148	51.941	14.25
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941	2.66

C1 and C2 are Simon's original optimality criteria

3.2.4 Various adaptive 2-stage optimal designs with 1 target response when $\alpha = 0.05$ and $\beta = 0.20$.

p_0	p_1	Optimal criteria	Method	s_1/n_1	s/n	$1-\alpha$	β	$E(N\|p_0)$	$E(N\|p_1)$	CPU time (mins)
0.05	0.20	C1	GS	0/10	3/29	0.953	0.199	17.624	26.960	0.09
			G-DPSO	1/15	4/40	0.968	0.197	19.274	35.822	2.51
		C2	GS	0/11	3/28	0.956	0.199	18.330	26.540	0.1
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540	2.32
0.20	0.35	C1	GS	5/22	19/72	0.951	0.200	35.368	63.855	15.15
			G-DPSO	5/22	19/72	0.951	0.200	35.368	63.855	2.87
		C2	GS	3/21	15/53	0.950	0.200	41.148	51.941	14.25
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941	2.66

C1 and C2 are Simon's original optimality criteria

- Uncertainty in the choice of p_1 can be an issue.

3.2.4 Various adaptive 2-stage optimal designs with 1 target response when $\alpha = 0.05$ and $\beta = 0.20$.

p_0	p_1	Optimal criteria	Method	s_1/n_1	s/n	$1-\alpha$	β	$E(N\|p_0)$	$E(N\|p_1)$	CPU time (mins)
0.05	0.20	C1	GS	0/10	3/29	0.953	0.199	17.624	26.960	0.09
			G-DPSO	1/15	4/40	0.968	0.197	19.274	35.822	2.51
		C2	GS	0/11	3/28	0.956	0.199	18.330	26.540	0.1
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540	2.32
0.20	0.35	C1	GS	5/22	19/72	0.951	0.200	35.368	63.855	15.15
			G-DPSO	5/22	19/72	0.951	0.200	35.368	63.855	2.87
		C2	GS	3/21	15/53	0.950	0.200	41.148	51.941	14.25
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941	2.66

C1 and C2 are Simon's original optimality criteria

- Uncertainty in the choice of p_1 can be an issue.
- Lin and Shih (Biometrics, 2004) allowed 2 alternative hypotheses and searched using a greedy search.

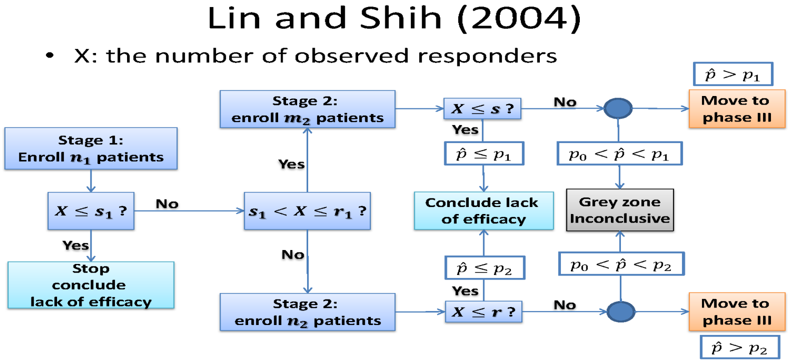
3.2.4 Various adaptive 2-stage optimal designs with 1 target response when $\alpha = 0.05$ and $\beta = 0.20$.

p_0	p_1	Optimal criteria	Method	s_1/n_1	s/n	$1-\alpha$	β	$E(N\|p_0)$	$E(N\|p_1)$	CPU time (mins)
0.05	0.20	C1	GS	0/10	3/29	0.953	0.199	17.624	26.960	0.09
			G-DPSO	1/15	4/40	0.968	0.197	19.274	35.822	2.51
		C2	GS	0/11	3/28	0.956	0.199	18.330	26.540	0.1
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540	2.32
0.20	0.35	C1	GS	5/22	19/72	0.951	0.200	35.368	63.855	15.15
			G-DPSO	5/22	19/72	0.951	0.200	35.368	63.855	2.87
		C2	GS	3/21	15/53	0.950	0.200	41.148	51.941	14.25
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941	2.66

C1 and C2 are Simon's original optimality criteria

- Uncertainty in the choice of p_1 can be an issue.
- Lin and Shih (Biometrics, 2004) allowed 2 alternative hypotheses and searched using a greedy search.
- Kim and Wong (Stat. Meth. in Med. Res., 2016) allowed 3 alternative hypotheses and searched using PSO.

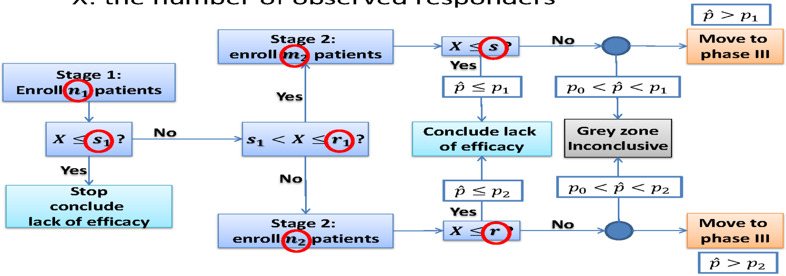
3.2.4a A Discrete Optimization Problem: Extended 2-stage Adaptive Phase II Trials with Two Targeted Alternatives (Lin and Shih, Biometrics, 2004)



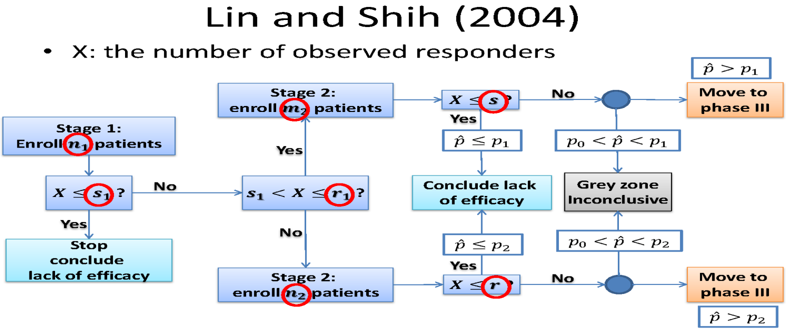
Depending on the quality of Stage 1 results, one of the 2 alternative hypotheses is tested to target p_1 more accurately.

Lin and Shih (2004)

- X: the number of observed responders



Depending on the quality of Stage 1 results, one of 2 alternative hypotheses for p_1 will be tested with possibly different sets of Type I and II constraints.



Depending on the quality of Stage 1 results, one of 2 alternative hypotheses for p_1 will be tested with possibly different sets of Type I and II constraints.

- Design problem for allowing 3 alternative hypotheses is much more complicated and was solved using a hybridized version of PSO (Kim and Wong, Stat. Meth. in Med. Res., 2016).

3.2.5 Web-based tools for Early Phase Trials

A helpful and useful website with more than 27,000 hits per year:

`https://biostatistics.mdanderson.org/SoftwareDownload/`

Many Bayesian optimal phase II designs can be generated from:

`https://trialdesign.org/`

Website for Finding Bayesian Optimal Designs for Phase II (BOP2) Trials: A specific example:

`https://trialdesign.org/one-page-shell.html#BOP2`

3.2.6 K-stage Designs under Simon's Framework

Need for more than 2-stage adaptive designs are motivated and studied in the following papers:

- 3-Stage Phase II Designs (Chen, SIM, 1997)
- 4-Stage Phase II Designs Permitting One Dose Escalation (Hanfelt, SIM, 1999)
- 4- and 5-Stage Phase II Oncology Trails (Tan and Xiong, Pharmaceutical Statistics, 2011)
- Single-Arm Multi-Stage Bayesian-Frequentist (SAMS-BF) Designs (Liu, PhD thesis, University of Kansas, 2022)

3.2.7 K-stage Designs under Simon's Framework

- For K-stage design, $K \geq 2$, the number of patients studied in each stage are $n_1, n_2, \ldots, n_K$, respectively.
- The total sample size is $N_{max} = \sum_k n_k$.
- The drug is rejected and we terminate the study if:
  - r_1 or fewer responses are observed at the first stage, or
  - $r_1 + r_2$ or fewer responses are observed at the second stage, or
  - $r_1 + r_2 + r_3$ or fewer responses are observed at the third stage, or
  - ...
  - $\sum_k r_k$ or fewer responses are observed at the Kth stage.

3.2.8 Probability of Early Termination for K-stage (Liu, 2022)

- The probability of early termination (PET) after the first stage is

$$PET_1(p) = B(r_1, n_1, p) = \sum_{x \le r_1} b(x, n_1, p)$$

- For $k = 2, \ldots, K$, the PET at the kth stage has the analytical form

$$PET_k(p) = \underbrace{\sum_{x_1=L_1}^{U_1} \sum_{x_2=L_2}^{U_2} \cdots \sum_{x_{k-1}=L_{k-1}}^{U_{k-1}}}_{k-1 \text{ Summations}} \left[\prod_{s=1}^{k-1} b\left(x_s, n_s, p\right)\right] B\left(\sum_{s=1}^{k} r_s - \sum_{s=1}^{k-1} x_s, n_k, p\right) \tag{5}$$

where

$$U_j = \min\left[n_j, r_k - \sum_{s=1}^{j-1} x_s\right] \text{ and } L_j = r_j + 1 - \sum_{s=1}^{j-1} x_s$$

- For the final stage (Kth stage) the value $PET_K(p)$ is the probability of rejecting a treatment at the end of the study.

3.2.9 Probability of Early Termination for K-stage

Under H_0, the expected sample size of the study is (Liu, 2022)

$$\begin{aligned} E(N \mid H_0) &= n_1 + (1 - PET_1(p))n_2 + (1 - PET_1(p) - PET_2(p))n_3 + \\ &\quad \cdots + (1 - PET_1(p) - PET_2(p) - \cdots PET_{K-1}(p))n_K \\ &= n_1 + \sum_{k=2}^{K} \left[1 - \sum_{s=1}^{k-1} PET_s(p) \right] n_k \end{aligned}$$

The total probability of rejecting a drug in the study is

$$P(\text{Reject the drug} \mid p) = \sum_{k=1}^{K} PET_k(p) \qquad (6)$$

3.2.10 Requirements of Finding K-stage Simon's Design

Given

- True response probability under the hypothesis $H_0 : p = p_0 = 0.2$ vs. $H_1 : p = p_1 = 0.4$
- Acceptable type I error $\alpha = 0.1$ and power $1 - \beta = 0.9$

Under Simon's frame work, find K-stage optimal and minimax design

$$r_1, n_1, r_2, n_2, \dots, r_K, n_K$$

- The maximal sample size N_{max} is chosen from 30 to 70, i.e. $N_{max} = \sum n_k \in \{30, 31, 32, ..., 69, 70\}$
- The minimal number of patients at the first stage is 10, i.e. $\min(n_1) = 10$
- For each stage, there must be at least 1 patient added into the study, i.e. $\min(\text{cohort size}) = \min(n_k) = 1$ for $k = 2, ..., K$

3.2.11 Objective Functions for K-stage Simon's Design

The objective function for Optimal K-stage design

$$\min_{\substack{N_{max} \\ n_1, n_2, \cdots, n_{K-1} \\ r_1, r_2, ..., r_k}} E(N|H_0) + M \times \max\{I(\text{t1e} > \alpha), I(\text{t2e} > \beta)\}$$

The objective function for Minimax K-stage design (Of all designs with minimal maximal sample size, find the best one with minimal expect sample size under H_0)

$$\min_{\substack{N_{max} \\ n_1, n_2, \cdots, n_{K-1} \\ r_1, r_2, ..., r_k}} \left[\frac{E(N|H_0)}{N_{max}} + N_{max}\right] + M \times \max\{I(\text{t1e} > \alpha), I(\text{t2e} > \beta)\}$$

where $M >> 0$ is the penalty on those designs not satisfying the requirements of type I error and power. Typically M can be very large, say $M = 10^8$.

3.2.12 PSO for K-stage Simon's Design

Under Simon's frame work, find K-stage optimal and minimax design

$$r_1, n_1, r_2, n_2, \ldots, r_K, n_K$$

We treat it as the $(2K+1)$-dimensional optimization problem with PSO's particle on the continuous domain

$$\text{particle} = (\Omega, \omega_1, \ldots, \omega_K, \phi_1, \ldots, \phi_K), \text{ where}$$

- $\Omega \in [30, 70]$ and $N_{max} = [\Omega]$
- $\omega_k \in [0, 1]$ satisfying $\sum \omega_k = 1$. Then the sample size at each look is $n_k = \left[\sum_{s=1}^{k} \omega_s \times N_{max}\right]$.
- $\phi_k \in [0, 1]$ and it is transferred as the cutoff of rejecting a drug by $r_k = [\phi_k \times n_k]$.

3.2.13 Numerical Results of 2-stage Simon's Design

PSO's results are consistent to Simon's results (Simon, 1989).

Algorithm	Optimal Design	E(N|H0)	min(f)	mean(f)	sd(f)	#(achieve)
Simon (1989)	3/17, 10/37	26.022	–	–	–	–
PSO	3/17, 10/37	26.022	26.022	26.022	0.000	100/100
QPSO	3/17, 10/37	26.022	26.022	26.025	0.026	99/100

Algorithm	Minimax Design	E(N|H0)	min(f)	mean(f)	sd(f)	#(achieve)
Simon (1989)	3/19, 10/36	28.263	–	–	–	–
PSO	3/19, 10/36	28.263	36.296	36.296	0.000	99/100
QPSO	3/19, 10/36	28.263	36.296	36.298	0.005	79/100

*#(achieve): times of the algorithm finding the best design over 100 replicates.

*CPU time to find the 2-stage optimal and minimax designs does not exceed 2.5 seconds. (Core(TM) i7-13700 2.10 GHz CPU, 128 GB RAM)

3.2.14 Numerical Results of 3-stage Simon's Design

Timothy Chen (1997) extend Simon's results to 3-stage. PSO can also reproduce Chen's results and outperforms the QPSO in finding minimax design.

Algorithm	Optimal Design	E(N|H0)	min(f)	mean(f)	sd(f)	#(achieve)
Chen (1997)	1/10, 6/26, 11/43	23.860	–	–	–	–
PSO	1/10, 6/26, 11/43	23.860	23.860	23.990	0.169	42/100
QPSO	1/10, 6/26, 11/43	23.860	23.860	24.011	0.191	42/100

Algorithm	Minimax Design	E(N|H0)	min(f)	mean(f)	sd(f)	#(achieve)
Chen (1997)	2/16, 5/26, 10/36	26.413	–	–	–	–
PSO	2/16, 5/26, 10/36	26.413	36.244	36.248	0.008	73/100
QPSO	2/16, 5/26, 10/36	26.413	36.244	36.256	0.011	12/100

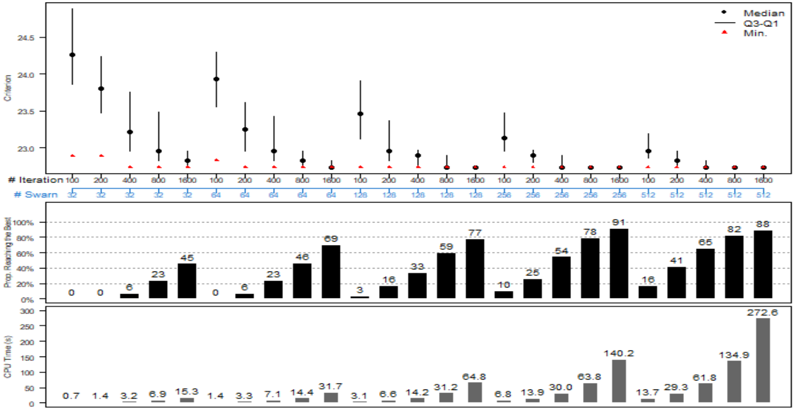
*#(achieve): times of the algorithm finding the best design over 100 replicates.

*CPU time to find the 3-stage optimal and minimax designs does not exceed 6.0 seconds. (Core(TM) i7-13700 2.10 GHz CPU, 128 GB RAM)

3.2.15 PSO Parameter Tuning for 4-Stage Optimal Design

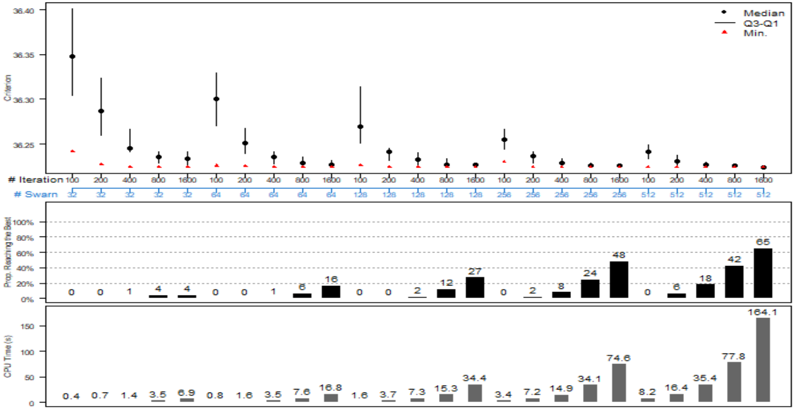
Experiment using 100 replications with different initial swarm.
To find a relatively optimal design with at least 78% chance, the recommendation of PSO parameters is:

$$\text{Swarm size} \geq 256 \text{ and Maximum iterations} \geq 800$$



3.2.16 PSO Parameter Tuning for 4-Stage Minimax Design

We find that finding minimax design is more challenged than doing for the optimal designs. In this case, we need more than 512 swarm size and more than 1600 iterations.



3.2.17 New Results of 4-stage Simon's Design by PSO

Status	4-stage Optimal Design	T1E	Power	E(N|H0)	#(achieve)
Best	1/10, 4/21, 7/30, 11/43	0.099	0.901	22.732	88/100
Top 2	2/14, 4/21, 7/30, 10/38	0.100	0.901	22.827	11/100
Top 3	1/11, 4/22, 6/26, 10/38	0.099	0.900	22.905	1/100

Status	4-stage Minimax Design	T1E	Power	E(N|H0)	#(achieve)
Best	1/13, 3/20, 6/29, 10/36	0.085	0.900	25.697	65/100
Top 2	1/13, 4/23, 7/31, 10/36	0.085	0.900	25.719	2/100
Top 3	1/14, 4/22, 6/29, 10/36	0.085	0.900	25.758	1/100
Top 4	2/17, 4/22, 6/29, 10/36	0.085	0.900	25.768	12/100
Top 5	2/16, 4/23, 7/32, 10/36	0.085	0.900	25.790	10/100
Top 6	1/14, 2/17, 5/25, 10/36	0.085	0.900	25.845	1/100
Top 7	2/16, 5/26, 6/29, 10/36	0.085	0.900	25.882	5/100
Top 8	1/14, 2/16, 5/26, 10/36	0.085	0.900	25.936	1/100
Top 9	2/16, 5/26, 8/34, 10/36	0.086	0.901	26.037	1/100
Top 10	2/16, 4/24, 5/26, 10/36	0.085	0.901	26.051	2/100

4 Summary

Other Nature-Inspired Algorithms:

- Differential evolution (1995)
- Bees algorithm (2006)
- Invasive weed optimization (2006)
- Artificial bee colony algorithm (2007)
- Monkey search (2007)
- Imperialist competitive algorithm (2007)
- Intelligent water drops algorithm (2009)
- Glowworm swarm optimization (2009)
- Cuckoo search (Yang & Deb, 2009, Journal of Mathematical Modeling and Numerical Optimization)
- Firefly algorithm (2009, 2010)
- Bat algorithm (2010)

and the list goes on and on...

4.1 Conclusions

- PSO methodology offers great promise and I believe represents a leap forward in the field of optimal experimental designs;

- Optimal designs should be more accessible now and hopefully optimal design ideas will be more widely used in practice, but not religiously;

- Nature-inspired meta-heuristic algorithms are very flexible because they are assumptions free and so they are general purpose optimization tools;

4.1 Conclusions

- PSO methodology offers great promise and I believe represents a leap forward in the field of optimal experimental designs;
- Optimal designs should be more accessible now and hopefully optimal design ideas will be more widely used in practice, but not religiously;
- Nature-inspired meta-heuristic algorithms are very flexible because they are assumptions free and so they are general purpose optimization tools;
- Nature-inspired algorithms are not problems free, and they include how to tune parameters for accelerated convergence, escape from a local optimum, lack of a rigorous proof of convergence, and there is a gargantuan/plethora of them.

4.2 Acknowledgement of collaborators:

Chen, PingYang (National Taipei University, New Taipei, Taiwan)

Chen, RayBing (National Cheng Kung University, Tainan, Taiwan)

Fang, Xinying (Penn State University at Hershey)

Lee, Jack (University of Texas - MD Anderson Cancer Center)

Ryeznik, Yevgen (Uppsala University)

Sverdlov, Alex (Novartis)

Zhou, Shouhao (Penn State University at Hershey)

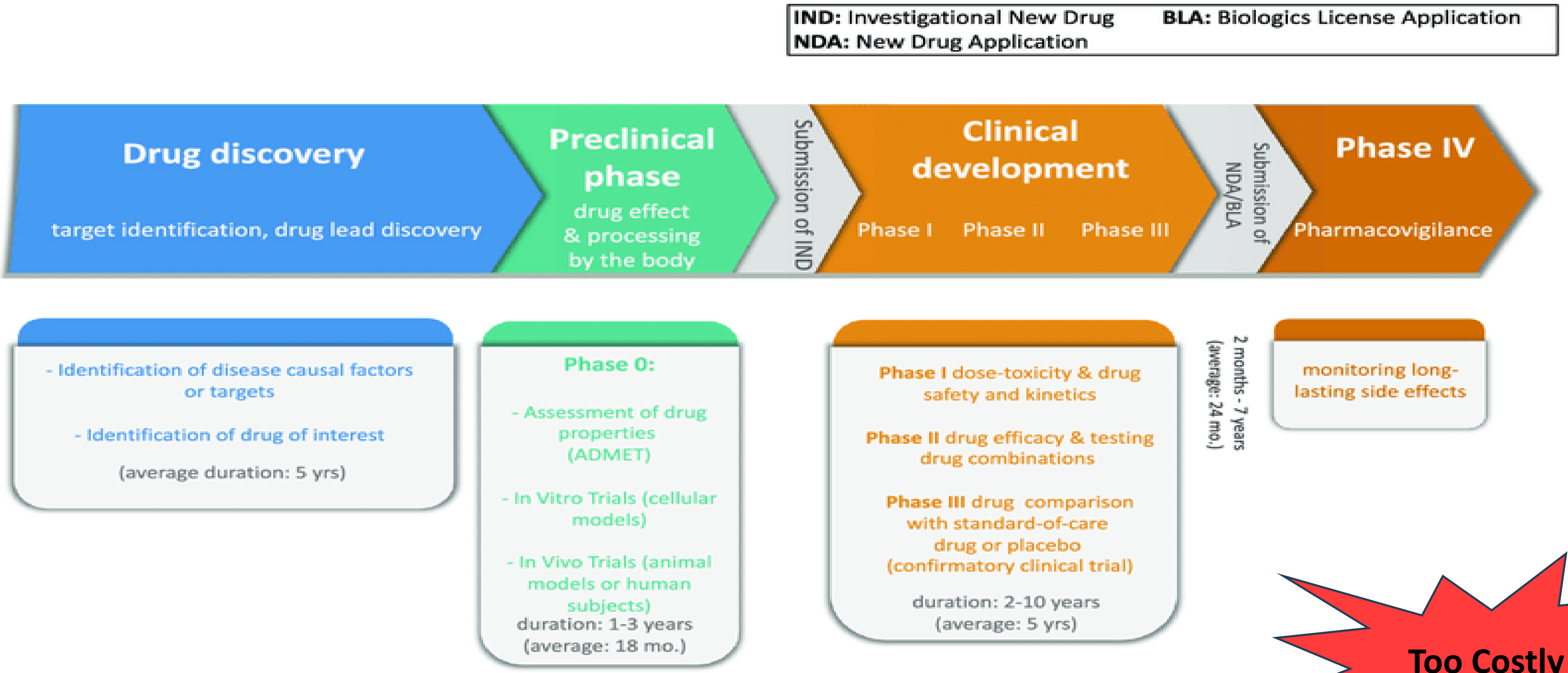
SCT 45TH ANNUAL MEETING

Boosting Bayesian Adaptive Platform Designs For Screening Safe and Efficacious Drugs via Particle Swarm Optimization

J. Jack Lee[1], Xinying Fang[2], Shouhao Zhou[2],
Ping-Yang Chen[3], Ray-Bing Chen[4], Weng Kee Wong[5]

1: University of Texas MD Anderson Cancer Center
2: Pennsylvania State University
3: National Taipei University, Taiwan
4: National Cheng-Kung University, Taiwan
5: University of California, Los Angeles (UCLA)

Drug Development Is A Costly and Arduous Process



Too Costly!

- Deloitte 2023 report found that the average cost of developing a new drug among the top 20 global biopharmas is approximately $2.3 billion.
- The time to develop a new drug is approximately 10-15 years.

Too Long!

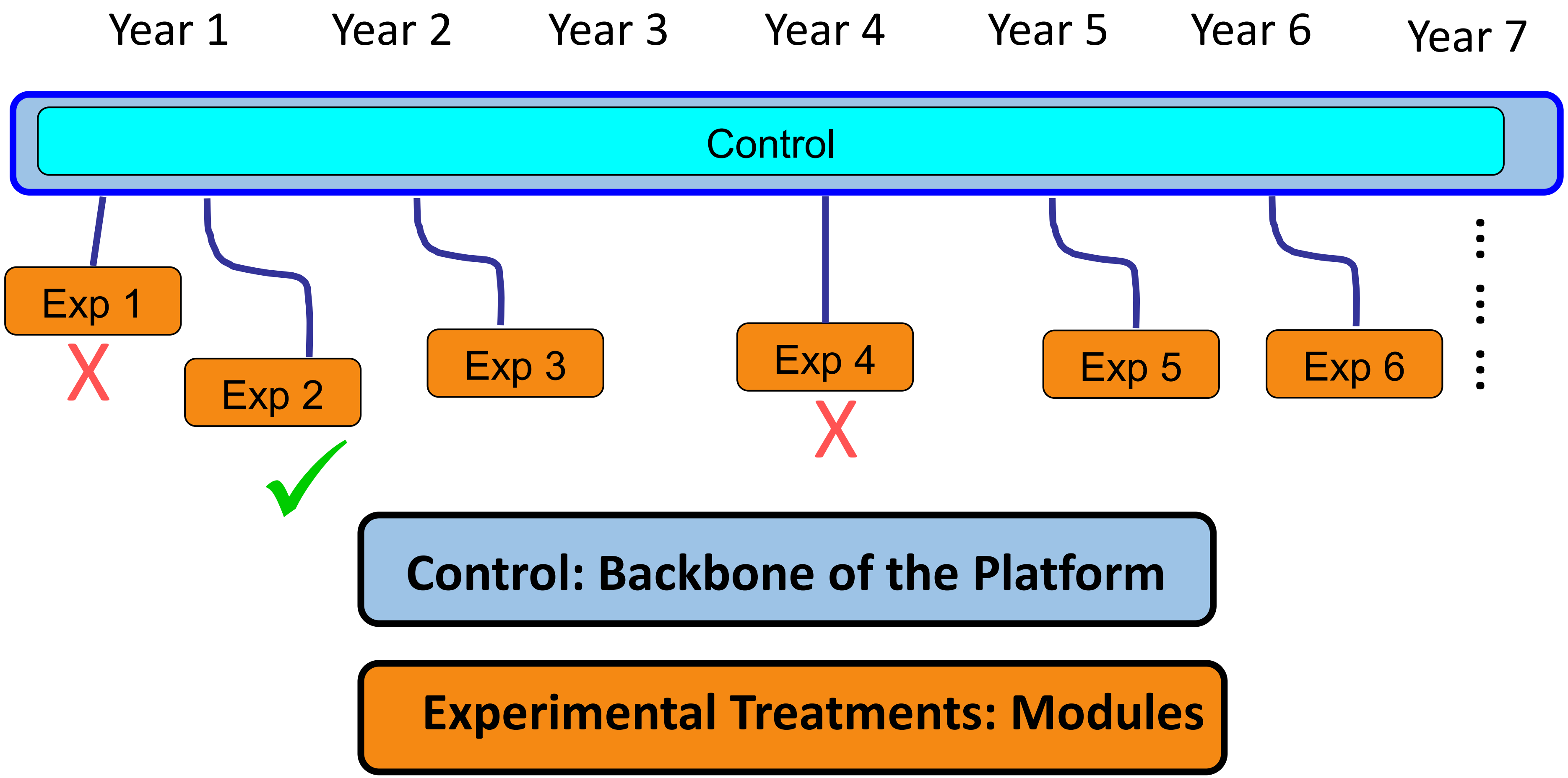
What Are Adaptive Designs?

Trials that use interim data to guide the study conduct

- Adaptive Phase I trials for dose finding and optimization
  - Continual Reassessment Method (CRM)
  - Bayesian Optimal Interval (BOIN) Design
  - Utility-based Bayesian Optimal Interval (U-BOIN) Design
- Adaptive Phase II design for assessing efficacy and/or toxicity with complex endpoints
  - BOP2: A Bayesian Optimal Phase 2 Design
- Adaptive estimation and decision making in Phase II or Phase III trials
  - Based on posterior probability, predictive probability, or probability of success
  - Adaptive decision making: Go / No-go decisions
  - Adaptive patient assignment to treatment, adaptive randomization
- Adaptive Multi-Arm, Multi-Stage (MAMS) Design
  - Platform design
  - Basket and umbrella designs, matching patients with treatments based on markers
  - Adaptive drop/graduate treatments due to toxicity, futility, and/or efficacy, add new treatments
- Utility-based decision theoretical approach
  - Optimizing decision with multi-facet problems

Bayesian 1-2-3 "Prior + Data → Posterior." It takes "We learn as we go" approach which is particularly suitable for adaptive designs.

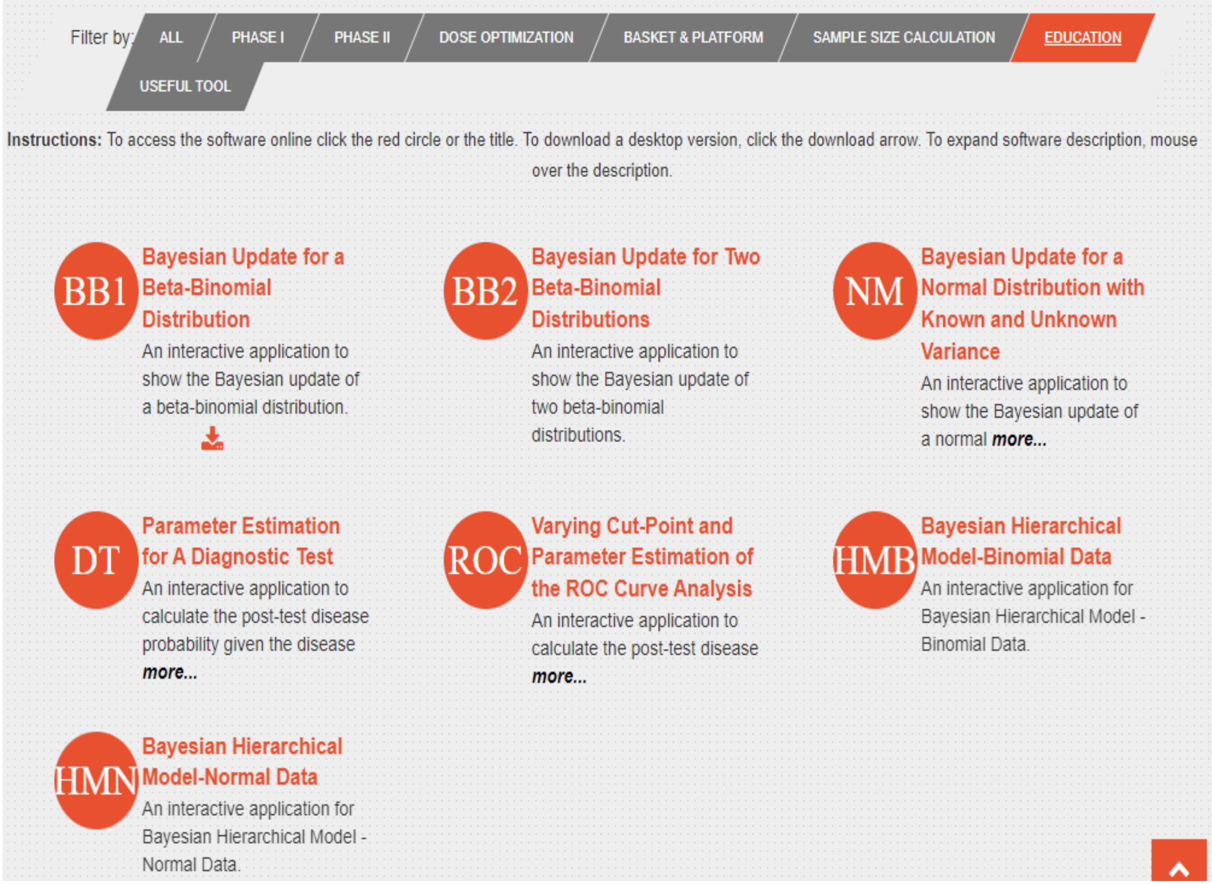
Bayesian Adaptive Platform Design

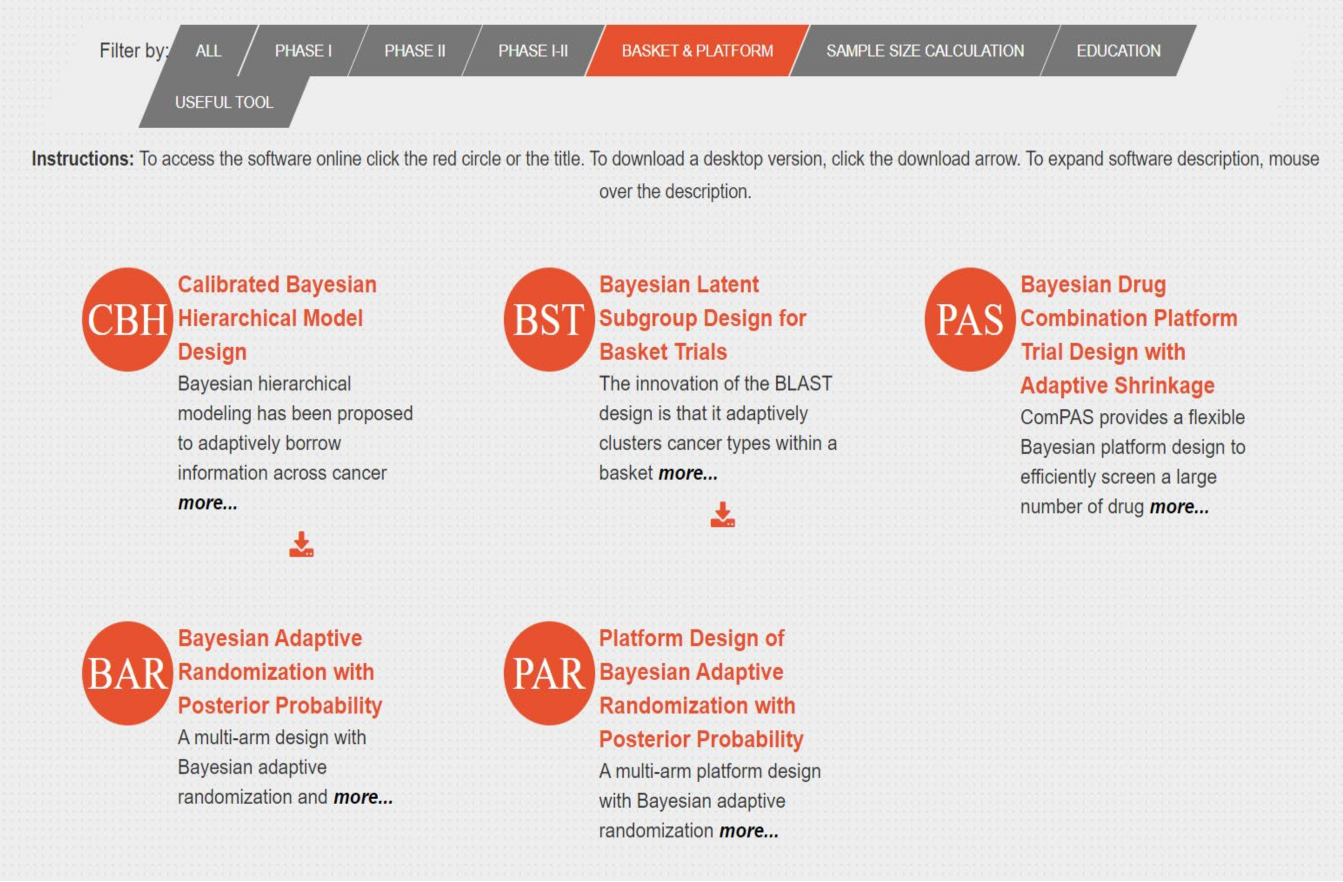


- Efficient continuous drug evaluation process
- No interruption of patient accrual by avoid white space between different trials

https://trialdesign.org/

Clinical Trials Design Software





https://trialdesign.org/

Bold Vision of Multi-arm, Multi-stage, Multi-marker Bayesian Adaptive Platform Designs

- Master Protocols with multiple treatment arms
  - Umbrella trials, Basket trials
- Biopsy required, biomarker guided
- Adaptive randomization
- Interim monitoring
  - Early stopping due to toxicity
  - Early stopping due to futility
  - Early stopping due to efficacy
  - Drop losers, graduate winners
  - Add new treatments, refine patient populations
- Seamless phase I/II, II/III, or I/II/III designs
  - Exploration, identification, and validation of winners
- Perpetual platform trials
- Efficient learning health care system
  - There is something available for every patient.
  - Provide precision medicine for every patient. Learn from every patient.

Video: Platform Trial with BAR: θ_1=0.2, θ_2=0.3, θ_3=0.4, θ_4=0.2, θ_5=0.3, θ_6=0.4

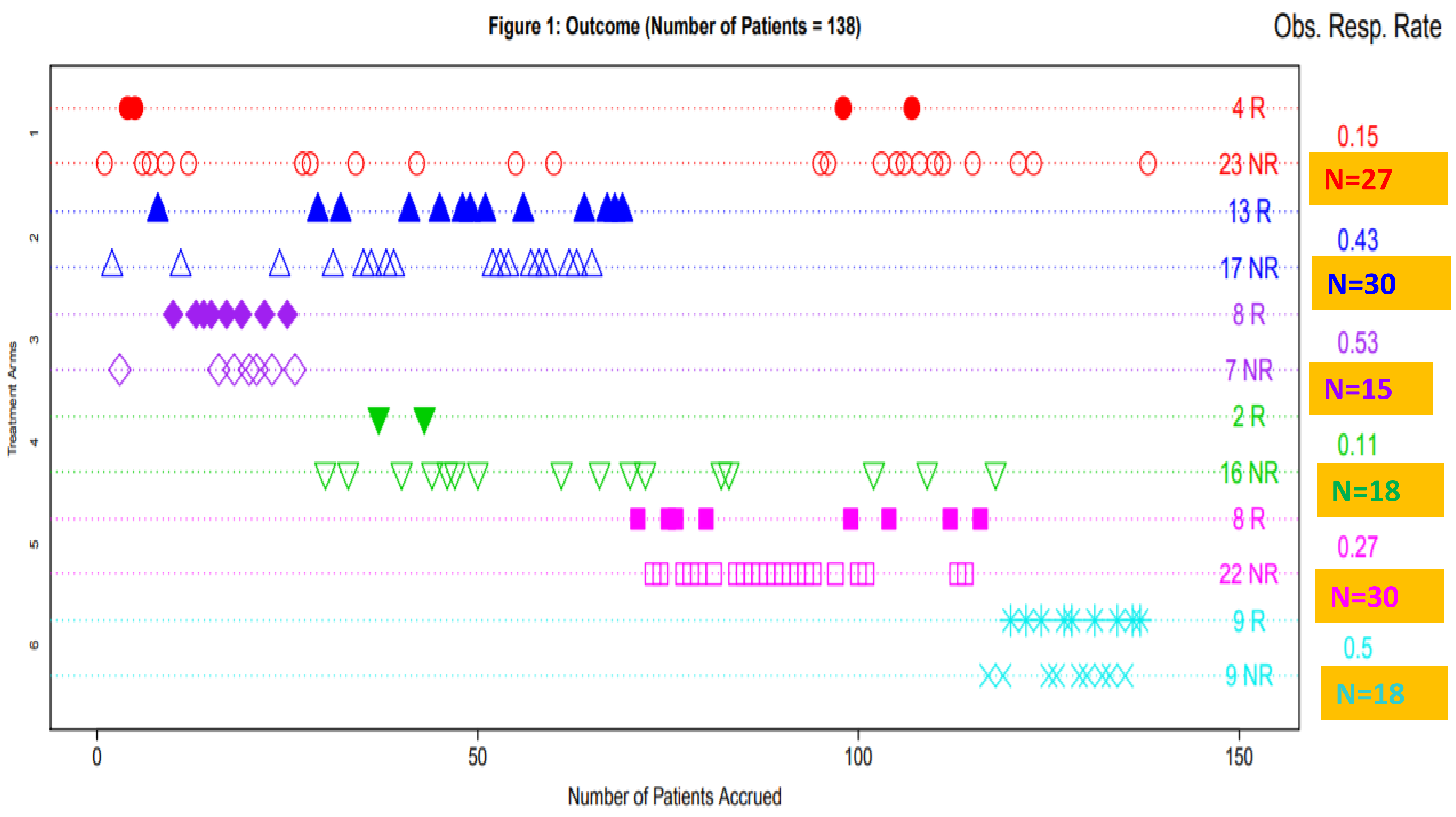


Figure 1: Outcome (Number of Patients = 138)
Obs. Resp. Rate
Treatment Arms
1
2
3
4
5
6
4 R
23 NR
13 R
17 NR
8 R
7 NR
2 R
16 NR
8 R
22 NR
9 R
9 NR
0.15
N=27
0.43
N=30
0.53
N=15
0.11
N=18
0.27
N=30
0.5
N=18
0
50
100
150
Number of Patients Accrued

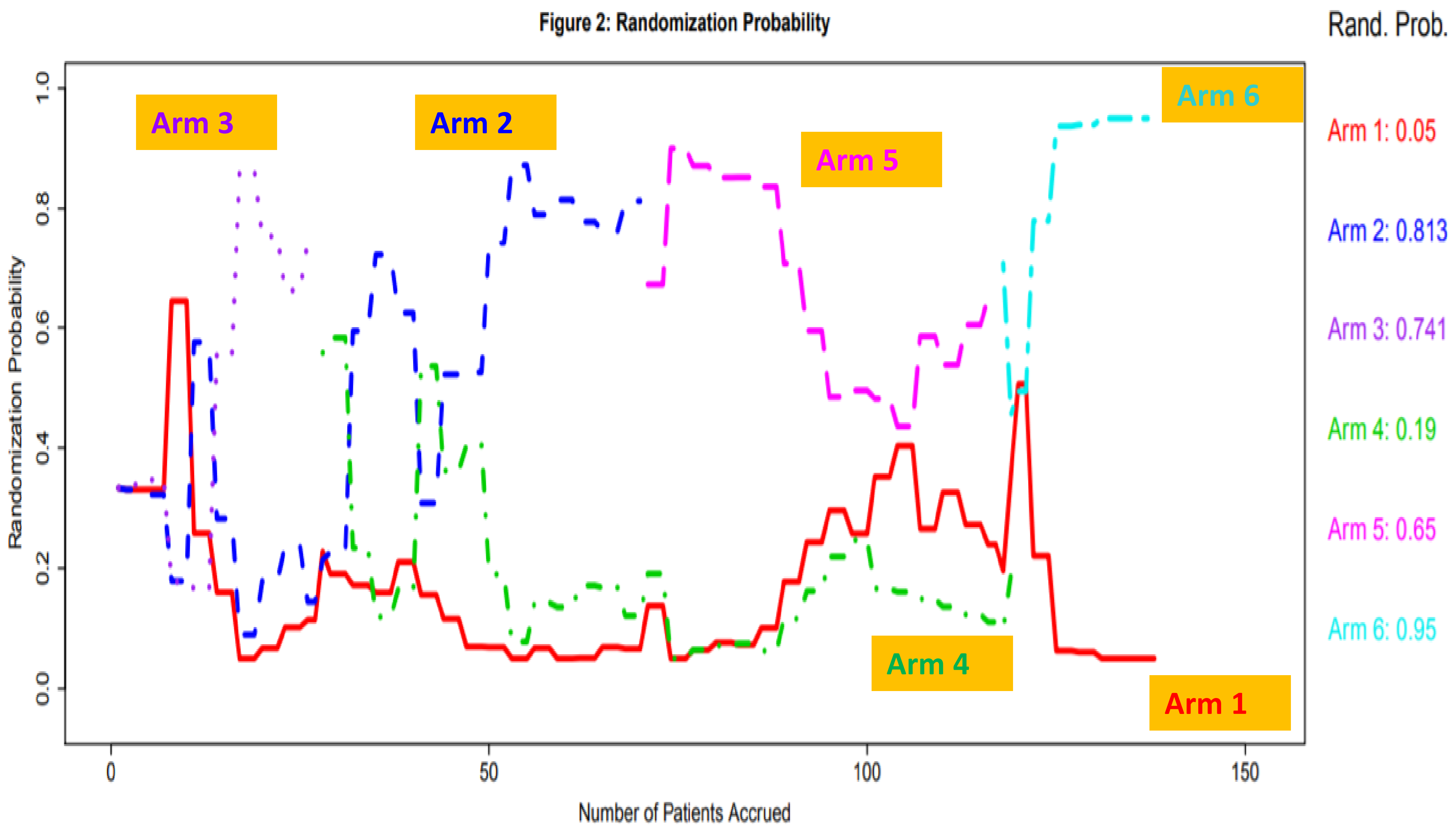
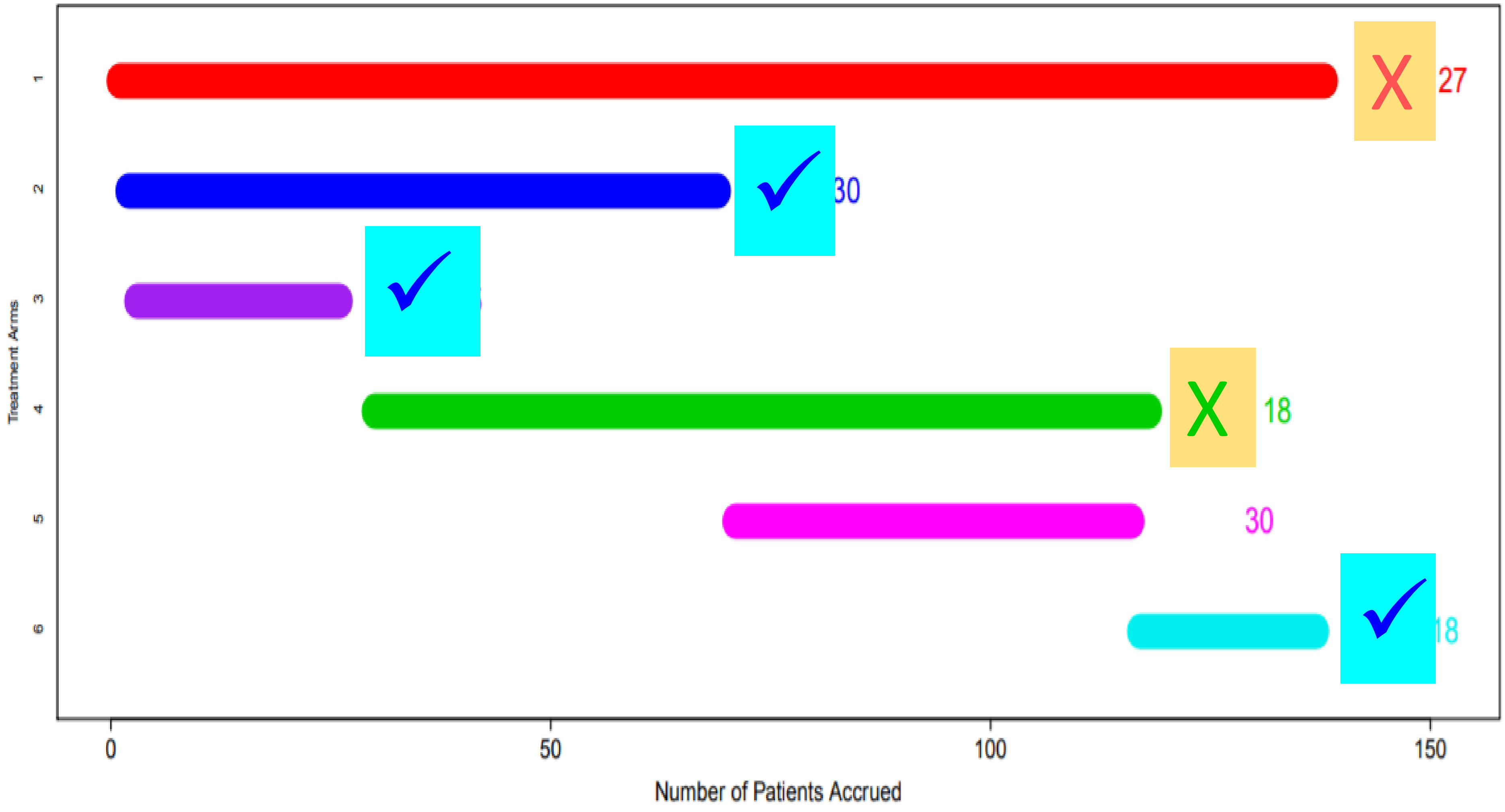


Figure 2: Randomization Probability
Rand. Prob.
Arm 1: 0.05
Arm 2: 0.813
Arm 3: 0.741
Arm 4: 0.19
Arm 5: 0.65
Arm 6: 0.95
Arm 3
Arm 2
Arm 5
Arm 6
Arm 4
Arm 1
Randomization Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
Number of Patients Accrued

Figure 3: Accrual Timeline (Number of Patients = 138)



Statistical Considerations for Platform Trials

- Can incorporate a common control arm or compare multiple arms with known references.
- What is the maximum number of patients in total? What is the accrual rate?
  - Minimum and maximum numbers of patients per arm
- How many arms can be evaluated at the same time?
- How many interim analyses and when to do them?
  - When is the 1st interim analysis? Cohort size for evaluation?
- What is the criteria for evaluation
  - Early stopping for toxicity, Early stopping for futility, Early stopping for efficacy
  - Null and alternative response rates? Probability strength for decision making?
- How to allocate patients to arms?
  - Equal randomization. Fixed ratio randomization
  - Outcome adaptive randomization
- How to evaluate the performance of a trial?
  - Control type I error and maintain statistical power?
  - How many drugs can be tested within a given timeframe?
  - What are the success rates?
    - True positive, false positive, true negative, false negative

BOP2: A Bayesian Optimal Design for Phase 2 Trials

- A unified framework for phase II trials with simple and complex efficacy and/or toxicity endpoints.
- Multiple endpoints: $Y \sim Multinomial\ (\theta_1, \theta_2, \ldots, \theta_K)$
  - $(\theta_1, \theta_2, \ldots, \theta_K) \sim\ Dir(a_1, a_2, \ldots., a_K)$
  - Given data: $\theta|D_n \sim Dir(a_1{+}x_1, a_2{+}x_2, \ldots., a_K{+}x_K)$
- Decision rule: Stop the trial if

$$Prob(b\theta \le \phi|D_n) > C(n)$$
$$C(n) = 1\ - \lambda(n/N)^{\gamma}$$

- Steps
  1. Elicit parameters under H_0, H_1
  2. Given number and sample size of interim analysis (n_1, n_2, …, n_k) and total N
  3. Find the set of (λ, γ) yielding type I error by grid search to

     (a) Optimize power, or

     (b) Minimize $E(N|H_0)$

Zhou H, Lee JJ, Yuan Y. BOP2: Bayesian optimal design for phase II clinical trials with simple and complex endpoints. Stat Med 2017.

Particle Swarm Optimization (PSO) for Multi-stage Designs

- PSO is a stochastic population-based global optimization method inspired by observing how a flock of birds move in the sky.
  - It is an evolutionary algorithm and stochastically evolves a group of particles (swarm) toward the best solution which mimicking observational behavior from nature.
- Recently, this class of algorithms has been gaining wide recognition for its ability to solve or nearly solve hard-to-optimize high dimensional problems in the real world.
- We apply PSO in multi-stage designs to find the optimal solution by varying many design parameters.
  - Design complexity increases exponentially as the number of design parameters increases linearly.
  - As the complexity increases, finding the optimal solution can quickly become unattainable using grid search or traditional optimization methods.

PSO in Bayesian Phase II Multi-Stage Designs

- Binomial Model and Hypothesis Testing:
  - $X_i \sim Binomial(n_i, \theta)$, where X_i is the number of responses, n_i is the sample size up to Stage $i, i = 1, \ldots, K$ and θ is the response rate.
  - Testing the following hypotheses $H_0\colon\ \theta \le p_0$ versus $H_1\colon\ \theta \ge p_1$
- BOP2 design with futility early stopping
  - Stop the trial at Stage k if $X_k \ \le\ r_k$ when $Prob(\theta \le \phi\ ) > C(n)$, where $C(n) = (\frac{n}{N})^{\gamma}$
  - Declare the treatment is
    - Not promising if
      - Meeting the futility early stopping criteria and the trial is stopped early for futility, i.e., lack of efficacy
      - At the end of the trial, $X_K \ \le r_K$
    - Promising if the trial does not stop early for futility and $X_K \ge r_K + 1$
- Design parameters
  - Response rates: $p_0 = 0.2$ and $p_1 = 0.4$
  - ϕ = 0.2
  - K-stage design with
    - $K = 2, 3, 4, \text{or}, 5$
    - Min$(n_1) = 10$
    - Min(cohort size) $= 1$
    - The maximum sample size: $N_{max} = n_K = 50$
    - Control error rates: Type I error α = 0.05 or 0.10 and power: 1-β = 0.80 or 0.90

PSO Methods

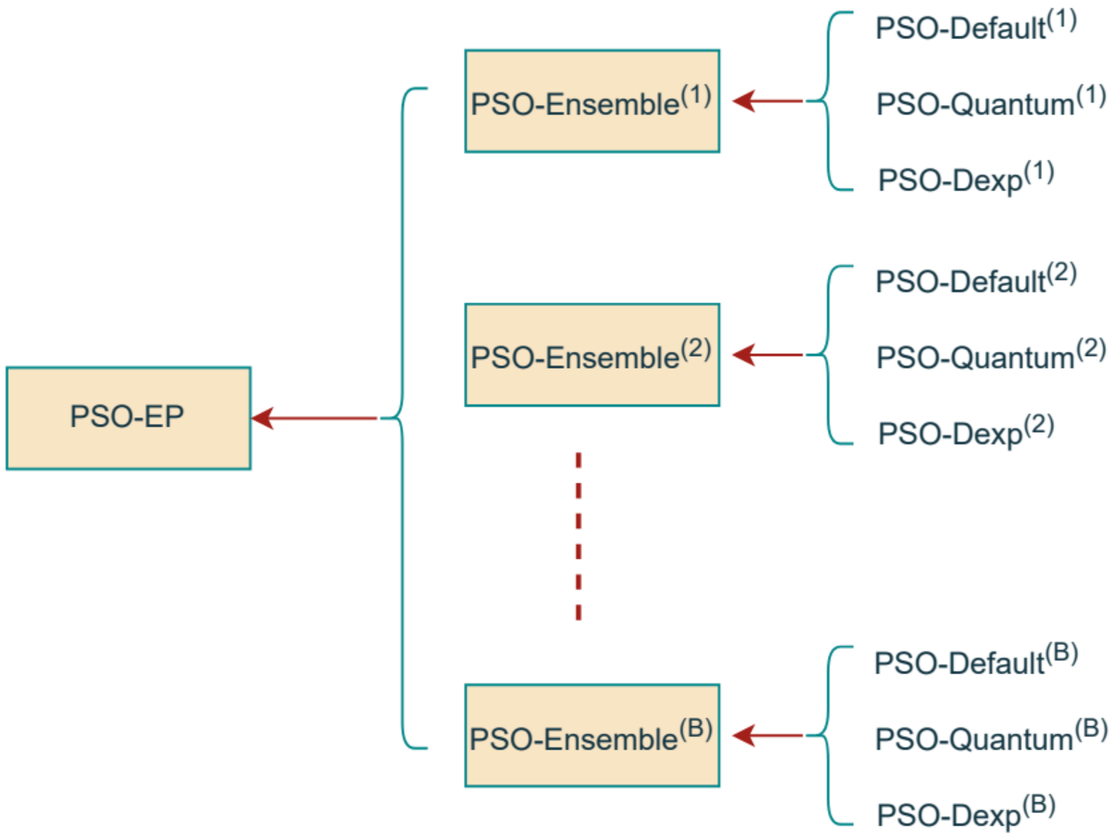
- PSO-Default : π_1
- PSO-Quantum : π_2
- PSO-DEXP : π_3

- PSO-Ensemble
  - Run individual PSO algorithms (default, quant, dexp) in parallel and pick the winner.
  - Improve *efficiency:*

$$p \geq \max(\pi_1, \pi_2, \pi_3)$$

- PSO-Ensemble Parallel (EP)
  - Run PSO-Ensemble in parallel, for a total of B replications.
  - Increase accuracy to achieve *global optimality:*

$$\lim_{B \to \infty} 1 - (1-p)^B = 1$$



Optimal 2-stage Design to Minimizing E(N| H_0)

- Utility function: $\bar{U}(\lambda, \gamma | w) = \bar{\boldsymbol{N}} + \boldsymbol{N} * \left(f_1(Power) + f_2(T1E)\right)$
- $p_0 = 0.2$ and $p_1 = 0.4$, Power = 0.90, T1E = 0.10

Method	Time (mean)	lambda (mean)	gamma (mean)	n1 (ncut1)	n2 (ncut2)	Power	Utility	Expected size	
Grid search	45.53 secs	0.83	0.515	17 (3)	37 (10)	0.903	26.022	26.022	
PSO-Default	3.91 secs	0.874	0.646	17 (3)	37 (10)	0.903	26.022	26.022	917/999
PSO-Quantum	3.37 secs	0.876	0.775	17 (3)	37 (10)	0.903	26.022	26.022	115/1000
PSO-Dexp	3.16 secs	0.876	0.724	17 (3)	37 (10)	0.903	26.022	26.022	626/999
PSO-Ensemble	3.91 secs	0.874	0.650	17 (3)	37 (10)	0.903	26.022	26.022	969/1000
PSO-EP	4.26 secs	0.876	0.658	17 (3)	37 (10)	0.903	26.022	26.022	100/100

- PSO-EP can always find the optimal solution which is identical with Simon's two-stage design.
- PSO-EP reduces the computing time by about **11** folds

Optimal 3-stage Design to Minimizing E(N| H_0)

- Utility function: $\bar{U}(\lambda, \gamma|w) = \bar{\boldsymbol{N}} + \boldsymbol{N} * \left(f_1(Power) + f_2(T1E)\right)$
- $p_0 = 0.2$ and $p_1 = 0.4$, Power = 0.90, T1E = 0.10

Method	Time (mean)	lambda (mean)	gamma (mean)	n1 (ncut1)	n2 (ncut2)	n3 (ncut3)	Power	Utility	Expected size	
Grid search	1.4 hrs	0.8	0.415	10 (1)	26 (6)	43 (11)	0.901	23.860	23.860	
PSO-Default	15.18 secs	0.863	0.552	10 (1)	26 (6)	43 (11)	0.901	23.860	23.860	26/999
PSO-Quantum	14.84 secs	0.847	0.529	10 (1)	26 (6)	43 (11)	0.901	23.860	23.860	10/995
PSO-Dexp	14.47 secs	0.863	0.564	10 (1)	26 (6)	43 (11)	0.901	23.860	23.860	53/999
PSO-Ensemble	15.08 secs	0.861	0.556	10 (1)	26 (6)	43 (11)	0.901	23.860	23.860	84/1000
PSO-EP	16.17 secs	0.859	0.557	10 (1)	26 (6)	43 (11)	0.901	23.860	23.860	100/100

- PSO-EP can always find the optimal solution which is identical with Chen's three-stage design. (Chen, *Statistics in Medicine* 1997)
- PSO-EP reduces the computing time by **312** folds

Optimal 5-stage Design to Minimizing E(N| H_0)

- Utility function: $\bar{U}(\lambda, \gamma | w) = \bar{\boldsymbol{N}} + \boldsymbol{N} * \left(f_1(Power) + f_2(T1E)\right)$
- $p_0 = 0.2$ and $p_1 = 0.4$, Power = 0.90, T1E = 0.10

Method	Time (mean)	lambda (mean)	gamma (mean)	n1 (ncut1)	n2 (ncut2)	n3 (ncut3)	n4 (ncut4)	n5 (ncut5)	Power	Utility	Expected size	
Grid search	>24 hrs	-	-	-	-	-	-	-	-	-	-	
PSO-Default-64swarms	21.12 secs	0.859	0.815	11 (1)	18 (3)	25 (5)	30 (7)	38 (10)	0.901	22.215	22.215	216/970
PSO-Default	40.17 secs	0.861	0.813	11 (1)	18 (3)	25 (5)	30 (7)	38 (10)	0.901	22.215	22.215	380/999
PSO-Quantum	47.94 secs	0.860	0.907	11 (1)	18 (3)	25 (5)	30 (7)	38 (10)	0.901	22.215	22.215	5/1000
PSO-Dexp	43.09 secs	0.862	0.849	11 (1)	18 (3)	25 (5)	30 (7)	38 (10)	0.901	22.215	22.215	132/1000
PSO-Ensemble	40.73 secs	0.862	0.817	11 (1)	18 (3)	25 (5)	30 (7)	38 (10)	0.901	22.215	22.215	468/1000
PSO-EP	52.14 secs	0.860	0.817	11 (1)	18 (3)	25 (5)	30 (7)	38 (10)	0.901	22.215	22.215	100/100

- The time for exact search is more than 24 hours. Thus, no exact solution is found yet.
- The time for PSO-Ensemble is about 41 seconds. Find the optimals in 468/1000 runs.
- The time for PSO-EP is about 52 seconds. Find the optimals in 100/100 runs and reduces the computing time by > **1600** folds

Optimal Platform Design

- Single boundary:
  - Futility early stopping

$$\begin{cases} \text{Futility:} & Pr(\theta > \phi + \delta_0 | D_n) < 1 - \lambda_f(\frac{n}{N})^{\gamma}, \text{ if } k \leq R \\ \text{Efficacy:} & Pr(\theta > \phi + \delta_0 | D_n) > \lambda_e, \text{ if } k = R + 1 \end{cases}$$

- Dual boundary:
  - Futility or Efficacy early stopping

The futility boundary is obtained through:

$$\begin{cases} Pr(\theta > \phi + \delta_0 | D_n) < 1 - \lambda_f(\frac{n}{N})^{\gamma_1} & , \text{for } k \leq R \end{cases}$$

The early graduation and final pass boundaries are:

$$\begin{cases} Pr(\theta > \phi + \delta_1(2 - (\frac{n}{N})^{\gamma_2}) | D_n) > \lambda_{e_1}(\frac{n}{N})^{\gamma_3} & , \text{for } k \leq R \\ Pr(\theta > \phi | D_n) > \lambda_{e_2} & , \text{for } k = R + 1 \end{cases}$$

- Utility function: $\bar{U}(\lambda, \gamma | w) = w\bar{\boldsymbol{N}}_{\boldsymbol{0}} + (\boldsymbol{1} - \boldsymbol{w})\bar{\boldsymbol{N}}_{\boldsymbol{1}} + \boldsymbol{N} * (f_1(Power) + f_2(T1E))$, with w=0.5
- Power = 0.80, T1E = 0.05
- Total sample size capped at 240
- Number of arms: 2; each arm has maximum 120 samples.
- True response rates for simulated drugs: 0.1, 0.2, 0.3, 0.4, 0.5
- Number of simulation replications: 1,000

Simon’s Design vs. Bayesian Adaptive Platform Design (BAPD)

Single boundary - Futility Early Stopping Only

	Cohort sizes	Boundaries	T1E	Power	EN0	EN1
Simon's	13, 43	3, 12	0.0496	0.8	20.6	37.9
BAPD	10,15,20,25,30,35,40	1,3,4,6,7,9,12	0.0341	0.8	18.3	36.5

Operating Characteristics

	Total_drug_used	Total_fut_drugs	Total_eff_drugs	Overall_accuracy	Fut_EarlyStop	Fut_accuracy	Eff_accuracy
Simon's	9.105 (1.588)	4.574 (1.882)	4.531 (0.929)	0.923 (0.092)	0.854 (0.207)	0.973 (0.096)	0.872 (0.165)
BAPD	10.148 (1.623)	5.074 (1.981)	5.074 (0.935)	0.93 (0.088)	0.957 (0.115)	0.985 (0.069)	0.869 (0.174)

Simon’s

True.theta	Fut	Indec	Eff	Exp.N	SD.N
0.1	1	0	0	13.987	5.351
0.2	0.95	0	0.05	**20.526**	13.005
0.3	0.592	0	0.408	30.405	14.806
0.4	0.197	0	0.803	**38.013**	11.169
0.5	0.047	0	0.953	41.635	6.252

BAPD

True.theta	Fut	Indec	Eff	Exp.N	SD.N
0.1	1	0	0	11.776	3.651
0.2	0.965	0	0.035	**18.246**	9.856
0.3	0.643	0	0.357	28.926	12.343
0.4	0.198	0	0.802	**36.613**	8.603
0.5	0.031	0	0.969	39.301	4.256

Fleming's Design vs. Bayesian Adaptive Platform Design (BAPD)

Dual boundaries - Futility and Efficacy Early Stopping

	Cohort sizes	Fut bounds	Eff bounds	T1E	Power	EN0	EN1
Flemming's	14, 37	3, 11	7, 12	0.048	0.802	20.675	27.067
BAPD	10 to 60 by 5	2,3,3,4,5,6,7,8,9,10,17	6,7,9,10,12,13,14,16,17,18,18	0.049	0.801	20.105	21.169

Operating Characteristics with Indecisive Decision

	Total_drug_	Total_fut_	Total_indec	Total_eff	Overall_acc	Fut_Early Stop	Eff_Early Stop	Fut_acc	Indec_acc	Eff_acc
Flemming's	11.31 (1.668)	4.525 (1.626)	2.26 (1.094)	4.525 (1.386)	0.741 (0.124)	0.526 (0.26)	0.526 (0.26)	0.975 (0.089)	0 (0)	0.877 (0.17)
BAPD	14.042 (2.766)	5.628 (1.888)	2.849 (1.184)	5.565 (1.779)	0.742 (0.109)	0.945 (0.12)	0.983 (0.068)	0.978 (0.076)	0 (0)	0.86 (0.159)

Operating Characteristics Without Indecisive Decision

	Total_drug	Total_fut	Total_eff	Overall_acc	Fut_EarlyStop	Eff_EarlyStop	Fut_acc	Eff_acc
Flemming's	11.862 (1.708)	5.965 (1.813)	5.897 (1.428)	0.928 (0.079)	0.534 (0.222)	0.534 (0.222)	0.977 (0.075)	0.877 (0.148)
BAPD	15.25 (2.689)	7.65 (2.041)	7.6 (1.874)	0.921 (0.07)	0.938 (0.103)	0.985 (0.053)	0.979 (0.058)	0.862 (0.136)

Fleming’s Design vs. Bayesian Adaptive Platform Design (BAPD)

Operating Characteristics

Fleming’s

True.theta	Fut	Indec	Eff	Exp.N	SD.N
0.1	0.957	0.043	0	14.989	4.666
0.2	0.699	0.253	0.048	**20.656**	10.43
0.3	0.355	0.265	0.381	26.699	11.437
0.4	0.123	0.075	0.802	**27.066**	11.393
0.5	0.029	0.006	0.966	22.411	11.078

BAPD

True.theta	Fut	Indec	Eff	Exp.N	SD.N
0.1	0.996	0.004	0	11.113	5.415
0.2	0.815	0.135	0.05	**20.045**	18.345
0.3	0.444	0.136	0.421	26.435	19.982
0.4	0.183	0.015	0.802	**21.176**	13.532
0.5	0.06	0	0.94	15.522	7.303

Summary

- Bayesian adaptive designs are flexible, efficient, and ethical.
  - Multi-stage designs allow early stopping for toxicity, futility, and/or efficacy to reach the correct decision sooner.
- Bayesian adaptive platform designs can evaluate multiple drugs efficiently.
  - However, there are many design parameters to choose from.
  - The dimension of design parameter can be huge.
  - Finding the optimal or near optimal design is not feasible by grid search.
- Particle Swarm Optimization (PSO) is highly efficient in finding the optimal or near optimal solutions.
  - PSO-EP can always find the optimal solution in our limited simulations.
- More in-depth studies are needed to find the best use of PSO in optimizing Bayesian adaptive platform designs and adaptive clinical trials in general.

Designing and Optimizing a Gene Therapy Clinical Development Program for a Rare Neurodegenerative Indication

Avery McIntosh

21 May 2024

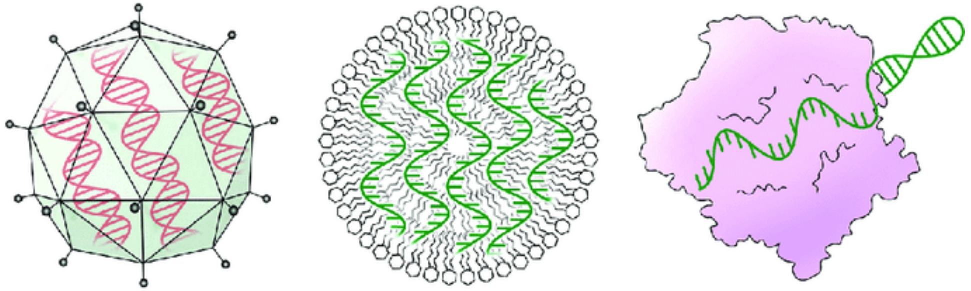
Disclaimer

- The views and opinions expressed are solely my own
- No proprietary information is presented nor are any specific programs discussed
- Nothing to disclose

Agenda

1. Brief recap of Gene Therapies (GTx)
2. Trial design differences for GTx vs other drug modalities
3. Challenges and opportunities in neurology
4. Opportunities for adaptivity

Gene therapies



Approach	Virus	Nanoparticle	Enzyme complex
Example	Adeno-associated virus (AAV) packaged with DNA encoding Cas9 & sgRNA	Liposomes encapsulating mRNA & sgRNA	Ribonucleoprotein (RNP) complex of Cas9 protein and sgRNA
Size	20 nm	50-500 nm	12 nm
Advantages	Extremely effective; prior use with classic gene therapy	Straightforward to prepare; low immunogenicity	Short lifetime and lower risk of off-target cutting

https://www.researchgate.net/publication/320339544_The_Promise_and_Challenge_of_In_Vivo_Delivery_for_Genome_Therapeutics

"*Gene therapy is a technique that modifies a person's genes to treat or cure disease*." –FDA

Major differences between cell/gene therapies and traditional pharmaceutical products (LMW/ other biologics):

- GTx are (so far) one-time administrations
- Source of safety signals is manifold: delivery mechanism, transgene insert, promoter, over/under expression
- ADME is a fundamentally different concept (biodistribution/shedding)
- CMC is major challenging for GTx
- Traditional study phases 1,2,3 often will not apply
- Dose finding is often constrained
- Often orphan, pediatric diseases, novel endpoints

FDA
Center for Biologics Evaluation and Research

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Committee for Advanced Therapies (CAT)

Pfizer

Development Landscape

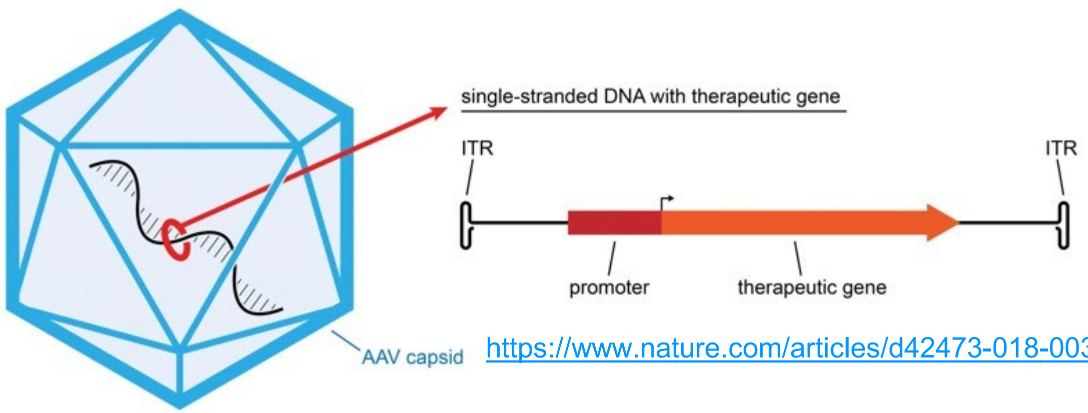
The majority of GTx to treat rare diseases in pipeline are oncology and neurology

www.asgct.org/publications/landscape-report

Global Status	Q1 2023	Q2 2023	Q3 2023	Q4 2023	Q1 2024
Preclinical	1,493	1,539	1,522	1,528	1,471
Phase I	245	240	256	270	301
Phase II	247	260	267	274	282
Phase III	30	30	30	33	35
Pre-registration	7	6	7	6	4
Total	**2,022**	**2,075**	**2,082**	**2,111**	**2,093**

Source: Pharmaprojects | Citeline, April 2024

Target Tissue/ RoA/ MoA are Variable

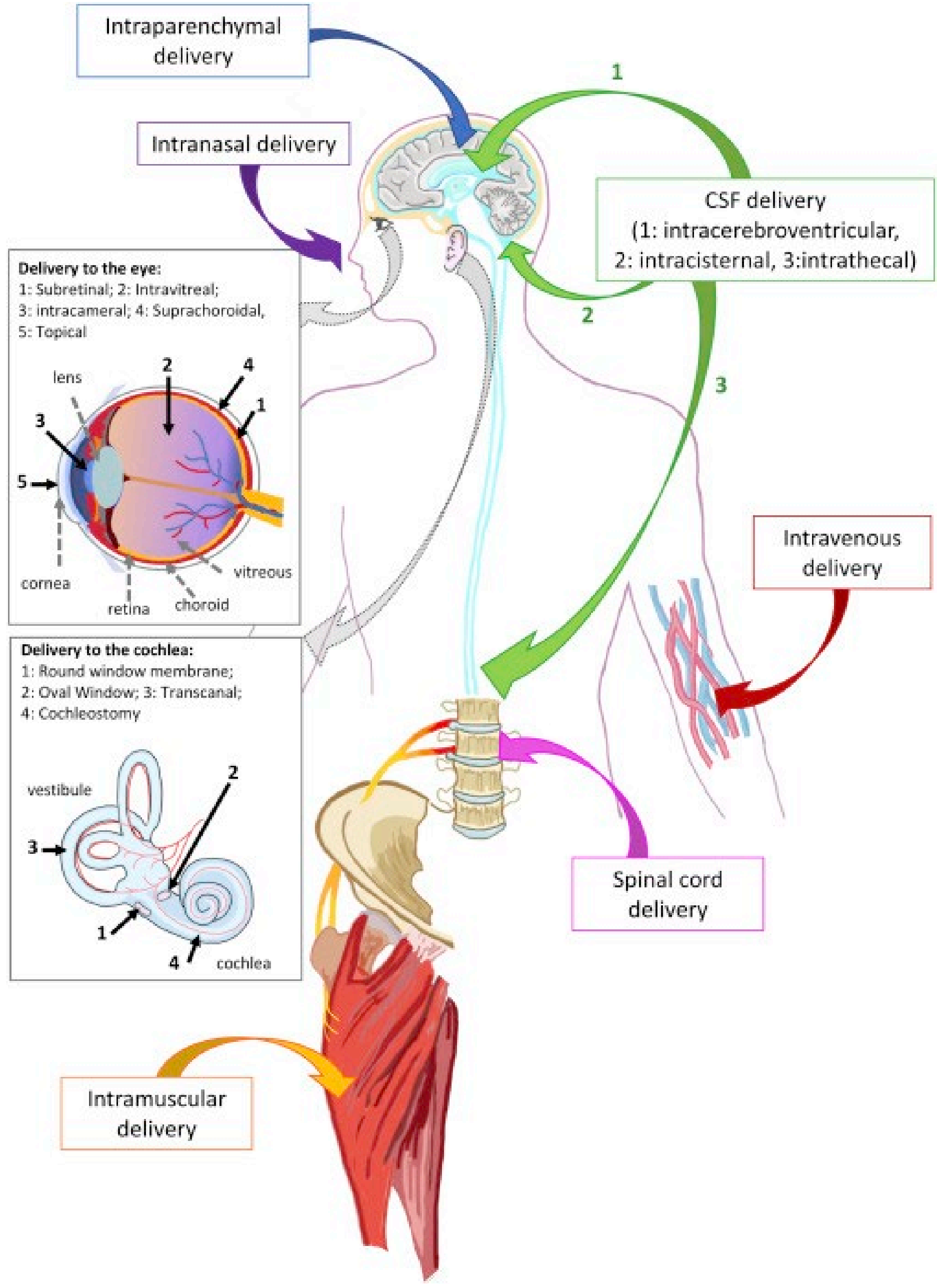


https://www.nature.com/articles/d42473-018-00307-6

Goal is to get “normal” amount of protein expression in the appropriate cell

For *in vivo* AAV GTx, different vector serotypes have different cell uptake affinities (AAV8: subretinal, AAV9: neuronal tissue, AAV2: kidney, etc.)

Very active space (capsid engineering)



Multiple MoAs (gene replacement, gene/base editing, insertion/excision, knockdown, epigenetic, etc.)

Hudrey E and Vendenberghe L. Neuron. 2019 Apr 3;102(1):263. doi: 10.1016/j.neuron.2019.03.020

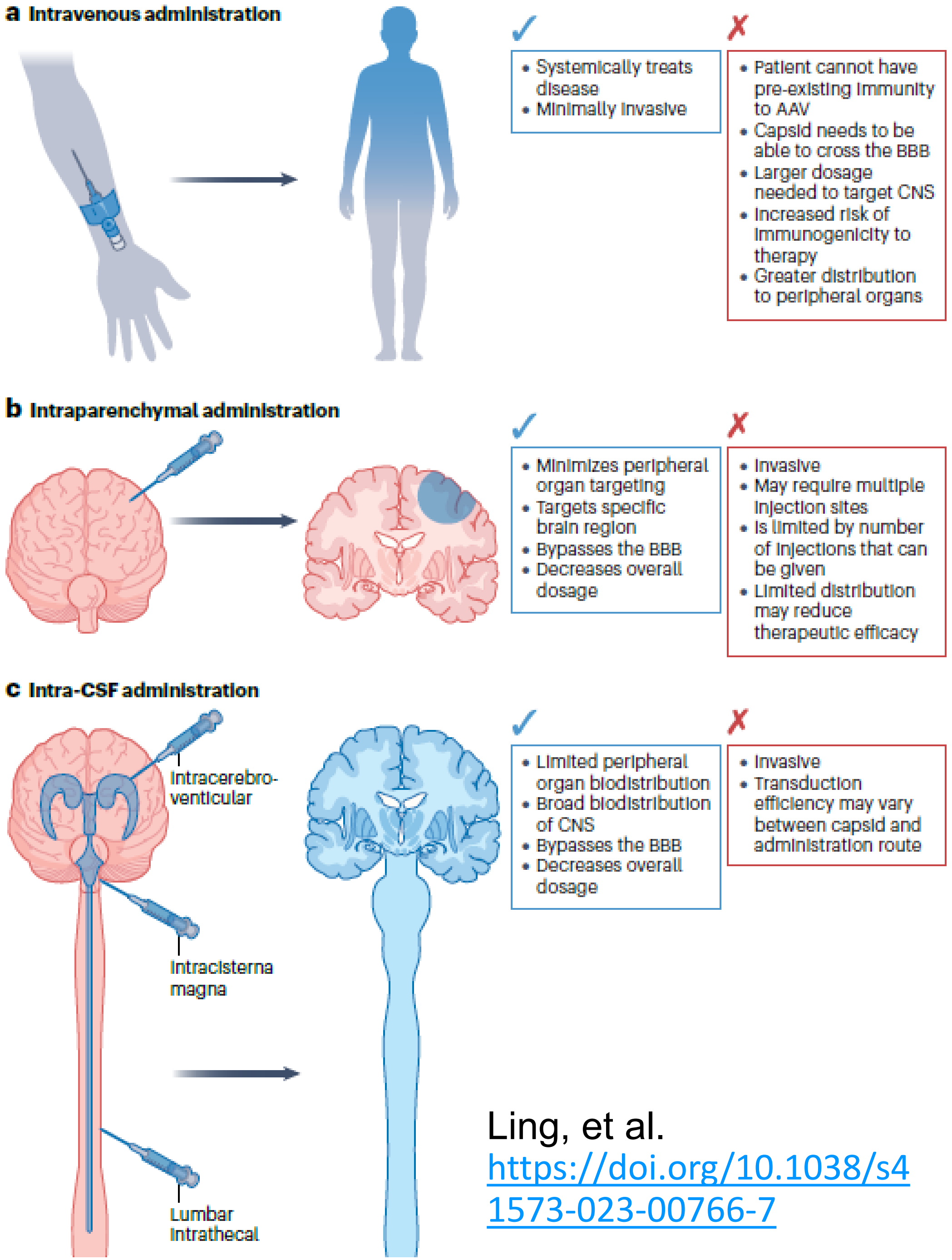
Gene Therapy for Neurological Diseases

Potential of a GTx intervention is most apparent for diseases of the CNS

Neurons are terminally differentiated, in contrast to the constantly dividing cells found in other organ systems

Thus, protein expression from an episomal gene cannot be diluted by cell division in CNS

Potentially more favorable safety profile (fewer vectors pass through liver vs systemic admin.)



Ling, et al.
https://doi.org/10.1038/s41573-023-00766-7

Pfizer

Challenges in GTx for Neurological Indications

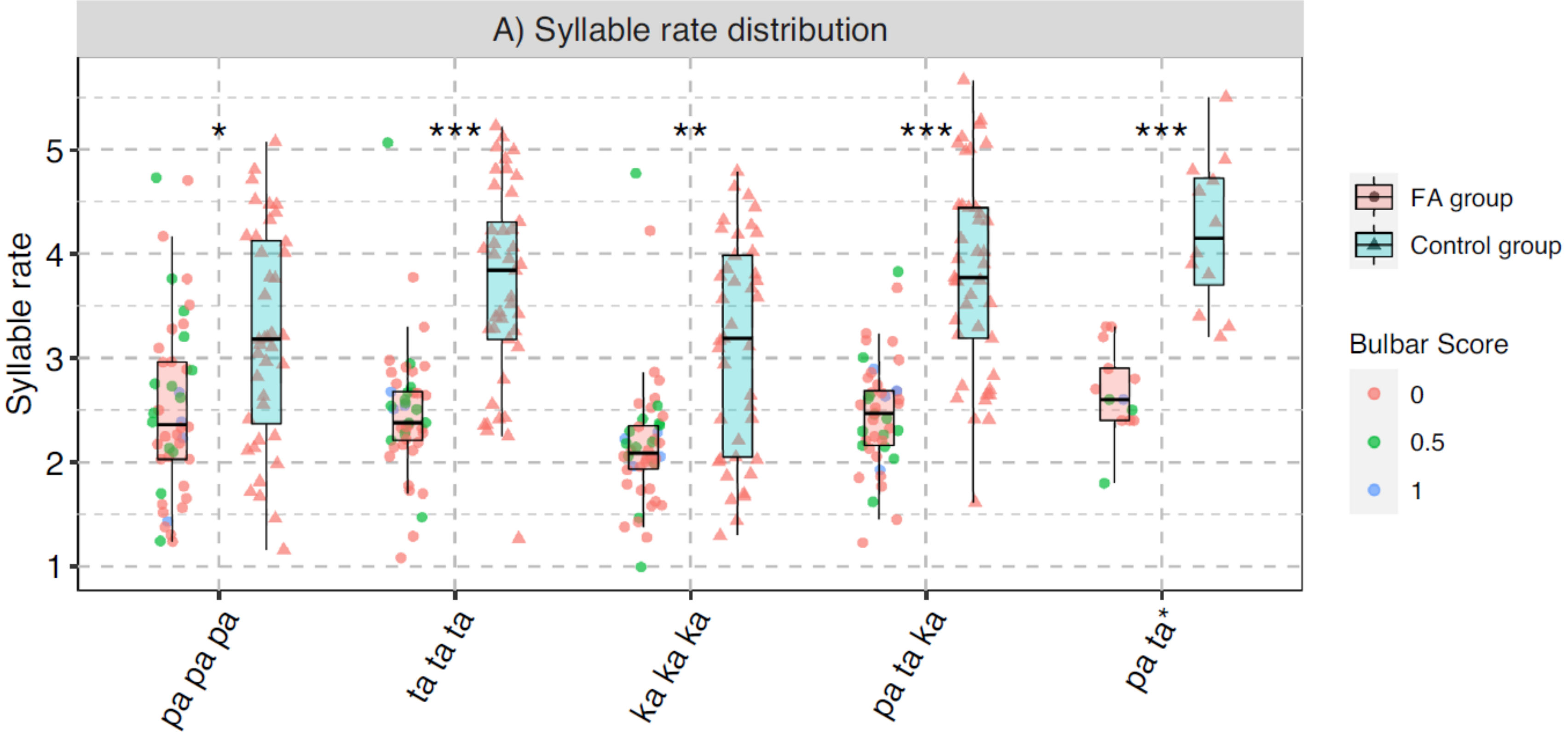
- For many CNS indications, understanding the pharmacodynamic effect will be challenging
- Reasons
  - lack of access to target tissue (invasive, hard to measure biomarkers or target engagement);
  - lack of understanding of pathogenesis (e.g. which cells/ tissues are implicated);
  - lack of validated biomarkers of pharmacodynamic effect (if the drug *is* working, may only know from functional or cognitive scales)
- This all makes *decision making* in this area very challenging. How do we know a drug works?

Tools for Neuro-muscular, -cognitive GTx

<u>Digital Endpoints</u>

- Actigraphy
- High freq cognitive testing

Using Natural History to link neuro-cog and fluid / imaging biomarkers



Mueller, et al. https://doi.org/10.1002/acn3.51438

John Shoffner [Astellas], "The Future of in vivo rAAV Gene Therapies for Rare Neurological Diseases." ISBN 9781032136554

Clinical Development Plan for Traditional Pharmaceuticals

(stepwise "learn & confirm")

Nonclinical studies	*Clinical studies*				
a. Biological activity b. Safety, tolerability c. PK d. PK-PD e. MPAD, NOAEL, FIH dose	b. Safety, tolerability c. PK d. PK-PD f. Maximum safe dose	a. Biological activity g. Dose range finding	h. Dose response for safety and efficacy i. Dose for confirmation	j. Confirmation of efficacy and safety k. Dose, subgroup for labeling	l. Long term efficacy and safety
Pharmacology Toxicology *In vitro, in vivo*	**Phase 1** *Healthy volunteers*	**Phase 2a** *Patients*	**Phase 2b** *Patients*	**Phase 3** *Patients*	**Phase 4** *Real world, Patients*

Pfizer

Oleksandr Sverdlov, Avery McIntosh, Jess LeClair, Sergey Aksenov. "Statistical Innovation for Gene Therapy Development: Clinical Trial Design and Analysis Considerations." ISBN 9781032136554

CDP for Gene Therapies (more condensed, no HVs)

Nonclinical studies

a. Biological activity
b. Safety, tolerability
c. PK
d. PK-PD
e. MPAD, NOAEL, FIH dose

Pharmacology
Toxicology
In vitro, in vivo

Clinical studies

a. Biological activity
b. Safety, tolerability
c. PK
d. PK-PD
f. Max safe dose
g. Dose range finding

Phase 1/2a
Patients

h. Dose response for safety and efficacy
j. Confirmation of efficacy and safety
k. Dose, subgroup for labeling

Phase 2b/3
Patients

l. Long term efficacy and safety

Phase 4
Real world, Patients

m. Disease progression

Natural History
Patients

Gray-highlighted text indicates changes relative to the standard sequence in figure on the previous slide

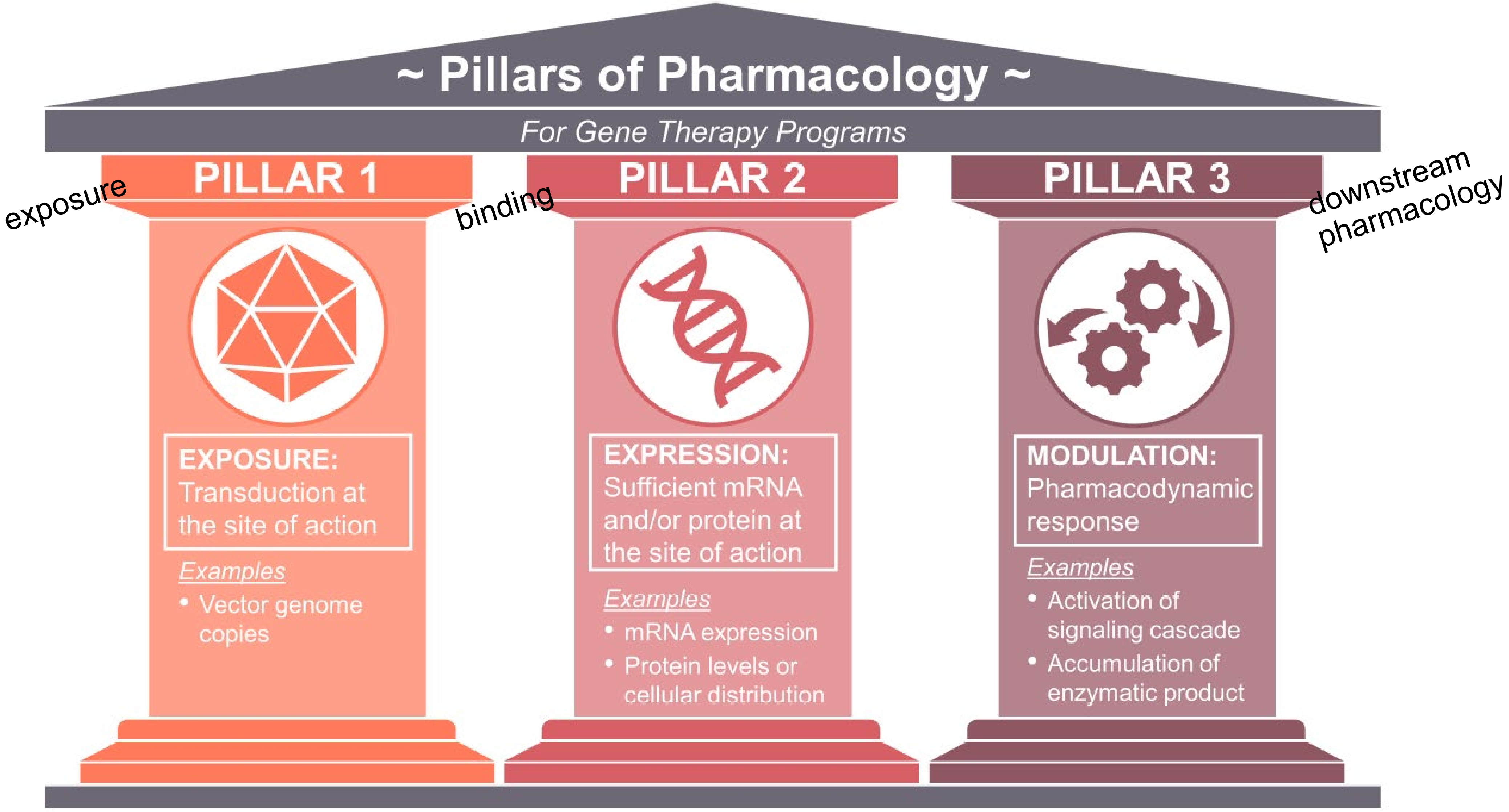
Oleksandr Sverdlov, Avery McIntosh, Jess LeClair, Sergey Aksenov. “Statistical Innovation for Gene Therapy Development: Clinical Trial Design and Analysis Considerations.” ISBN 9781032136554

Seven Hallmarks of Gene Therapy Trial Design

1. Disease modifying (curative??)
2. Usually rare disease
3. 1x administration
4. No consensus on what endpoint measures pharmacologic activity
5. Challenging safety monitoring
6. May be challenging to dose placebo in a trial
7. Long-term safety and efficacy difficult to predict

Oleksandr Sverdlov, Avery McIntosh, Jess LeClair, Sergey Aksenov. "Statistical Innovation for Gene Therapy Development: Clinical Trial Design and Analysis Considerations." ISBN 9781032136554

Defining Success by Pharmacology Principles



Mohamed Hassanein, Kelly A Fader, David Beidler. "Biomarkers in Gene Therapy Development for Rare Diseases." ISBN 9781032136554

The Real Hurdle is Sponsor Governance

The normal framework of pre-specifying success criteria at each stage should be relaxed for sponsor governance purposes in some cases

Neuro indications will often have multiple imaging, digital, functional, fluid biomarker scales

Not always clear which should be used to indicate PD activity or clinical efficacy

Post-hoc testing of multiple endpoints can be hugely informative (win ratio, MDRI, Claggett method, Wei-Lachin multivariate one-sided test, etc.)

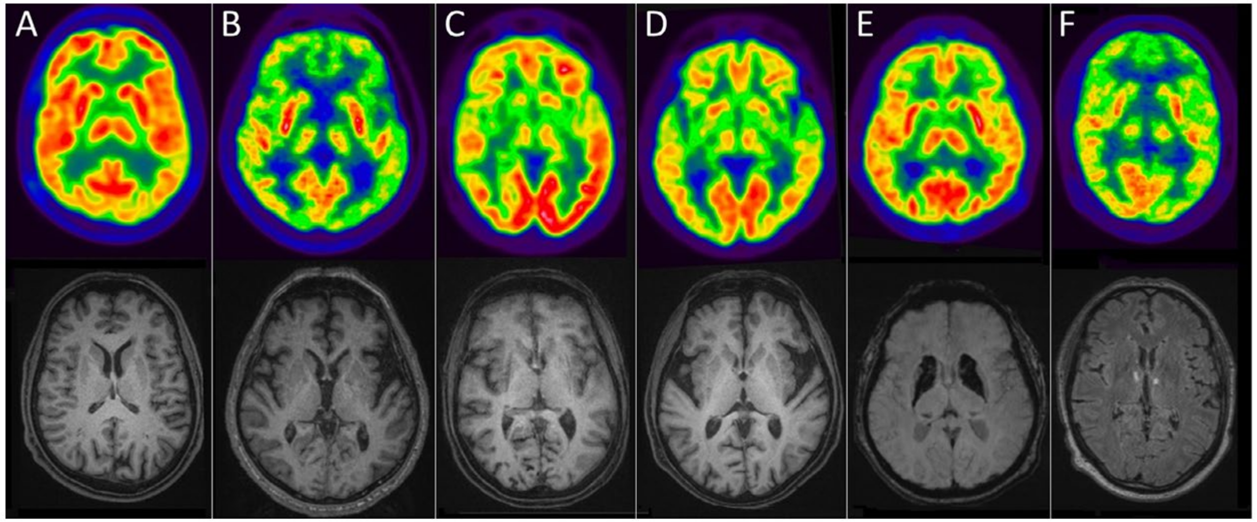


Figure 2 A number of ^{18}F-FDG PET/MR cases showing striatal hypometabolism (upper row) in different conditions. (A) healthy control (to be used as reference), (B) progressive supranuclear palsy, (C) multiple system atrophy with predominant parkinsonism, (D) Huntington disease, (E) FAHR's disease and (F) thalamic bilateral lacunar infarct. On the lower row corresponding anatomic images: T1 isotropic MPRAGE (A-D), Susceptibility-Weighted Imaging (SWI) (E) and T2-Flair (F).

Cecchin et al.
https://doi.org/10.1053/j.semnuclmed.2021.03.003

Recommendations for Clinical Development of Neuro GTx

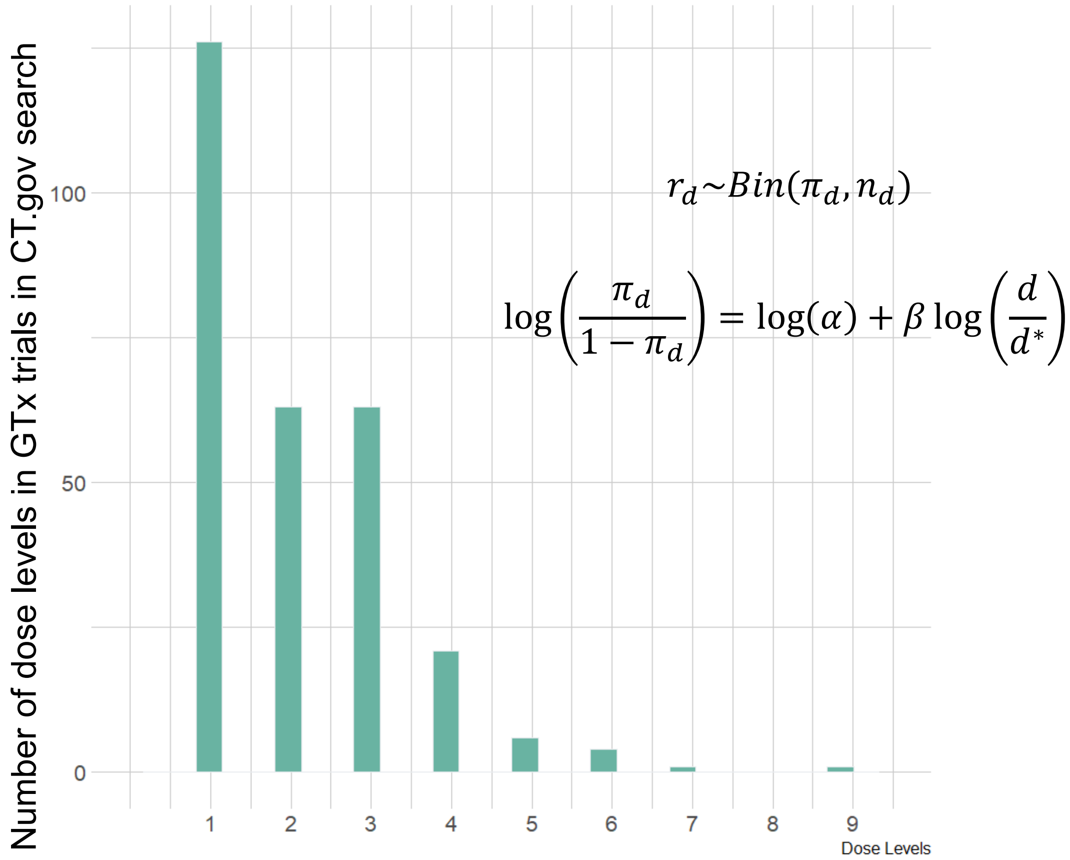
1. Utilization of patient registries and natural history studies to identify subjects.
2. Collaboration with patient advocacy groups to create educational materials providing high-quality information on gene therapy.
3. Implementation of adaptive/flexible trial designs to reduce study durations and cohort size (Bothwell et al., 2018).
4. Leverage surrogate endpoints that are reasonably likely to predict clinical benefit.
5. Adoption of decentralized clinical trial solutions (e.g., telehealth visits, electronic collection of adverse event data, home visits via visiting nurse) to help reduce the patient burden of participating in trials.
6. Inclusion of endpoints related to health economics to preemptively address payer concerns and build a solid rationale for reimbursement.
7. Utilization of real-world data to monitor safety and efficacy over extended time periods.

Petra Kaufmann [Vigil], Amanda Haidet-Phillips [Sarepta]. "Bringing Gene Therapy to Patients: A Clinical Development Perspective Based on Brain and Neuromuscular Diseases." ISBN 9781032136554

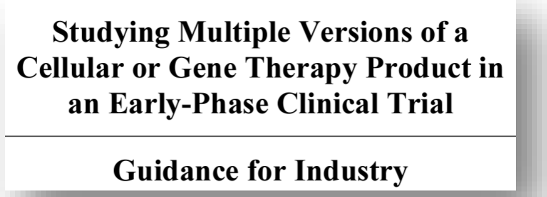
Example Areas for Adaptivity
(Trial & CDP level)

1. Quantitative dose finding *(paper in preparation)*
2. Vectors /ph1 platform trial (TBD)
3. Adaptive endpoints (nusinersen)
4. **Platform trials for long term follow-up *(published)***

1

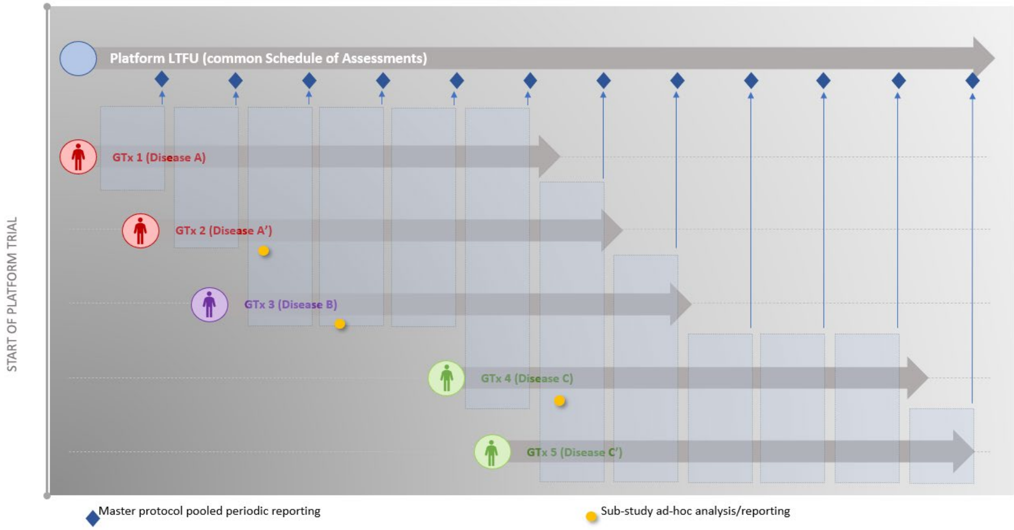


2

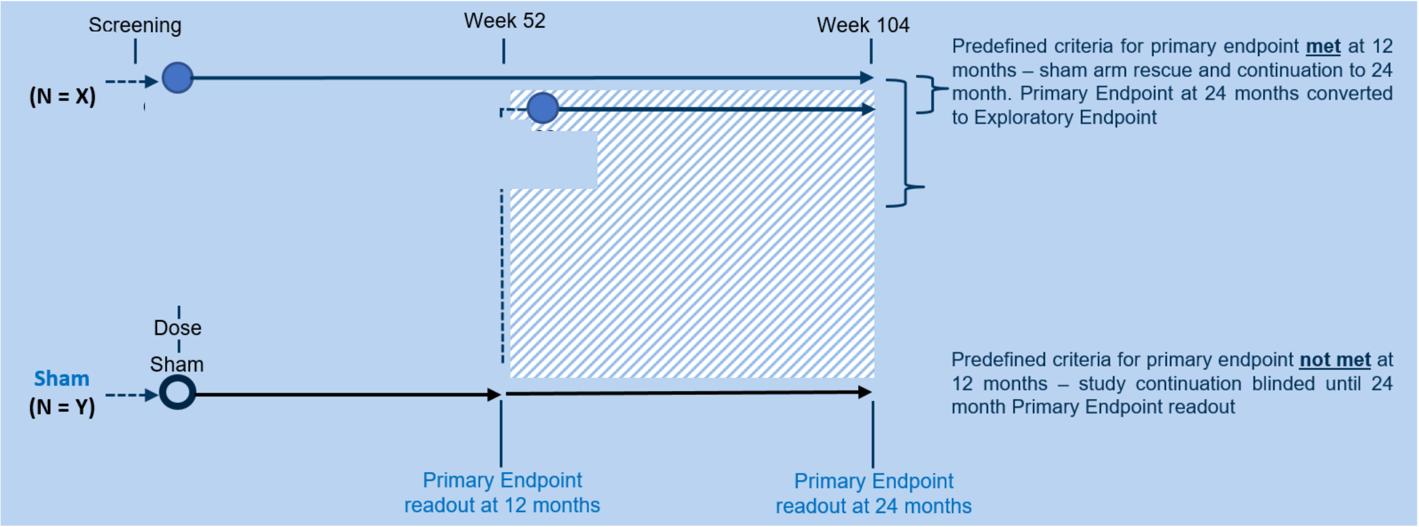

Studying Multiple Versions of a Cellular or Gene Therapy Product in an Early-Phase Clinical Trial

Guidance for Industry

4



3



Pfizer

Long Term Follow-Up (LTFU) for Gene Therapy

Why do GTx trials need long term follow-up?

- The long term safety profile of GTx products is still uncertain
- Long term data is required to fully assess benefit-risk profile
- Want to quantify the length of efficacy: 5,10 years? Lifetime?
- Assess adverse events due to the vectors:
  - Viral reactivation, immune reactions, off-target effects (e.g., dorsal root ganglion damage)
  - Risk of cancer from activating oncogenes if there is integration into the genome
  - Off-target edits from gene editing
- Collect data on long term biodistribution and viral shedding

Long Term Follow-Up (LTFU) for Gene Therapy

Why do GTx trials need long term follow-up?

- The long term safety profile of GTx products is still uncertain
- Long term data is required to fully assess benefit-risk profile
- Want to quantify the length of efficacy: 5,10 years? Lifetime?
- Assess adverse events due to the vectors:
  - Viral reactivation, immune reactions, off-target effects (e.g., dorsal root ganglion damage)
  - Risk of cancer from activating oncogenes if there is integration into the genome
  - Off-target edits from gene editing
- Collect data on long term biodistribution and viral shedding

Challenges

- FDA / EMA require sponsors to enroll patients administered a GTx product into LTFU study
- 5 – 15 years of follow-up
- Unprecedented length of engagement w/ patients: risk of loss to follow-up and lack of protocol adherence

Long Term Follow-Up (LTFU) for Gene Therapy

Why do GTx trials need long term follow-up?

- The long term safety profile of GTx products is still uncertain
- Long term data is required to fully assess benefit-risk profile
- Want to quantify the length of efficacy: 5,10 years? Lifetime?
- Assess adverse events due to the vectors:
  - Viral reactivation, immune reactions, off-target effects (e.g., dorsal root ganglion damage)
  - Risk of cancer from activating oncogenes if there is integration into the genome
  - Off-target edits from gene editing
- Collect data on long term biodistribution and viral shedding

Challenges

- FDA / EMA require sponsors to enroll patients administered a GTx product into LTFU study
- 5 – 15 years of follow-up
- Unprecedented length of engagement w/ patients: risk of loss to follow-up and lack of protocol adherence

Innovative solutions

- Platform trials / Master protocols
- Robust Bayesian hierarchical models (EXNEX) for borrowing safety information across gene therapy modalities
- Time-to-event models for adverse events
- Using existing patient registries
- Decentralized trials and use of electronic devices for data capture

Platform Trials for LTFU Allow Adaptivity in CDPs

Scientific	Pooling standardized data can address major unanswered safety questions (e.g., same assays, durations, aligned schedules of assessments for biopsies / samples)
Patient Access	A robust safety package for follow-on geographies can enhance HTA dossiers for successful reimbursement / access
Efficiency	Increased regulator confidence in safety across a therapeutic class can reduce the clinical evidence needed for adjacent populations (e.g., older, younger, different phenotype)
Pharmacologic	Clinical pharmacology models of exposure, persistence, and other dynamic parameters can be informed by longer term human data pooled across appropriate classes
Future Development	Robust safety for new products in a class (e.g., gene editing) could inform benefit-risk and increase likelihood of approval in new but adjacent indications / modality

Clinical Pharmacology & Therapeutics

Article

Practical and Statistical Considerations for the Long Term Follow-Up of Gene Therapy Trial Participants

Maximilian Rohde, Seoan Huh, Vanessa D'Souza, Steven Arkin, Erika Roberts, Avery McIntosh

First published: 27 October 2023 | https://doi.org/10.1002/cpt.3087

How /Why to Pool AE Rates Across a Class for GTx

- The European Commission's guideline on summary product characteristics (SmPC) classifies AEs in five frequency categories:

  - very rare (< 0.01%)
  - rare (< 0.1%)
  - uncommon (< 1%)
  - common (< 10%)
  - very common (≥ 10%)

- Accurate estimation of anything but "very common" and "common" is infeasible for LTFU trials that may have < 100 subjects

- The key to this limitation is in statistical tools that "borrow strength" from similar categories within a cluster

Bayesian Hierarchical Modeling (BHM)

- Hierarchical statistical models are appropriate when there is more than one level of structure or hierarchy in the data
- Strong scientific rationale to support the hypothesis that classes of gene therapy products have similar adverse event profiles:
  - Mechanism of action
  - Route of administration
  - Vector
- **For a platform trial containing related sub-studies, we should borrow information on adverse event rates (where appropriate)**
- Bayesian modeling is well-suited to hierarchical models because prior knowledge can inform the degree of information borrowing and MCMC methods can fit complex models

Neuenschwander, B., Wandel, S., Roychoudhury, S., & Bailey, S. (2016). Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical statistics*, *15*(2), 123-134.

EXNEX

- In BHMs, sharing is determined by how much data was collected in each trial
  - Trials with less data borrow more strongly from the other trials

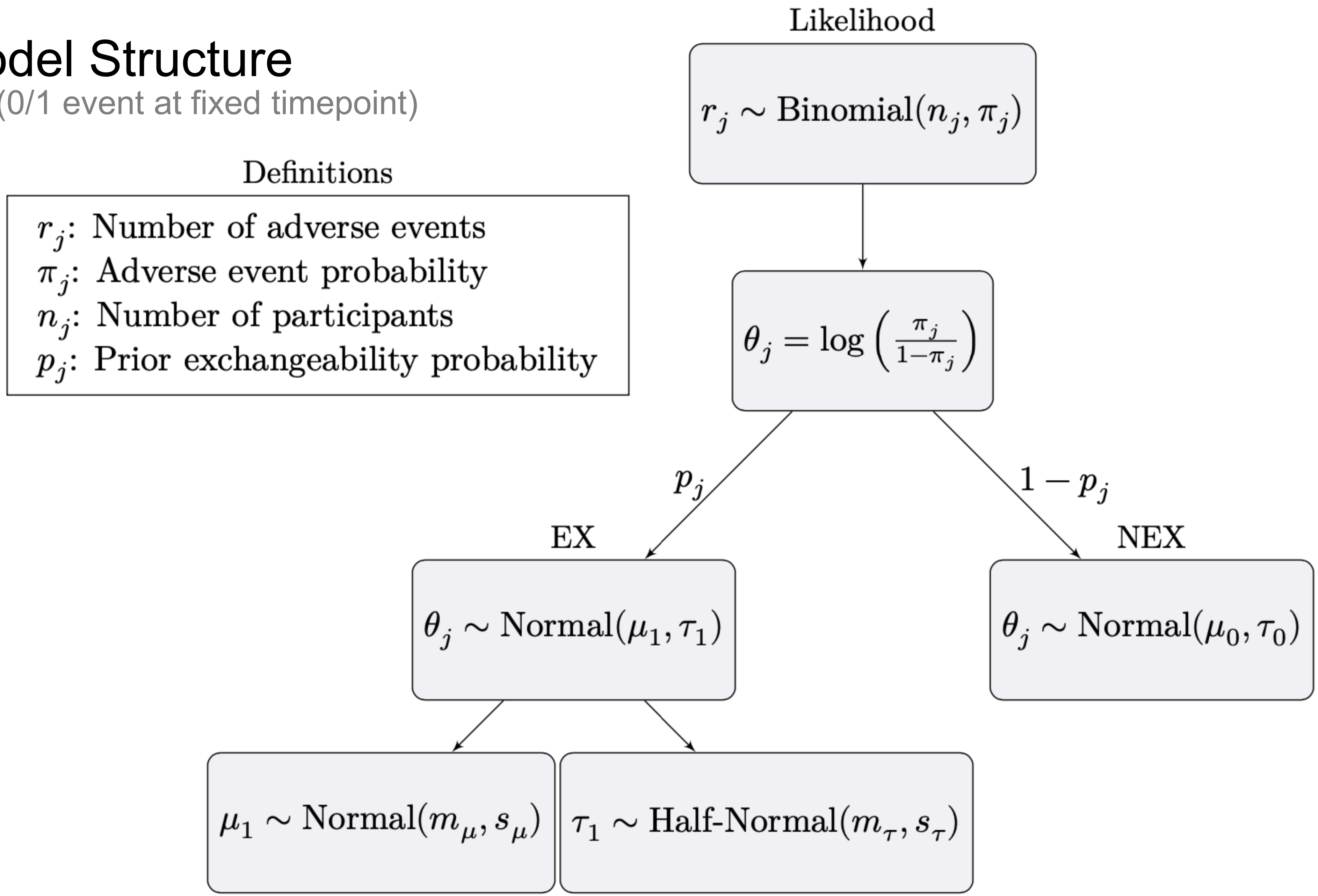
- Bayesian hierarchical models:
  - Perform well when the trials are "exchangeable" (i.e., cluster around a common rate)
  - Perform poorly if any of the trials has an extreme event rate compared to the others

- **EXNEX** ("Exchangeable/Non-Exchangeable") is an extension of BHMs that is more robust to outlier clusters
  - Mixture model where each trial is "exchangeable" with the others in platform with probability p_j or not exchangeable with *any* with probability (1- p_j)

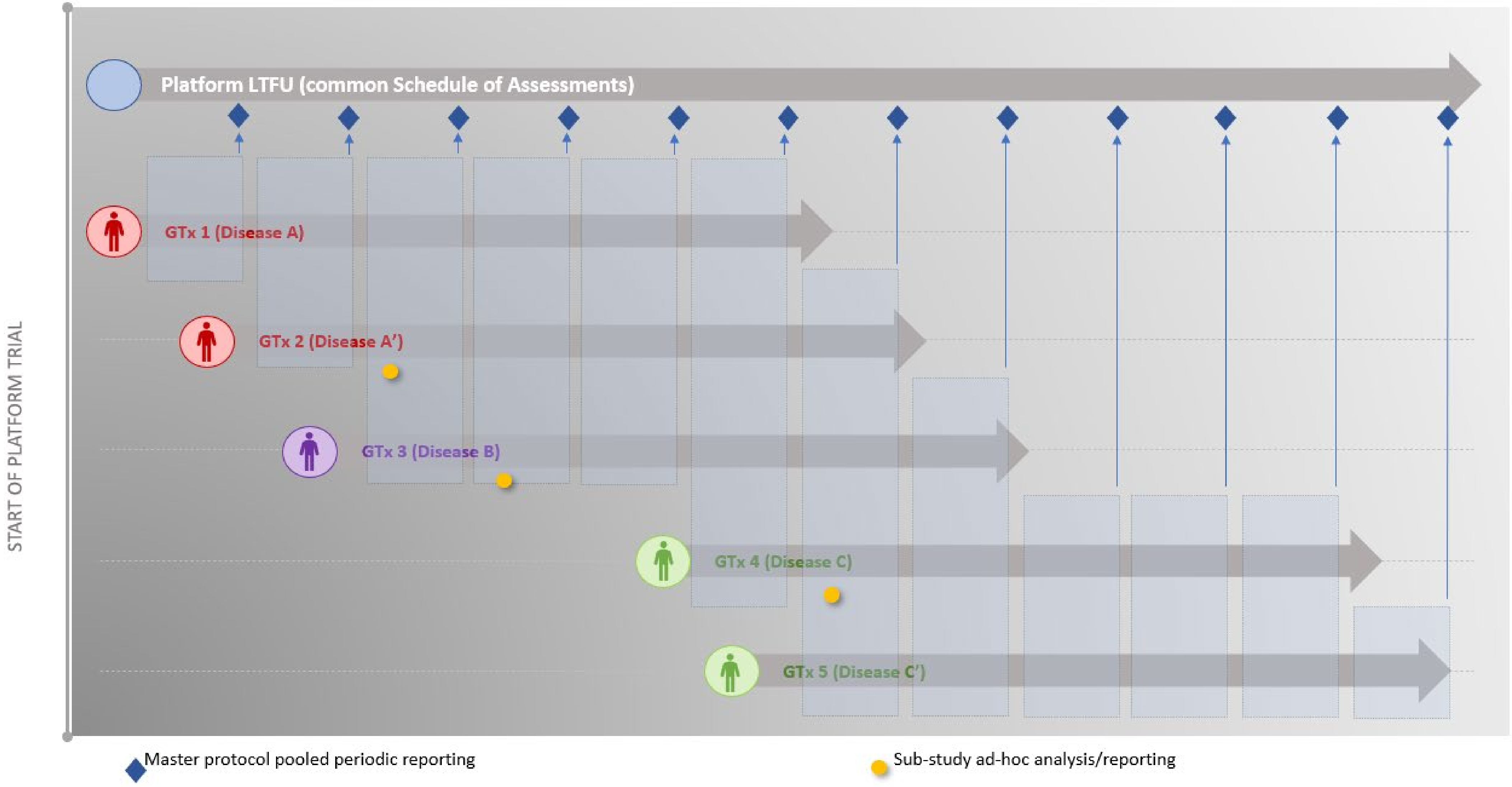
Neuenschwander, B., Wandel, S., Roychoudhury, S., & Bailey, S. (2016). Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical statistics*, *15*(2), 123-134.

EXNEX Model Structure

Binary outcome (0/1 event at fixed timepoint)



Adverse Events with Varying Follow-Up Times

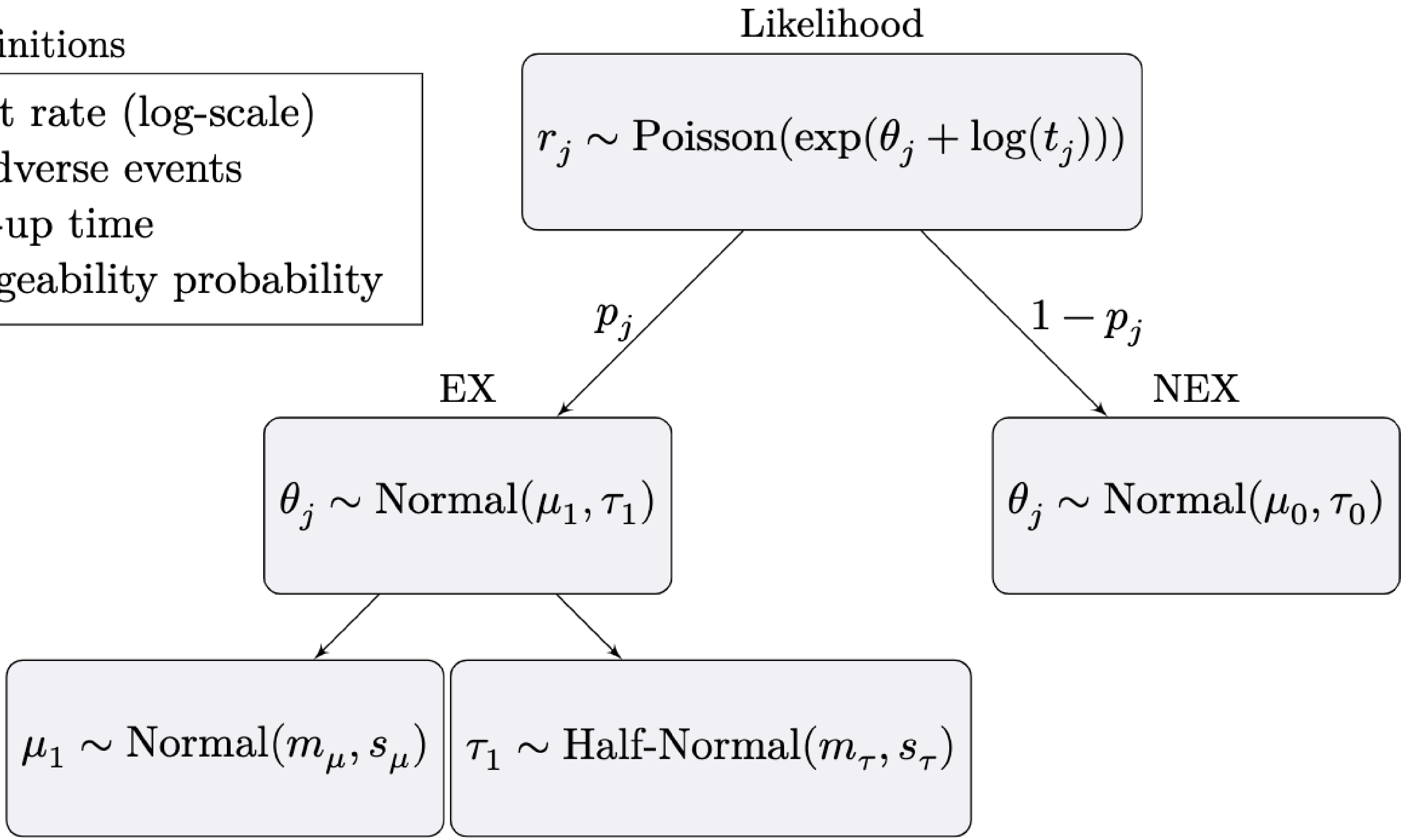


EXNEX Model Structure

Count outcome (use *offset* for varying follow-up times)

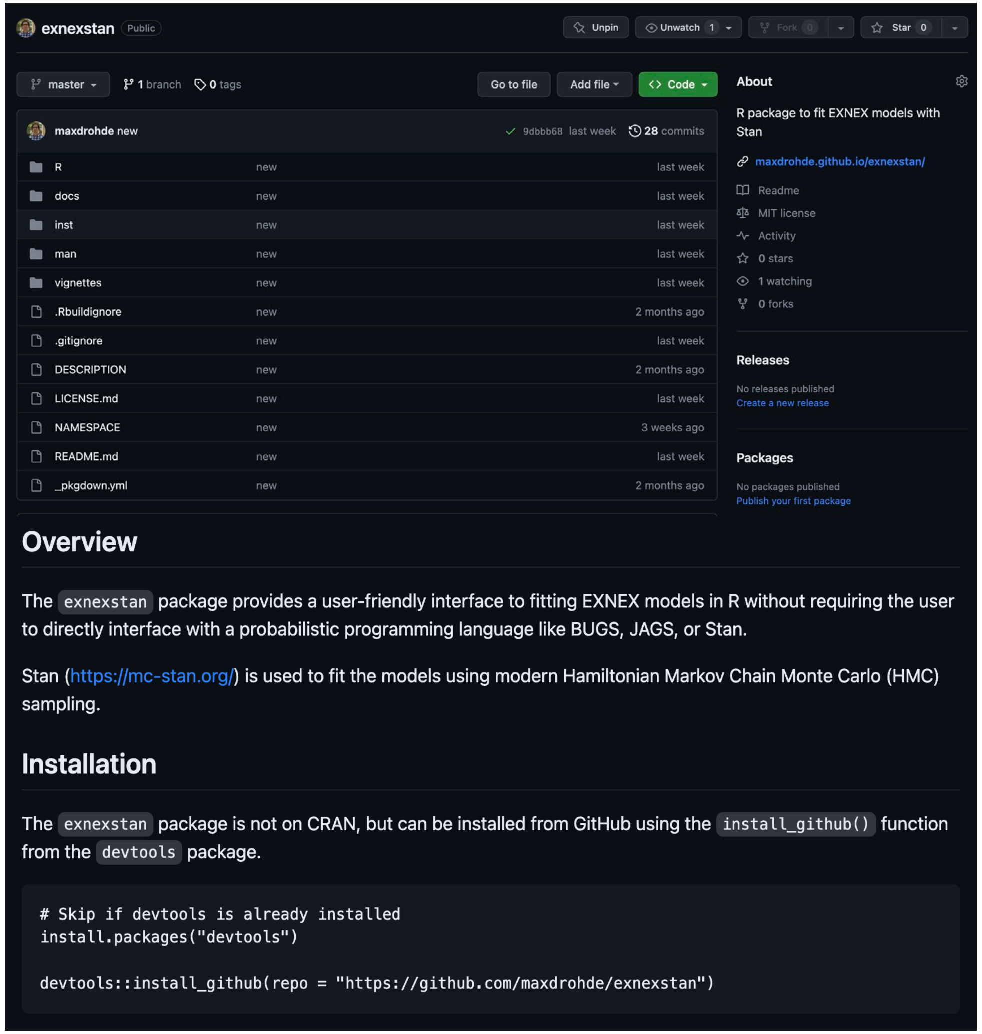
Definitions

θ_j: Adverse event rate (log-scale)
r_j: Number of adverse events
t_j: Strata follow-up time
p_j: Prior exchangeability probability



Fitting EXNEX models with the *exnexstan* R package

GitHub page


exnexstan Public

Unpin | Unwatch 1 | Fork 0 | Star 0

master | 1 branch | 0 tags | Go to file | Add file | Code

maxdrohde new · 9dbbb68 last week · 28 commits

R	new	last week
docs	new	last week
inst	new	last week
man	new	last week
vignettes	new	last week
.Rbuildignore	new	2 months ago
.gitignore	new	last week
DESCRIPTION	new	2 months ago
LICENSE.md	new	last week
NAMESPACE	new	3 weeks ago
README.md	new	last week
_pkgdown.yml	new	2 months ago

About

R package to fit EXNEX models with Stan

maxdrohde.github.io/exnexstan/

Readme
MIT license
Activity
0 stars
1 watching
0 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Overview

The `exnexstan` package provides a user-friendly interface to fitting EXNEX models in R without requiring the user to directly interface with a probabilistic programming language like BUGS, JAGS, or Stan.

Stan (https://mc-stan.org/) is used to fit the models using modern Hamiltonian Markov Chain Monte Carlo (HMC) sampling.

Installation

The `exnexstan` package is not on CRAN, but can be installed from GitHub using the `install_github()` function from the `devtools` package.

```
# Skip if devtools is already installed
install.packages("devtools")

devtools::install_github(repo = "https://github.com/maxdrohde/exnexstan")
```

R package vignettes

`exnexstan`: Binary data and package overview

Code

AUTHOR
Maximilian Rohde

PUBLISHED
July 25, 2023

Background

The `exnexstan` package implements the EXNEX model for binary data introduced in "Robust exchangeability designs for early phase clinical trials with multiple strata" by Neuenschwander et al. (2015) (https://onlinelibrary.wiley.com/doi/10.1002/pst.1730) using Stan. The `cmdstanr` package is used to interface R with the Stan probabilistic programming language that fits the models using Markov Chain Monte Carlo (MCMC)[1].

EXNEX models are an extension of Bayesian hierarchical models (BHMs). Bayesian hierarchical models are commonly used to analyze data from related studies, such as strata in a basket trial, since the partial pooling resulting from BHMs is often a good compromise between complete stratification and complete pooling. However, BHMs can perform poorly if some strata are not exchangeable with the other strata.

EXNEX is a mixture model that allows for each strata the possibility of being exchangeable (with probability p_j) with the other strata, or nonexchangeable with the other strata (with probability (1 - p_j)). This increases the robustness of the model to certain strata being not exchangeable with the others. More than two exchangeability groups may be specified in the model, although they can be difficult to fit depending on the amount of data available. Currently, `exnexstan` only supports a single exchangeability group.

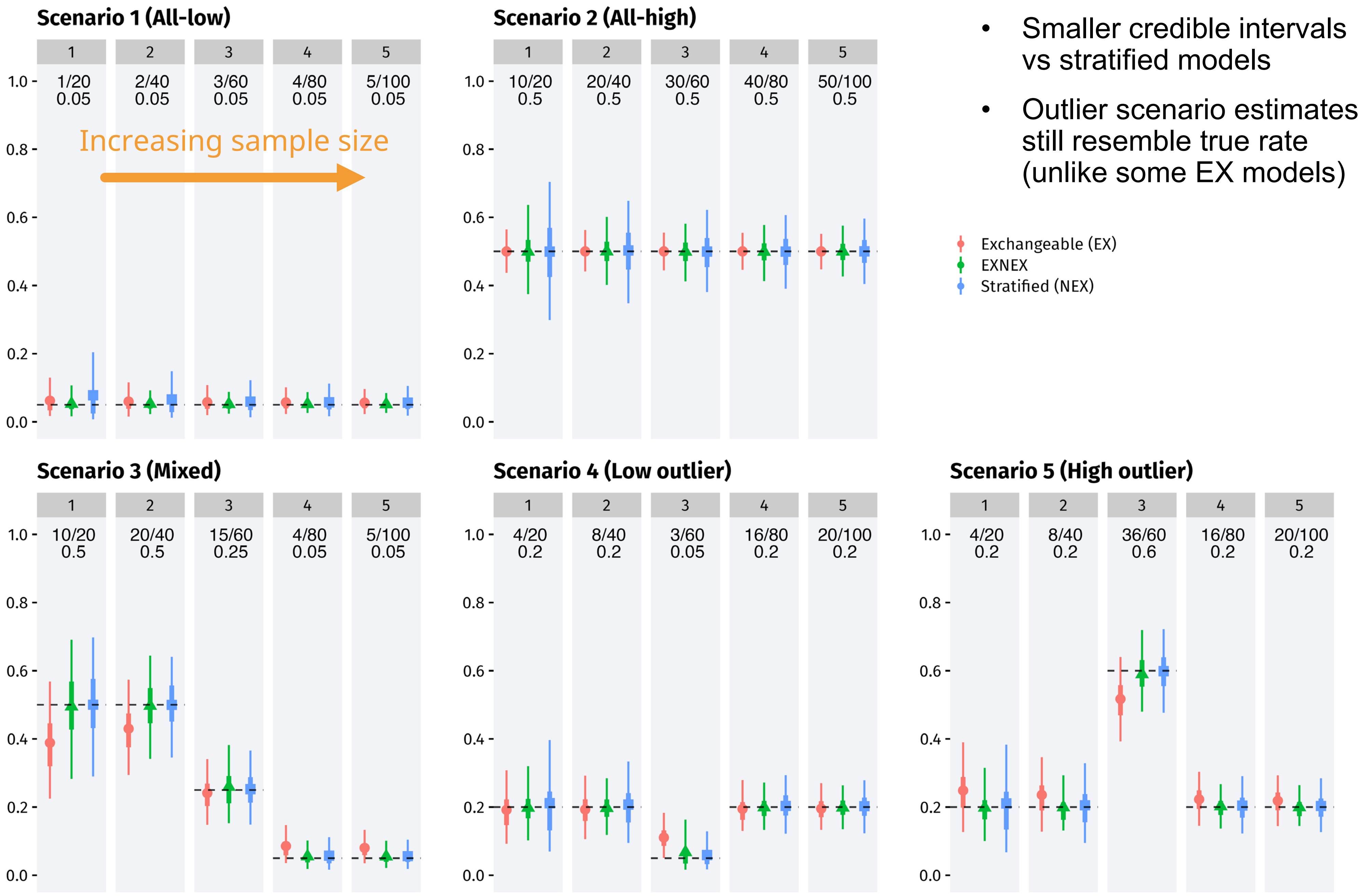
We write out the model in mathematical notation below. For clarity, we use the names for the prior values as given in the code.

$$
\begin{aligned}
Z_j &\sim \text{Bernoulli}(p_j) && \text{(Indicator variable of EX vs NEX)}\\
\theta_j &\sim \text{Normal}(\text{mean} = \mu_{Z_j}, \text{sd} = \tau_{Z_j}) && \text{(Response probability on log-odds scale)}\\
\mu_0 &= \texttt{nex\_prior\_mean} && \text{(NEX mean)}\\
\tau_0 &= \texttt{nex\_prior\_sd} && \text{(NEX standard deviation)}
\end{aligned}
$$

EXNEX Scenario

Binary outcome

- 1,2,3,4,5 columns are platform sub-studies
- Numbers at top of each column are #AEs / cohort sample size, and associated rate
- Dotted lines are true event rate



- Smaller credible intervals vs stratified models
- Outlier scenario estimates still resemble true rate (unlike some EX models)

More on Gene Therapy Drug Development

Available from https://www.routledge.com and other booksellers

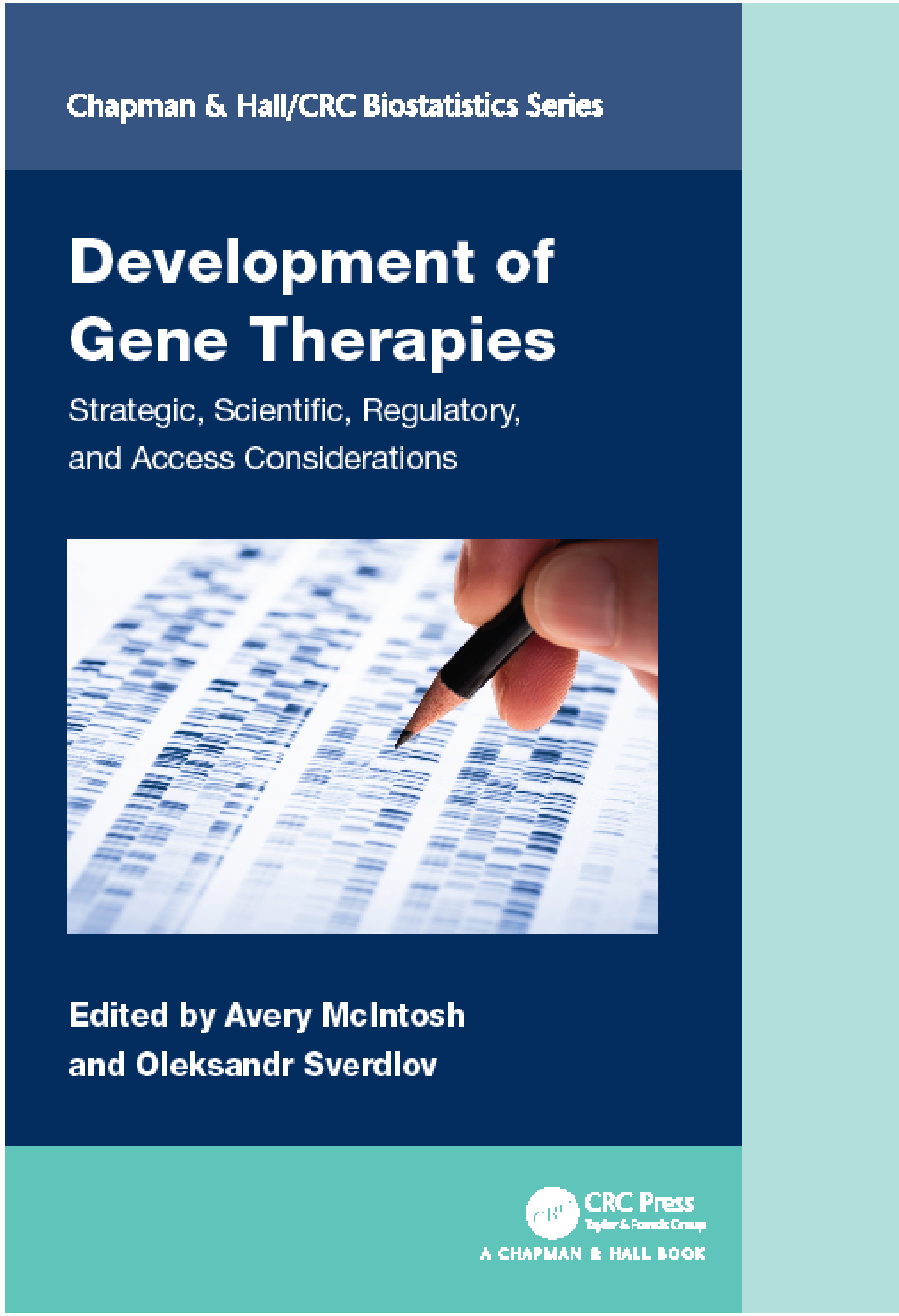
19 chapters from experts in industry and academia, with a focus on strategic and operational considerations from multi-stakeholder perspectives

Three recent publications on GTx trial design & analysis:

Clinical Pharmacology & Therapeutics

Clinical Pharmacology & Therapeutics

ANNALS of Clinical and Translational Neurology Open Access



Thank You