Design and Analysis of Drop-the-Losers Studies in the Rare Disease Setting

Ina Jazić[1], Xiaoyan Liu[1], Glen Laird[1]

[1]Vertex Pharmaceuticals, Inc

NERDS 2020 Webinar Series #4

DISCLOSURE/DISCLAIMER

- The opinions expressed in this presentation and on following slides are those of the speaker, and do not necessarily represent those of the associated affiliation

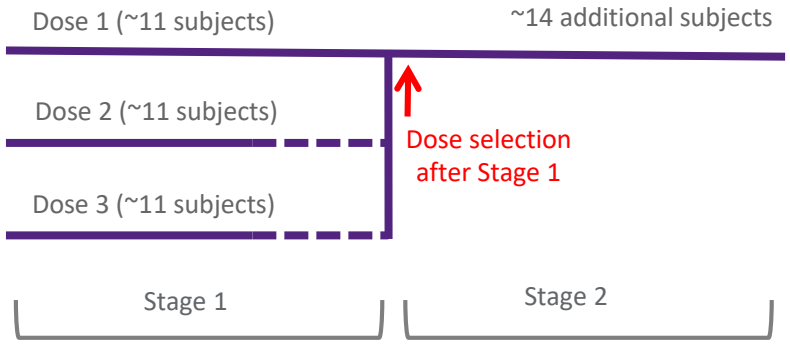
- No confidential data is disclosed

OUTLINE

- Motivation
- Introduction to drop-the-losers designs and existing methods
- Unconditional exact likelihood (UEL) framework
- Simulation results
- Additional design considerations
- Robustness to deviations from prespecified selection rule
- Future directions

MOTIVATION: EXAMPLE RARE DISEASE PIVOTAL PH2B/3 STUDY

3 doses of drug in Stage 1; 1 dose chosen to enter Stage 2
Endpoint: complete remission rate at a defined timepoint



Justification for non-inclusion of control arm

- Ultra-rare disease
- Expected spontaneous response/remission rate very low (<10%)

Key features

- Binary endpoint
- No control arm
- No early stopping for efficacy

INTRODUCTION: DROP-THE-LOSERS DESIGNS

Goal: combine Ph2 and Ph3 trials under a single protocol with known upfront sample size, speeding up clinical development

Stage 1: n_1 subjects assigned to each of k treatments, and possibly a control arm

- $X_i \sim Bin(n_1, \pi_i)$: total number of Stage 1 responses under treatment i. Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$.
- Select a single treatment $\boldsymbol{\tau}$ with highest number of observed responses $\max\{X_1, \dots, X_k\}$ to proceed to Stage 2 for further evaluation
  - Ties are broken through pre-specified treatment preferences, such that τ is the treatment with the smallest index among those with the highest observed response

Stage 2: n_2 subjects assigned to treatment τ, and possibly a control arm

- $Y \sim Bin(n_2, \pi_\tau)$: total number of Stage 2 responses under treatment τ

Use data from both stages to draw conclusions about the selected treatment

INTRODUCTION: EXISTING METHODS

Methods for drop-the-losers designs for the binomial case:

- Tappin (1992) derived UMVUE* for response rate in selected arm. UMVUE exists if ties broken according to a pre-specified order, but not if broken randomly.
- **Sill and Sampson (2009)** extended earlier work to the binomial, following conditional approach of Tappin (1992) to derive UMPCU** test and interval estimation procedure.

Closed testing – agnostic to endpoint:

- Closed testing procedure using p-value combination tests frequently used in this setting; described in **Posch *et al.* (2005)**, a generalization of Bauer and Kieser (1999)

*- uniformly minimum variance unbiased estimator

**-uniformly most powerful conditionally unbiased

UNCONDITIONAL EXACT LIKELIHOOD (UEL)

We propose the unconditional exact likelihood (UEL) framework for the analysis of drop-the-losers studies with binary endpoints and no control arm .

The test and confidence interval are based on the distribution of the statistic of interest: $\boldsymbol{Z_\tau = X_\tau + Y}$, the total number of responses in the *entire study* for the treatment selected after Stage 1.

- Recall: $\boldsymbol{X_\tau}$ and $\boldsymbol{Y}$ are the Stage 1 and 2 responses for arm τ, respectively

UEL: DISTRIBUTION OF Z_τ

$\boldsymbol{Z_\tau = X_\tau + Y}$. By change-of-variable, $f_{Z_\tau}(Z_\tau = z) = f_Y(Z_\tau - X_\tau = z - x)$. The constraints are $\boldsymbol{0 \le X_\tau \le n_1}$ and $0 \le Y \le n_2$. The latter constraint can be transformed into $\boldsymbol{Z_\tau - n_2 \le X_\tau \le Z_\tau}$.
Then, $f_{Z_\tau}(Z_\tau = z)$ can be expressed as the joint distribution of X_τ and Y marginalized over X_τ:

$$f_{Z_\tau}(Z_\tau = z) = f_Y(Z_\tau - X_\tau = z - x) = \sum_{x=\max(0,z-n_2)}^{\min(n_1,z)} f(X_\tau = x, Y = z - x)$$

$$= \sum_{x=\max(0,z-n_2)}^{\min(n_1,z)} \sum_{j=1}^{k} f(X_\tau = x, Y = z - x, \tau = j)$$

Since X_τ and Y are independent, this can be rewritten as:

$$f_{Z_\tau}(Z_\tau = z) = \sum_{x=\max(0,z-n_2)}^{\min(n_1,z)} \sum_{j=1}^{k} f(X_\tau = x, \tau = j) f(Y = z - x | \tau = j)$$

Probability of treatment j being selected and having x responses in Stage 1

Ordinary binomial probability for Stage 2 responses

UEL: DISTRIBUTION OF Z_τ

$$f_{Z_\tau}(Z_\tau = z) = \sum_{x=\max(0,z-n_2)}^{\min(n_1,z)} [\underbrace{\binom{n_1}{x}\binom{n_2}{z-x}\pi_1^z(1-\pi_1)^{n_1+n_2-z}\prod_{s=2}^{k}P(X_s \le x)}_{\tau=1}$$

$$+ \underbrace{\sum_{j=2}^{k-1}[\binom{n_1}{x}\binom{n_2}{z-x}\pi_j^z(1-\pi_j)^{n_1+n_2-z}\prod_{s_1=1}^{j-1}P(X_{s_1} < x)\prod_{s_2=j+1}^{k}P(X_{s_2} \le x)]}_{\tau \in [2,k-1]}$$

$$+ \underbrace{\binom{n_1}{x}\binom{n_2}{z-x}\pi_k^z(1-\pi_k)^{n_1+n_2-z}\prod_{s=1}^{k-1}P(X_s < x)]}_{\tau=k}$$

Tie-breaking rule:

Treatment 1
preferred over
Treatment 2
...
preferred over
Treatment k

UEL: DISTRIBUTION OF Z_τ

$$f_{Z_\tau}(Z_\tau = z) = \sum_{x=\max(0,z-n_2)}^{\min(n_1,z)} [\binom{n_1}{x}\binom{n_2}{z-x}\pi_1^z(1-\pi_1)^{n_1+n_2-z}\prod_{s=2}^{k}P(X_s \le x)$$

$\tau = 1$

$$+\sum_{j=2}^{k-1}[\binom{n_1}{x}\binom{n_2}{z-x}\pi_j^z(1-\pi_j)^{n_1+n_2-z}\prod_{s_1=1}^{j-1}P(X_{s_1} < x)\prod_{s_2=j+1}^{k}P(X_{s_2} \le x)]$$

$\tau \in [2, k-1]$

$$+\binom{n_1}{x}\binom{n_2}{z-x}\pi_k^z(1-\pi_k)^{n_1+n_2-z}\prod_{s=1}^{k-1}P(X_s < x)]$$

$\tau = k$

Tie-breaking rule:

Treatment 1
preferred over
Treatment 2
...
preferred over
Treatment k

$\tau = 1$

$$f(X_\tau = x, \tau = 1) = f(X_1 = x, X_2 \le x, \dots, X_k \le x) = P(X_1 = x)\prod_{s=2}^{k}P(X_s \le x)$$

UEL: DISTRIBUTION OF Z_τ

$$f_{Z_\tau}(Z_\tau = z) = \sum_{x=\max(0,z-n_2)}^{\min(n_1,z)} [\binom{n_1}{x}\binom{n_2}{z-x}\pi_1^z(1-\pi_1)^{n_1+n_2-z}\prod_{s=2}^{k}P(X_s \le x)$$

$\tau = 1$

$$+\sum_{j=2}^{k-1}[\binom{n_1}{x}\binom{n_2}{z-x}\pi_j^z(1-\pi_j)^{n_1+n_2-z}\prod_{s_1=1}^{j-1}P(X_{s_1} < x)\prod_{s_2=j+1}^{k}P(X_{s_2} \le x)]$$

$\tau \in [2, k-1]$

$$+\binom{n_1}{x}\binom{n_2}{z-x}\pi_k^z(1-\pi_k)^{n_1+n_2-z}\prod_{s=1}^{k-1}P(X_s < x)]$$

$\tau = k$

Tie-breaking rule:

Treatment 1
preferred over
Treatment 2
...
preferred over
Treatment k

$\tau \in [2, k-1]$

$$f(X_\tau = x, \tau = j) = f(X_1 < x, \dots, X_{j-1} < x, X_j = x, X_{j+1} \le x, \dots, X_k \le x)$$

$$= P(X_j = x)\prod_{s_1=1}^{j-1}P(X_{s_1} < x)\prod_{s_2=j+1}^{k}P(X_{s_2} \le x)$$

UEL: DISTRIBUTION OF Z_τ

$\tau = 1$

$$f_{Z_\tau}(Z_\tau = z) = \sum_{x=\max(0,z-n_2)}^{\min(n_1,z)} \Big[\binom{n_1}{x}\binom{n_2}{z-x} \pi_1^z (1-\pi_1)^{n_1+n_2-z} \prod_{s=2}^{k} P(X_s \le x)$$

$\tau \in [2, k-1]$

$$+ \sum_{j=2}^{k-1} \Big[\binom{n_1}{x}\binom{n_2}{z-x} \pi_j^z (1-\pi_j)^{n_1+n_2-z} \prod_{s_1=1}^{j-1} P(X_{s_1} < x) \prod_{s_2=j+1}^{k} P(X_{s_2} \le x) \Big]$$

$$+ \binom{n_1}{x}\binom{n_2}{z-x} \pi_k^z (1-\pi_k)^{n_1+n_2-z} \prod_{s=1}^{k-1} P(X_s < x) \Big]$$

$\tau = k$

Tie-breaking rule:

Treatment 1
preferred over
Treatment 2
...
preferred over
Treatment k

$\tau = k$

$$f(X_\tau = x, \tau = k) = f(X_1 < x, \dots, X_{k-1} < x, X_k = x) = P(X_k = x) \prod_{s=1}^{k-1} P(X_s < x)$$

UEL: HYPOTHESIS TEST

Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$. Global null hypothesis: $\boldsymbol{\pi} = (\pi_0, \dots, \pi_0)$.

Define a hypothesis test based on the **distribution of $\boldsymbol{Z_\tau}$ under the global null hypothesis** *without specification of the selected arm.*

- In the uncontrolled case, π_0 is pre-specified – clinically meaningful threshold or control response rate from natural history study or historical data
- Type 1 error control under the global null provides type 1 error control in the strong sense
  - Fixing one arm as π_0, decreasing or increasing the response rates for the other two arms both lead to a lower type 1 error rate
- Throughout these slides, one-sided level $\alpha/2$ tests are conducted

The null hypothesis is rejected if $\boldsymbol{P[Z_\tau \geq z | \pi = (\pi_0, \dots, \pi_0)] < \alpha/2}$, where z is the total number of observed responses for arm τ.

UEL: HYPOTHESIS TEST

Example: let $\pi_0 = 0.1, n_1 = n_2 = 15$, and $k = 3$.

- If the observed number of responses in arm τ is 6, then the *p*-value for the UEL test is $P[Z_\tau \geq 6|\boldsymbol{\pi} = (0.1,0.1,0.1)] = 0.163$.
- For $\alpha/2 = 0.025$, this test would **fail to reject the null**.
- The critical value is the smallest z such that $P[Z_\tau \geq z|\boldsymbol{\pi} = (0.1,0.1,0.1)] < 0.025$.
  - In this case, the critical value is 8, with a corresponding *p*-value of 0.020.
- R functions in `DropTheLosersUEL` for the p-value and critical value can be found below:

```
> pval_ztau(z=6, pi0=0.1, ntrts=3, n1=15, n2=15)
[1] 0.1634468
> crit_ztau(pi0=0.1, ntrts=3, n1=15, n2=15)
[1] 8
> pval_ztau(z=8, pi0=0.1, ntrts=3, n1=15, n2=15)
[1] 0.02043866
```

TEST VS. CONFIDENCE INTERVAL

The hypothesis test is based on the marginal distribution of Z_τ under the global null hypothesis, **regardless of which arm was selected for Stage 2**.

- Aligns with scientific interest for the test

However, when constructing a confidence interval for Z_τ, scientific interest lies in a reasonable interval for the response rate for **the particular dose that was selected**.

- Using a marginal interpretation is not as helpful

UEL: CONFIDENCE INTERVAL

To construct a $100 \times (1-\alpha)$ % CI for π_τ,

- **Lower bound:** $P(Z_{\boldsymbol{\tau}} \geq z | \boldsymbol{\tau}, \boldsymbol{\pi} = \boldsymbol{\pi}^*_{\boldsymbol{LB}}) = \alpha/2$
- **Upper bound:** $P(Z_{\boldsymbol{\tau}} \leq z | \boldsymbol{\tau}, \boldsymbol{\pi} = \boldsymbol{\pi}^*_{\boldsymbol{UB}}) = \alpha/2$

The forms of $\boldsymbol{\pi}^*_{\boldsymbol{LB}}$ and $\boldsymbol{\pi}^*_{\boldsymbol{UB}}$ must be prespecified.

- The distribution of Z_τ depends on all elements of $\boldsymbol{\pi}$. In a confidence interval for π_τ, the remaining $k-1$ response rates are nuisance parameters.
- A natural interval estimator uses the **observed Stage 1 response rates** as **plug-in estimators** for these nuisance parameters.

For example, if $\tau = 1$, a confidence interval (π_L, π_U) for π_τ can be constructed by setting $\boldsymbol{\pi}^*_{\boldsymbol{LB}} = (\pi_L, \widehat{\pi_2}, \ldots, \widehat{\pi_k})$ and $\boldsymbol{\pi}^*_{\boldsymbol{UB}} = (\pi_U, \widehat{\pi_2}, \ldots, \widehat{\pi_k})$, where $\widehat{\pi_2} = \mathrm{X}_2/\mathrm{n}_1$, etc.

UEL: CONFIDENCE INTERVAL

Continuing with the previous example:

Let $\pi_0 = 0.1, n_1 = n_2 = 15$, and $k = 3$. Suppose that $\tau = 1$ and $Z_\tau = 6$, and that Stage 1 responses of 2 were observed for both of the non-selected arms ($X_2 = X_3 = 2$). A 95% CI for π_1 can be calculated as follows:

- The lower bound $\boldsymbol{\pi_L = 0.057}$ satisfies $P\left[Z_\tau \geq 6 \middle| \tau = 1, \boldsymbol{\pi} = \left(\pi_L, \frac{2}{15}, \frac{2}{15}\right)\right] = 0.025.$
- The upper bound $\boldsymbol{\pi_U = 0.370}$ satisfies $P\left[Z_\tau \leq 6 \middle| \tau = 1, \boldsymbol{\pi} = \left(\pi_U, \frac{2}{15}, \frac{2}{15}\right)\right] = 0.025.$
- The R function in `DropTheLosersUEL` for the confidence interval can be found below:

```
> ci_ztau(resp.vec=c(6,2,2), tau=1, n1=15, n2=15, two.sided=TRUE, alpha=0.05)
[1] 0.0567627 0.3701172
```

UEL: POINT ESTIMATION

A point estimator based on the median of this confidence interval was also investigated

- Similar properties to the UMVUE; slightly less variability, but modest bias when an “incorrect” treatment is selected

In our opinion, the **UMVUE from Tappin (1992) suffices as a point estimator** in this setting:

- Theoretical justification
- Closed-form expression
- Intuitive appeal of unbiasedness (even conditionally on a poor choice of treatment)

SIMULATIONS: SETUP

Simulations were performed to compare operating characteristics of the proposed methods to the following existing methods:

- Naïve exact binomial
- Posch *et al.* (2005) with Šidak multiplicity adjustment
- Sill & Sampson (2009) modified for the uncontrolled case.

The comparisons were made based on:

- Type 1 error
- Power (both probability to reject based on any treatment *and* to reject based on the truly best treatment)
- Coverage probability of 95% CI
- Mean width of 95% CI

SIMULATIONS: SETUP

Power calculations for the UEL test were performed analytically using an R function in `DropTheLosersUEL`:

```
> pwr_ztau(pi0=0.1, ntrts=3, pi.alt=c(0.4,0.3,0.2), n1=12, n2=12)
$`Critical Value`
[1] 7

$Power
[1] 0.9213124

$`Power by Treatment`
[1] 0.69753869 0.19227586 0.03149788
```

Probability to reject any treatment

Probability to reject the truly best treatment

Power calculations for the other tests (and all CI computations) were based on simulations with 10,000 repetitions.

The base case investigated was $n_1 = n_2 = 12$ and $\pi_0 = 0.1$, although these were varied in some instances. All simulations displayed use $k = 3$; results for $k = 2$ and $k = 4$ are in the backup slides.

SIMULATIONS: COMPARATOR METHODS

Naïve exact binomial

Conduct a single standard exact binomial test and CI, combining data from Stage 1 and Stage 2 without accounting for the selection step.

This incorrectly treats the naïve point estimate for π_τ, $\left(\frac{Z_\tau}{n_1+n_2}\right)$, as unbiased.

SIMULATIONS: COMPARATOR METHODS

Posch *et al.* (2005) with Šidak multiplicity adjustment

Closed testing approach using two-stage combination test, providing strong control of type 1 error:

- $\boldsymbol{p}$: Stage 1 *p*-value for selected treatment from one-sided exact binomial test vs. π_0
- $\boldsymbol{p^{adj}} = 1 - (1-p)^k$: Šidak-adjusted *p*-value; as easy to implement as Bonferroni but slightly more powerful
- $\boldsymbol{q}$: Stage 2 p-value for selected treatment from one-sided exact binomial test vs. π_0 (no multiplicity adjustment needed since only arm τ proceeds to Stage 2)
- Test based on combined p-value applying weighted inverse normal combination function to p^{adj} and q:

$$C\left(p^{adj}, q\right) = 1 - \phi\left[v\phi^{-1}\left(1 - \boldsymbol{p^{adj}}\right) + w\phi^{-1}(1-q)\right]$$

  - Often use $v^2 = \frac{n_1}{n_1+n_2}$ and $\mathrm{w}^2 = \frac{n_2}{n_1+n_2}$ (proportional to relative sample sizes in each stage)
- **Reject if $\boldsymbol{C\left(p^{adj}, q\right) < \alpha/2}$**

Confidence interval: contains all π such that the combination of stagewise exact binomial tests vs. π does not reject based on the observed data

SIMULATIONS: COMPARATOR METHODS

Modified Sill and Sampson (2009)

Define θ_τ and θ_0 as the log-odds of response under selected treatment τ and control. **Condition on $\boldsymbol{Q}$, the observed ordering of Stage 1 responses $\boldsymbol{X}$**. Let $\boldsymbol{X}^*$ contain all members of $\boldsymbol{X}$ except X_τ. For generic θ , define

$$f_Q(Z|\theta, \boldsymbol{X}^*) = \frac{\exp\{Z\theta\} \sum_{x=\max(L,Z-n_2)}^{\min(n_1,Z)} \binom{n_1}{x}\binom{n_2}{Z-x}}{\sum_{z=L}^{n_1+n_2} \exp\{z\theta\} \sum_{x=\max(L,z-n_2)}^{\min(n_1,z)} \binom{n_1}{x}\binom{n_2}{z-x}}$$

L represents the lower bound of the Stage 1 responses for τ. For the three-arm case:

$$L = \begin{cases} \max(X_2, X_3) \ \ if\ \tau = 1 \\ \max(X_1 + 1, X_3) \ \ if\ \tau = 2 \\ \max(X_1 + 1, X_2 + 1) \ \ if\ \tau = 3 \end{cases}$$

Then, **reject $\boldsymbol{H_0}$ if $\boldsymbol{Z_\tau \geq Z_U}$ where**

$$\alpha/2 \geq \sum_{z=Z_U}^{n_1+n_2} f_Q(z|\theta_0, \boldsymbol{X}^*)$$

Confidence interval : contains all θ such that substituting in for θ_0 above provide upper and lower tail probabilities above $\alpha/2$

SIMULATIONS: TYPE 1 ERROR

π_0	Total sample size	$n_1; n_2$	Type 1 error (%)			
			Naïve	Posch	S&S	UEL
0.1	48	12;12	2.2	0.4	1.6	2.0
	60	15;15	2.2	0.6	1.3	2.0
	100	25;25	6.6	0.7	1.6	2.4
0.3	48	12;12	3.3	0.9	1.8	1.0
	60	15;15	4.8	1.2	2.0	1.7
	100	25;25	3.6	1.2	2.1	1.5
0.5	48	12;12	2.9	1.0	0.8	0.9
	60	15;15	5.5	1.0	1.3	2.2
	100	25;25	4.0	1.1	1.4	2.1

$\pi_0 = \pi_1 = \pi_2 = \pi_3; \quad \alpha/2 = 0.025$

Nominal type 1 error: 2.5%

Naïve test fails to control type 1 error, due to its overestimation of π_τ. Not listed in remaining tables.

The other three tests **control type 1 error**

- Šidak-adjusted Posch generally more conservative than S&S or UEL
- No consistent trends across π_0 or sample size

SIMULATIONS: POWER (ALL EQUAL)

$\pi_0 = 0.1; \quad \alpha/2 = 0.025$

(π_1, π_2, π_3)	Total sample size	$n_1; n_2$	Probability to reject any treatment (%)			P(τ is the best treatment)	Prob to reject H0 for the best treatment (%)		
			Posch	S&S	UEL		Posch	S&S	UEL
(0.3,0.3,0.3)	48	8;24	81.3	71.8	**90.5**	--	--	--	--
		10;18	81.3	61.2	**82.4**	--	--	--	--
		12;12	69.2	48.6	**84.3**	--	--	--	--
		14;6	67.2	31.7	**69.5**	--	--	--	--
	60	15;15	81.5	62.3	**91.2**	--	--	--	--
	100	25;25	97.7	82.3	**99.0**	--	--	--	--

For these and subsequent scenarios, the most powerful test is highlighted in **purple**; if it outperforms the next-most-powerful test by ≥5%, it is highlighted in **green**.

In general:

a) **If one treatment clearly better, UEL ≥ S&S ≥ Posch**
b) **If multiple treatments equal or nearly equal, UEL ≥ Posch ≥ S&S**

SIMULATIONS: POWER (ONE TREATMENT BETTER)

$\pi_0 = 0.1; \quad \alpha/2 = 0.025$

(π_1, π_2, π_3)	Total sample size	$n_1; n_2$	Probability to reject any treatment (%)			P(τ is the best treatment)	Prob to reject H0 for the best treatment (%)		
			Posch	S&S	UEL		Posch	S&S	UEL
(0.4,0.1,0.1)	48	8;24	87.9	90.9	**92.2**	93.7	87.8	90.9	**91.9**
		10;18	86.5	88.9	**89.5**	95.3	86.3	88.9	**89.3**
		12;12	78.3	86.1	**88.9**	96.5	78.2	86.1	**88.5**
		14;6	72.3	77.2	**74.8**	97.4	72.2	77.1	**74.6**
	60	15;15	89.4	92.0	**94.4**	97.8	89.3	92.0	**94.1**
	100	25;25	98.9	99.1	**99.4**	99.5	98.9	99.1	**99.3**
(0.1,0.1,0.4)	48	8;24	79.3	79.3	**81.3**	81.6	79.3	79.3	**80.5**
		10;18	82.2	81.0	**82.8**	86.7	82.2	81.0	**82.4**
		12;12	76.4	79.1	**84.9**	90.3	76.4	79.1	**84.3**
		14;6	71.3	69.0	**73.7**	92.9	71.3	69.0	**73.5**
	60	15;15	87.5	86.9	**91.5**	93.9	87.4	86.8	**91.1**
	100	25;25	98.3	98.2	**98.7**	98.6	98.3	98.2	**98.5**

SIMULATIONS: POWER (MULTIPLE TREATMENTS SIMILAR)

$\pi_0 = 0.1; \quad \alpha/2 = 0.025$

(π_1, π_2, π_3)	Total sample size	$n_1; n_2$	Probability to reject any treatment (%)			P(τ is the best treatment)	Prob to reject H0 for the best treatment (%)		
			Posch	S&S	UEL		Posch	S&S	UEL
(0.4,0.3,0.3)	48	8;24	93.0	85.5	**96.8**	62.7	60.0	58.6	**62.1**
		10;18	92.8	78.4	**93.5**	63.6	60.5	56.5	**61.6**
		12;12	85.6	67.0	**94.2**	64.6	57.6	50.3	**62.3**
		14;6	82.8	48.5	**85.7**	65.6	56.4	39.0	**58.1**
	60	15;15	93.4	80.2	**97.5**	66.1	63.5	58.9	**65.3**
	100	25;25	99.7	92.6	**99.9**	70.9	70.0	68.8	**70.9**
(0.3,0.3,0.4)	48	8;24	89.0	81.0	**94.6**	40.8	**40.6**	38.9	**40.6**
		10;18	89.8	72.7	**90.3**	44.0	**43.2**	38.7	**43.2**
		12;12	82.6	59.9	**92.0**	46.7	43.5	35.1	**45.8**
		14;6	80.6	41.2	**83.6**	49.1	43.7	26.5	**45.6**
	60	15;15	91.3	75.7	**96.3**	50.2	48.6	44.2	**49.9**
	100	25;25	99.4	90.5	**99.8**	59.2	**59.4**	58.2	59.1

$\pi_0 = 0.1; \quad \alpha/2 = 0.025$

SIMULATIONS: POWER (MULTIPLE TREATMENTS SIMILAR, CONT'D)

(π_1, π_2, π_3)	Total sample size	$n_1; n_2$	Probability to reject any treatment (%)			P(τ is the best treatment)	Prob to reject H0 for the best treatment (%)		
			Posch	S&S	UEL		Posch	S&S	UEL
(0.4,0.3,0.2)	48	8;24	90.4	85.5	**94.9**	70.4	68.0	67.4	**69.6**
		10;18	90.5	80.2	**91.3**	71.7	67.8	64.8	**69.0**
		12;12	82.7	70.7	**92.1**	73.0	64.0	59.1	**69.8**
		14;6	79.8	53.9	**81.9**	74.3	61.9	47.2	**63.6**
	60	15;15	91.6	81.6	**96.2**	74.9	71.4	67.2	**73.6**
	100	25;25	99.2	94.0	**99.7**	79.8	79.2	78.2	**79.8**

UEL test is as or more powerful than Posch, Sill and Sampson [S&S] tests in almost all scenarios examined

SIMULATIONS: CONFIDENCE INTERVALS

(π_1, π_2, π_3)	Coverage probability of 95% CI for π_τ (%)			Mean width of 95% CI for π_τ		
	Posch	S&S	UEL	Posch	S&S	UEL
(0.1,0.1,0.1)	99.6	98.4	98.3	0.388	0.355	0.290
(0.3,0.3,0.3)	99.1	97.6	97.8	0.499	0.447	0.413
(0.4,0.1,0.1)	99.6	99.0	96.8	0.501	0.422	0.408
(0.1,0.1,0.4)	99.6	99.0	97.4	0.497	0.426	0.408
(0.4,0.3,0.3)	99.3	98.0	97.8	0.508	0.454	0.426
(0.3,0.3,0.4)	99.1	97.9	98.0	0.506	0.458	0.426
(0.4,0.3,0.2)	99.3	98.2	97.5	0.506	0.446	0.421

$\pi_0 = 0.1; \quad n_1 = n_2 = 12$

In the scenarios displayed here, the 95% CIs based on the UEL framework:

- **Achieve >95% coverage**, as do Posch and S&S
- On average are the **narrowest** across the three methods

PRACTICAL CONSIDERATIONS

The following slides investigate design-related and other practical considerations for drop-the-losers designs in this context:

- **Choice of n_1 and n_2** for fixed total sample size
- Persistent discreteness and the **normal approximation** to the UEL test
- Robustness to **deviations from the pre-specified treatment selection** rule

CHOICE OF n_1 AND n_2

The choice of n_1 and n_2 is not trivial and can notably impact power, even for a fixed total sample size.

(π_1, π_2, π_3)	Total sample size	$n_1; n_2$	Power to reject any treatment (%)		
			Posch	S&S	UEL
(0.4,0.3,0.2)	48	8;24	90.4	85.5	**94.9**
		10;18	90.5	80.2	**91.3**
		12;12	82.7	70.7	**92.1**
		14;6	79.8	53.9	**81.9**

`DropTheLosersUEL` contains a tool to help trialists decide how to allocate sample size to the two stages, by plotting power and the probability of choosing a desired treatment across different combinations. This function produces plots similar to those on the next slide.

```
> plot_power_pselecttrt(pi0=0.1, pi=c(0.4,0.1,0.1), tau=1, totalN=48)
```

$\pi_0 = 0.1; \alpha/2 = 0.025$; Total sample size: 48

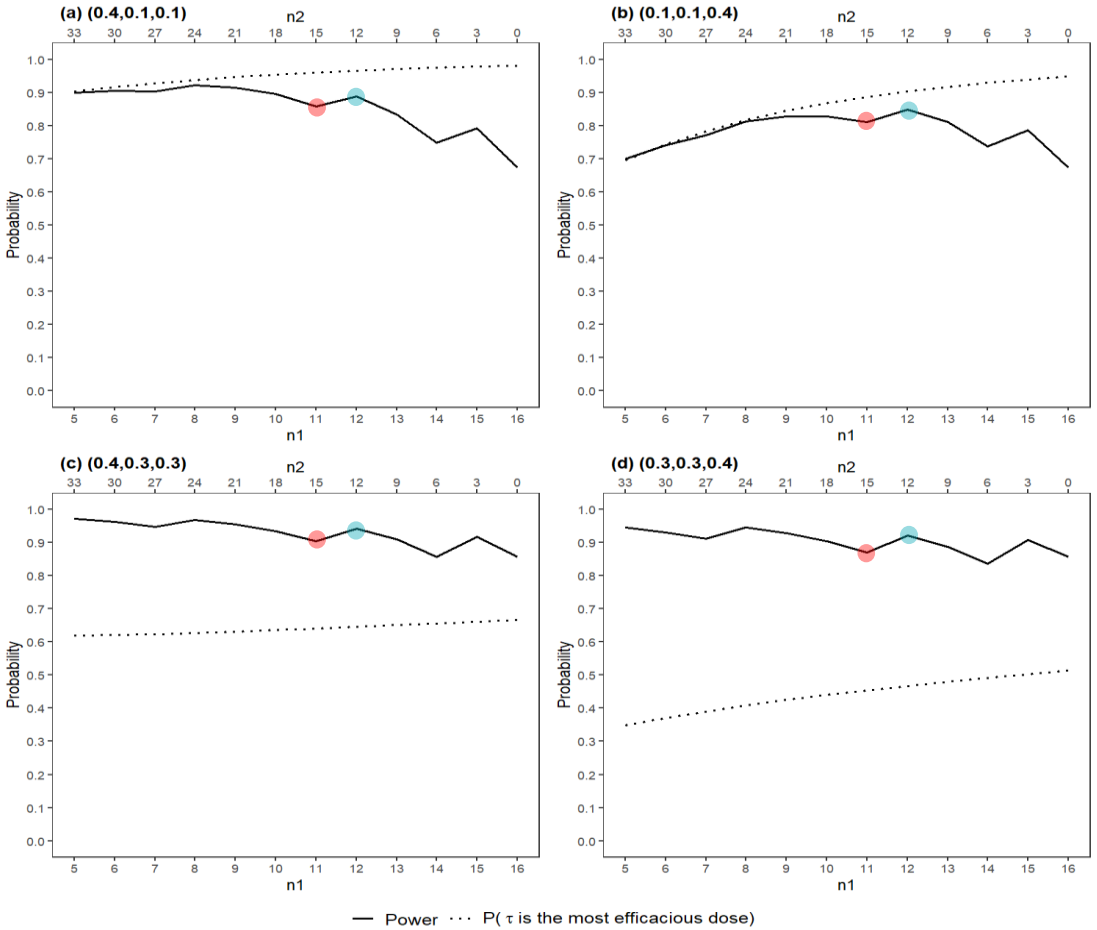
CHOICE OF n_1 AND n_2

Discreteness in small sample sizes leads to jagged overall power curves.

Consistent spikes/dips across scenarios: e.g., $n_1 = n_2 = 12$ outperforms $n_1 = 11; n_2 = 15$.

Chance of choosing most efficacious dose to proceed to Stage 2 (dotted line) increases with n_1.

Overall study power (solid line) increases with n_1 *to a point*. Overly large values of n_1 drive down $n_1 + n_2$ for fixed sample size, lowering power.

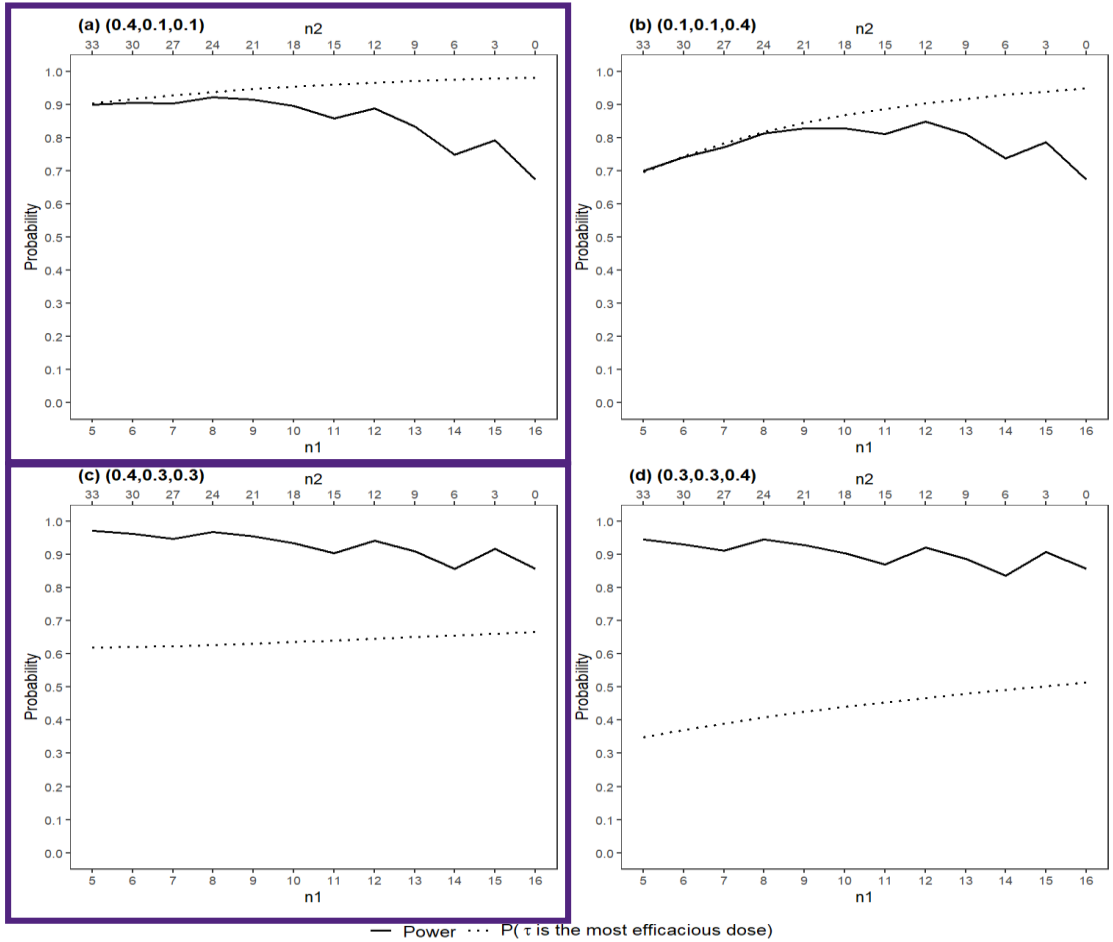


$\pi_0 = 0.1; \alpha/2 = 0.025$; Total sample size: 48

CHOICE OF n_1 AND n_2

If the most efficacious treatment is preferred in ties, τ tends to be selected correctly even if n_1 is small.

The choice of n_1 and n_2 is driven more by the power curve.



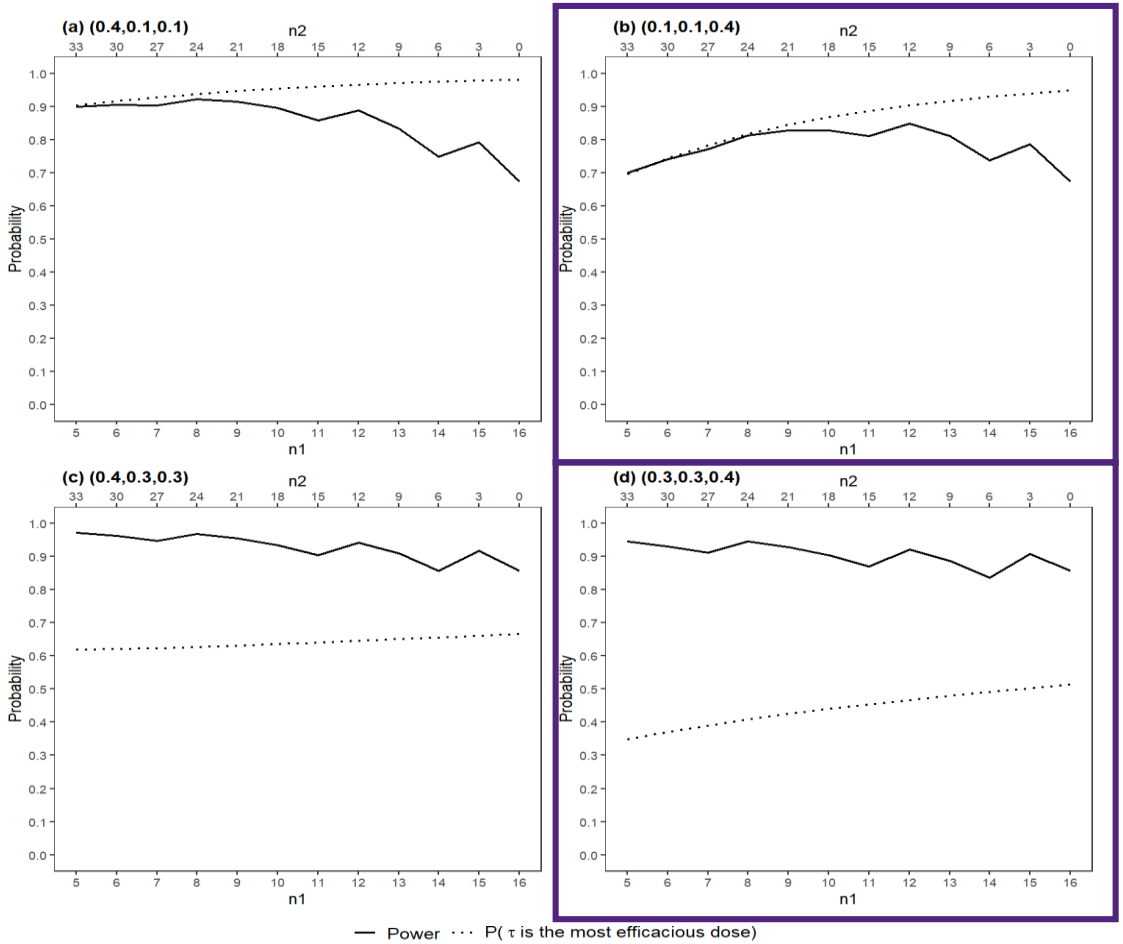
$\pi_0 = 0.1; \alpha/2 = 0.025$; Total sample size: 48

CHOICE OF n_1 AND n_2

If the most efficacious treatment is preferred in ties, τ tends to be selected correctly even if n_1 is small.

The choice of n_1 and n_2 is driven more by the power curve.

If the most efficacious treatment is not preferred in ties, the higher chance of ties for small n_1 leads to a greater probability of carrying forward a worse treatment.



$\pi_0 = 0.1; \alpha/2 = 0.025$; Total sample size: 48

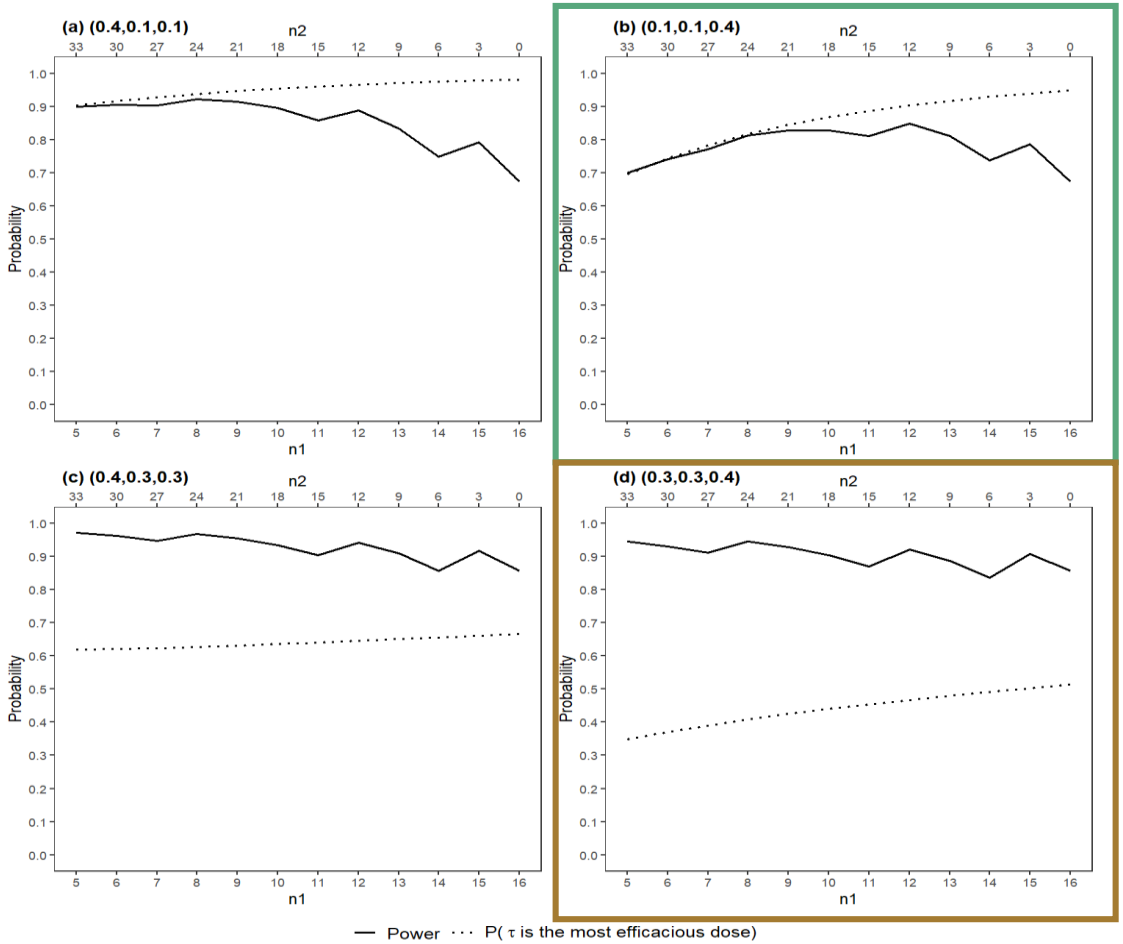
CHOICE OF n_1 AND n_2

If the most efficacious treatment is preferred in ties, τ tends to be selected correctly even if n_1 is small.

The choice of n_1 and n_2 is driven more by the power curve.

If the most efficacious treatment is not preferred in ties, the higher chance of ties for small n_1 leads to a greater probability of carrying forward a worse treatment.

- If one treatment arm is considerably superior, a smaller n_1 may still suffice.
- If all arms are reasonably efficacious, a larger n_1 is needed to distinguish the best arm.



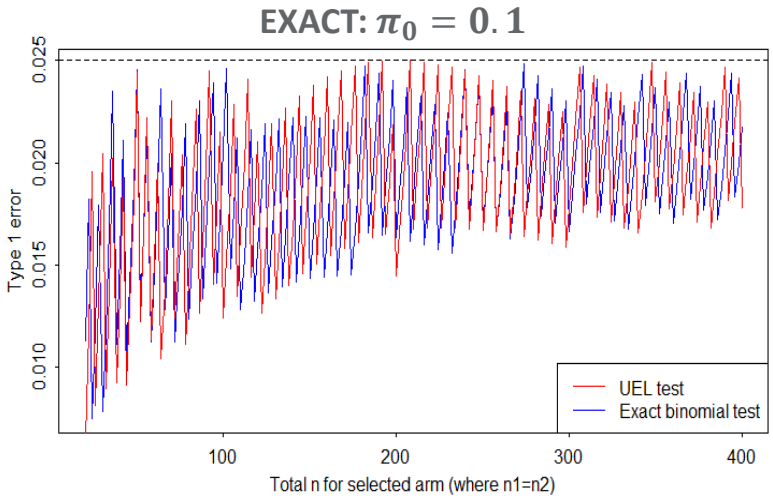
CHOICE OF n_1 AND n_2

Recommendations

In designing a drop-the-losers trial in the rare disease setting, trialists should:

- Check overall study power to ensure that the allocation of n_1 and n_2 represents a spike in power and not a dip
- Decide on a clinically meaningful difference in response rate
  - Difference Δ at which it is important to be adequately powered to pick the most efficacious dose with response rate π over another treatment with a non-null response rate $\pi - \Delta$

NORMAL APPROXIMATION



Discreteness persistently affects the type 1 error, even for large sample sizes ($n_1 = n_2 = 200$).

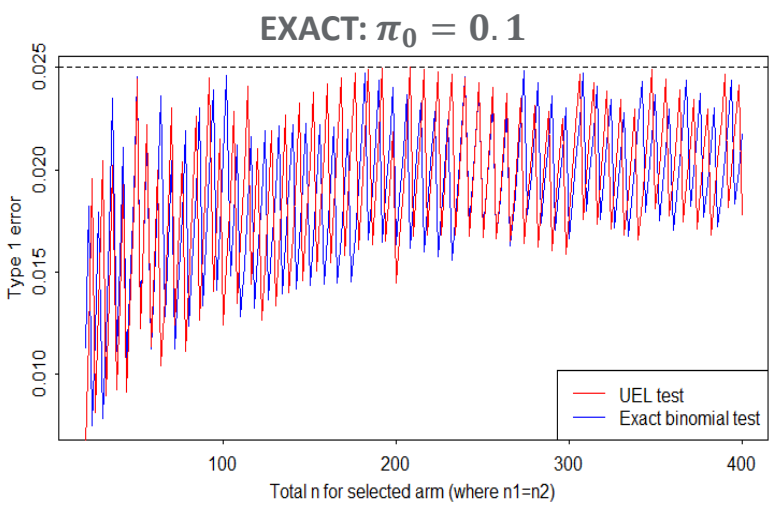
Exact binomial known to be conservative; same is true for UEL test.

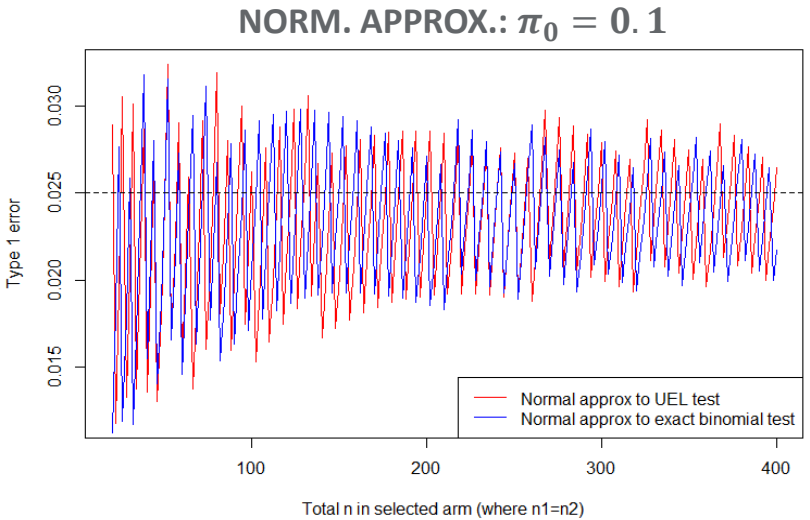
Normal approximation to the UEL test can be developed:

- Mean (μ_τ) and variance (σ_τ^2) of Z_τ under the global null can be computed from its distribution (implemented in `DropTheLosersUEL`)
- Let Z_τ^* be the normal analog of Z_τ under the global null. $Z_\tau^* \sim N(\mu_\tau, \sigma_\tau^2)$
- The test rejects if $\mathbf{P[Z_\tau^* \geq (z - 0.5) | \pi = (\pi_0, \ldots, \pi_0)] < \alpha/2}$

Confidence intervals can also be constructed based on continuity-corrected Z_τ^*

NORMAL APPROXIMATION





The normal approximations to the UEL test and exact binomial test display similar behavior, demonstrating approximate type 1 error control under $\pi_0 = 0.1$.

- Norm. approx. to UEL test: type 1 error > 0.03 rare for $\pi_0 \geq 0.05$ and never occurs for $\pi_0 \geq 0.2$

Standard rule of thumb for normal approximation to the binomial can be adapted to this setting:
$(n_1 + n_2)\pi_0 \geq 5$ and $(n_1 + n_2)(1 - \pi_0) \geq 5$.

NORMAL APPROXIMATION

$\pi_0 = 0.1; \quad \alpha/2 = 0.025$

(π_1, π_2, π_3)	$n_1; n_2$	Type 1 error/Power to reject any treatment (%)	
		UEL	Norm. approx.
(0.1,0.1,0.1)	30;30	1.5	1.5
	32;32	1.0	2.6
(0.3,0.1,0.1)	30;30	92.9	92.9
	32;32	93.1	95.6
(0.1,0.1,0.3)	30;30	90.7	90.7
	32;32	91.3	93.6
(0.3,0.2,0.2)	30;30	93.4	93.4
	32;32	93.1	96.3
(0.2,0.2,0.3)	30;30	90.6	90.6
	32;32	90.3	94.4
(0.3,0.2,0.1)	30;30	93.1	93.1
	32;32	93.0	96.0

For large enough sample sizes, the normal approximation is on average more powerful than the UEL test.

Individual cases may vary

- $n_1 = n_2 = 30$: UEL and normal approximation equally powerful
- $n_1 = n_2 = 32$: up to 4% gain in power from normal approximation

The p-value, critical value, power, and CI functions in `DropTheLosersUEL` contain a `norm.approx=TRUE` option.

DEVIATIONS FROM SELECTION RULE

Recall: Null distribution of Z_τ assumes that treatment with highest number of responses at Stage 1 is selected for Stage 2

- In case of ties, preference is Treatment 1 over Treatment 2 over Treatment 3

In practice, the team may wish to select a different treatment for Stage 2 (e.g., due to safety, manufacturing concerns).

Simulation performed to assess performance of UEL procedures when the "second-place" treatment is chosen if its observed Stage 1 response rate is within 10% of the "first-place" treatment $[n_1 = n_2 = 12]$

Conclusion: UEL procedures robust to modest deviations from the pre-specified selection rule

- Some loss in power, as expected; most notably for (0.4,0.1,0.1)
- 95% CIs maintain nominal coverage

DEVIATIONS FROM SELECTION RULE

(π_1, π_2, π_3)	Type 1 error/ Probability to reject any treatment (%)		Coverage probability of 95% CI for π_τ (%)		Mean width of 95% CI for π_τ	
	No violation	10% violation	No violation	10% violation	No violation	10% violation
(0.1,0.1,0.1)	2.0	1.8	98.3	97.1	0.290	0.254
(0.3,0.3,0.3)	84.3	77.7	97.8	96.5	0.413	0.402
(0.4,0.1,0.1)	88.9	76.1	96.8	97.4	0.408	0.384
(0.1,0.1,0.4)	84.9	80.3	97.4	97.1	0.408	0.397
(0.4,0.3,0.3)	94.2	87.5	97.8	97.1	0.426	0.414
(0.3,0.3,0.4)	92.0	88.1	98.0	96.5	0.426	0.419
(0.4,0.3,0.2)	92.1	82.3	97.5	96.7	0.421	0.406

$\pi_0 = 0.1;\ \ n_1 = n_2 = 12;\ \ \alpha/2\ = 0.025$

FUTURE DIRECTIONS: CONTROLLED CASE

While the uncontrolled case is realistic for the rare disease setting, it may be useful to extend the UEL framework to the controlled case. However, there are several complicating factors:

- Specific null hypothesis (i.e., the value of all π) needs to be chosen to construct the null distribution of $Z_\tau - Z_0$ (the difference in responses between the selected arm and the control)
  - Most conservative case (all $\pi = 0.5$) led to an unacceptable loss of power, while all $\pi = \widehat{\pi_0}$ failed to control the type 1 error
  - One option: fit reduced model s.t. $\pi_0 = \pi_1 = \cdots = \pi_k$; derive null distribution of $Z_\tau - Z_0$ based on MLE, but computing MLE may not be straightforward
- Constructing CI for difference in proportion based on $Z_\tau - Z_0$ is challenging
  - $Z_\tau - Z_0$ will have different distribution for $\pi_0 = 0.1; \pi_\tau = 0.3$ compared to $\pi_0 = 0.3; \pi_\tau = 0.5$. Also depends on the non-selected treatment arms.

SUMMARY

- Proposed **unconditional exact likelihood (UEL) procedure** for hypothesis testing and inference for drop-the-losers designs for a binary endpoint *in the absence of a control arm*
  - Based on distribution of Z_τ (total # responses in Stage 1 and 2 for selected arm) under global null
  - Performance vs. Šidak-adjusted Posch, modified Sill and Sampson:
    - UEL test **controls type 1 error** and is **as or more powerful** than other methods
    - CIs based on UEL test **achieve nominal coverage** and **tend to be more narrow** than CIs based on other methods
- Explored additional design and analysis considerations for settings where UEL could be applied
  - Stage 1 and Stage 2 sample size affect overall power and probability of selecting correct arm for Stage 2
  - Normal approximation to UEL test is available
  - UEL test robust to modest violations of prespecified selection rule

Manuscript submitted to *Statistics in Biopharmaceutical Research*

Design and Analysis of Drop-the-Losers Studies using Binary Endpoints in the Rare Disease Setting

The drop-the-losers (DtL) design was introduced as an adaptive approach combining a phase 2 and phase 3 trial under a single protocol, thereby gaining efficiency over a traditional clinical development approach. Such a design assesses k treatments in a first stage, with the treatment exhibiting the greatest efficacy selected to continue to a second stage, at the end of which the null hypothesis is tested based upon the selected arm. These designs may be particularly useful in the rare disease setting, where conserving sample size is paramount. We propose an unconditional exact likelihood (UEL) testing and inference procedure for these designs for a binary endpoint, using small sample sizes. The operating characteristics of this approach are compared to existing testing and confidence interval construction approaches in the literature. Special considerations for small sample sizes are evaluated, including the choice of first and second stage sample sizes and the effect of ties.

Keywords: adaptive design; interim analysis; drop-the-losers design; two-stage design; rare disease; binary endpoint; small sample size

R package posted to GitHub

Package 'DropTheLosersUEL'

May 26, 2020

Type Package

Title Implementation of unconditional exact likelihood (UEL) testing and inference procedure for drop-the-losers designs for binary endpoints in the rare disease setting

Version 0.1.0

Author Ina Jazic, Xiaoyan Liu, Glen Laird

Maintainer Ina Jazic <Ina_Jazic@vrtx.com>

Description The drop-the-losers design was proposed to establish efficacy for one of several treatment arms more efficiently than the traditional development paradigm of separate studies for treatment selection and confirmation. A pre-specified number of patients receive each treatment in Stage 1, at the end of which the experimental treatment exhibiting the most efficacy proceeds to Stage 2, after which a hypothesis test of efficacy is conducted incorporating data from both stages.

REFERENCES

Bauer, P., and Kieser, M. (1999), “Combining different phases in the development of medical treatments within a single trial,” *Statistics in Medicine*, 18, 1833–1848.

Posch, M., Koenig, F., Branson, M., Brannath, W., Dunger-Baldauf, C., and Bauer, P. (2005), “Testing and estimation in flexible group sequential designs with adaptive treatment selection,” *Statistics in Medicine*, 24, 3697–3714.

Sampson, A. R., and Sill, M. W. (2005), “Drop-the-Losers Design: Normal Case,” *Biometrical Journal*, 47, 257–268.

Sill, M. W., and Sampson, A. R. (2009), “Drop-the-losers design: Binomial case,” *Computational Statistics & Data Analysis*, 53, 586–595.

Tappin, L. (1992), “Unbiased estimation of the parameter of a selected binomial population,” *Communications in Statistics - Theory and Methods*, Taylor & Francis, 21, 1067–1083.

BACKUP

POINT ESTIMATION: UMVUE

Additional notation

- Let $i_1, \dots, i_k$ denote the indices of X_j (Stage 1 responses) when sorted. In case of a tie, the smaller index comes first. Selected dose $\tau = i_1$.
- Let $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_k, Z_{k+1}, \dots, Z_{2k}) = (X_{i_1} + Y, X_{i_2}, \dots, X_{i_k}, i_1, \dots, i_k)$
  - Note: this is the complete/sufficient statistic for the vector $\boldsymbol{\pi}$ used in the UMVUE (Tappin 1992)

UMVUE (Tappin 1992):

$$\widehat{\pi_\tau} = \begin{cases} \dfrac{Z_1}{n_1 + n_2} \; if \; Z_1 - Z_2 > n_2 \\ \dfrac{1}{n_2} \dfrac{\sum_{y=0}^{\boldsymbol{z_1 - z_2}} y \binom{n_2}{y} \binom{n_1}{z_1 - y}}{\sum_{y=0}^{\boldsymbol{z_1 - z_2}} \binom{n_2}{y} \binom{n_1}{z_1 - y}} \; if \; Z_1 - Z_2 \le n_2 \; and \; \boldsymbol{i_1 < i_2} \\ \dfrac{1}{n_2} \dfrac{\sum_{y=0}^{\boldsymbol{z_1 - z_2 - 1}} y \binom{n_2}{y} \binom{n_1}{z_1 - y}}{\sum_{y=0}^{\boldsymbol{z_1 - z_2 - 1}} \binom{n_2}{y} \binom{n_1}{z_1 - y}} \; if \; Z_1 - Z_2 \le n_2 \; and \; \boldsymbol{i_1 > i_2} \end{cases}$$

Note: $\widehat{\pi_\tau^U} = 0$ if $Z_1 - Z_2 = 0$ (for $i_1 < i_2$) or $Z_1 - Z_2 - 1 = 0$ (for $i_1 > i_2$), even if Stage 1 responses were observed. Tappin (1992) does not give an expression for the variance.

SILL AND SAMPSON (2009): FULL DERIVATION

Let $\boldsymbol{X}^*$ be all Stage 1 responses besides that of treatment τ. Let the Stage 2 response be $Y \sim Bin(n_2, \pi_\tau)$. Let the combined Stage 1 and Stage 2 responses for the control arm be $Y_0 \sim Bin(n_0, \pi_0)$, where n_0 is the total number of subjects assigned to control in both stages. Define $Z = X_\tau + Y$ and define $T = Z + Y_0$. Let $\theta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$. Various alternatives can be specified through different values of $\Delta_\tau = \theta_\tau - \theta_0$.

Let Q be the ordering of $X_1, \dots, X_k$ after Stage 1 – all expressions below are conditional on this ordering. Let L denote a potential lower bound for X_τ in the equations below. If $\tau = 1$, then $L = \max\{X_2, \dots, X_k\}$. If $\tau = k$, then $L = \max\{X_1 + 1, \dots, X_{k-1} + 1\}$. For other values of τ, $L = \max\{X_1 + 1, \dots, X_{\tau-1} + 1,\ X_{\tau+1}, \dots, X_k\}$.

The joint distribution of $\boldsymbol{X}^*$ and Z can be written as:

$$f_Q(\boldsymbol{X}^*, Z) = \left[K(\pi)^{-1} I_Q^*(\boldsymbol{X}^*)\Pi\right] \pi_\tau^Z (1-\pi_\tau)^{n_1+n_2-z} \sum_{x=\max(L, Z-n_2)}^{\min(n_1, Z)} \binom{n_1}{x}\binom{n_2}{Z-x}$$

$$= K^* \pi_\tau^Z (1-\pi_\tau)^{n_1+n_2-Z} h(Z) = K^* (1-\pi_\tau)^{n_1+n_2} \exp\{Z\theta_\tau\}\, h(Z)$$

where $K(\pi)$ is the probability of observing ordering Q (does not need to be computed), $I_Q^*(\boldsymbol{X}^*)$ is the indicator function for $\boldsymbol{X}^*$ on Q, and $\Pi = \prod_{i \neq \tau} \binom{n_1}{X_i} \pi_i^{X_i} (1-\pi_i)^{n_1 - X_i}$.

SILL AND SAMPSON (2009): FULL DERIVATION

Since the responses for the control are independent of the responses for the other treatments, the joint distribution can be easily expanded to accommodate Y_0:

$$f_Q(\boldsymbol{X}^*, Z, Y_0) = f_Q(\boldsymbol{X}^*, Z)\binom{n_0}{Y_0}\pi_0^{Y_0}(1-\pi_0)^{n_0-Y_0} = K^*(1-\pi_\tau)^{n_1+n_2}\binom{n_0}{Y_0}(1-\pi_0)^{n_0}h(Z)\exp\{Z\theta_\tau + Y_0\theta_0\}$$

To reframe in terms of $T = Z + Y_0$, such that Z is sufficient for Δ_τ:

$$f_Q(\boldsymbol{X}^*, Z, T) = K^*(1-\pi_\tau)^{n_1+n_2}\binom{n_0}{T-Z}(1-\pi_0)^{n_0}h(Z)\exp\{Z\Delta_\tau + T\theta_0\}$$

Then, the conditional distribution used to define the UMPCU test is the following, where the denominator can be obtained by marginalizing $f_Q(\boldsymbol{X}^*, Z, T)$ over possible values of Z.

$$f_Q(Z|\Delta_\tau, \boldsymbol{X}^*, T) = \frac{f_Q(\boldsymbol{X}^*, Z, T)}{f_Q(\boldsymbol{X}^*, T)} = \frac{\binom{n_0}{T-Z}\exp\{Z\Delta_\tau\}\sum_{x=\max(L,Z-n_2)}^{\min(n_1,Z)}\binom{n_1}{x}\binom{n_2}{Z-x}}{\sum_{z=\max(L,T-n_0)}^{\min(n_1+n_2,T)}\left[\binom{n_0}{T-z}\exp\{z\Delta_\tau\}\sum_{x=\max(L,z-n_2)}^{\min(n_1,z)}\binom{n_1}{x}\binom{n_2}{z-x}\right]}$$

SILL AND SAMPSON (2009): MODIFICATION FOR UNCONTROLLED CASE

Since there is no control arm, a modified version of the Sill and Sampson (2009) test can simply proceed from $f_Q(\boldsymbol{X}^*, Z)$ given above, where $f_Q(\boldsymbol{X}^*)$ can be obtained by marginalizing $f_Q(\boldsymbol{X}^*, Z)$ over possible values of Z.

$$f_Q(Z|\theta_\tau, \boldsymbol{X}^*) = \frac{f_Q(\boldsymbol{X}^*, Z)}{f_Q(\boldsymbol{X}^*)} = \frac{\exp\{Z\theta_\tau\}\sum_{x=\max(L,Z-n_2)}^{\min(n_1,Z)}\binom{n_1}{x}\binom{n_2}{Z-x}}{\sum_{z=L}^{n_1+n_2}\left[\exp\{z\theta_\tau\}\sum_{x=\max(L,z-n_2)}^{\min(n_1,z)}\binom{n_1}{x}\binom{n_2}{z-x}\right]}$$

SIMULATION RESULTS: K=2

(π_1, π_2)	Prb of rejecting any dose (%)			Prb of rejecting the best dose (%)			Coverage probability (%)			Width of CI		
	Posch	S&S	UEL	Posch	S&S	UEL	Posch	S&S	UEL	Posch	S&S	UEL
(0.1,0.1)	0.8	1.2	1.4	-	-	-	99.2	98.8	98.7	0.355	0.354	0.283
(0.3,0.3)	71.3	54.5	77.0	-	-	-	99.3	97.6	97.4	0.468	0.435	0.402
(0.4,0.1)	83.9	87.9	89.6	83.7	87.9	88.8	99.1	99.1	96.6	0.478	0.419	0.405
(0.1,0.4)	83.3	84.1	87.2	83.1	84.1	87.6	99.2	99.0	97.0	0.475	0.424	0.407
(0.4,0.3)	89.1	74.7	92.7	69.6	63.3	72.0	99.1	98.3	97.5	0.481	0.444	0.417
(0.3,0.4)	87.6	69.9	90.4	59.5	50.5	60.7	99.1	98.1	97.7	0.480	0.448	0.418

SIMULATION RESULTS: K=4

$(\pi_1, \pi_2, \pi_3, \pi_4)$	Prb of rejecting any dose (%)			Prb of rejecting the best dose (%)			Coverage probability (%)			Width of CI		
	Posch	S&S	UEL	Posch	S&S	UEL	Posch	S&S	UEL	Posch	S&S	UEL
(0.1,0.1,0.1,0.1)	0.7	1.7	2.5	-	-	-	99.3	98.3	98.4	0.411	0.350	0.293
(0.3,0.3,0.3,0.3)	75.0	44.0	88.4	-	-	-	98.9	96.7	98.1	0.520	0.451	0.420
(0.4,0.1,0.1,0.1)	78.8	85.4	88.2	78.6	85.2	87.8	99.4	98.8	96.9	0.518	0.424	0.410
(0.1,0.1,0.1,0.4)	75.4	75.3	83.0	75.1	75.2	82.4	99.4	99.0	97.6	0.512	0.425	0.408
(0.4,0.3,0.3,0.3)	88.0	62.9	95.3	52.1	43.7	54.8	99.0	97.2	98.0	0.527	0.459	0.432
(0.3,0.3,0.3,0.4)	84.4	52.1	93.4	35.8	27.7	36.9	98.9	96.7	98.0	0.525	0.460	0.431

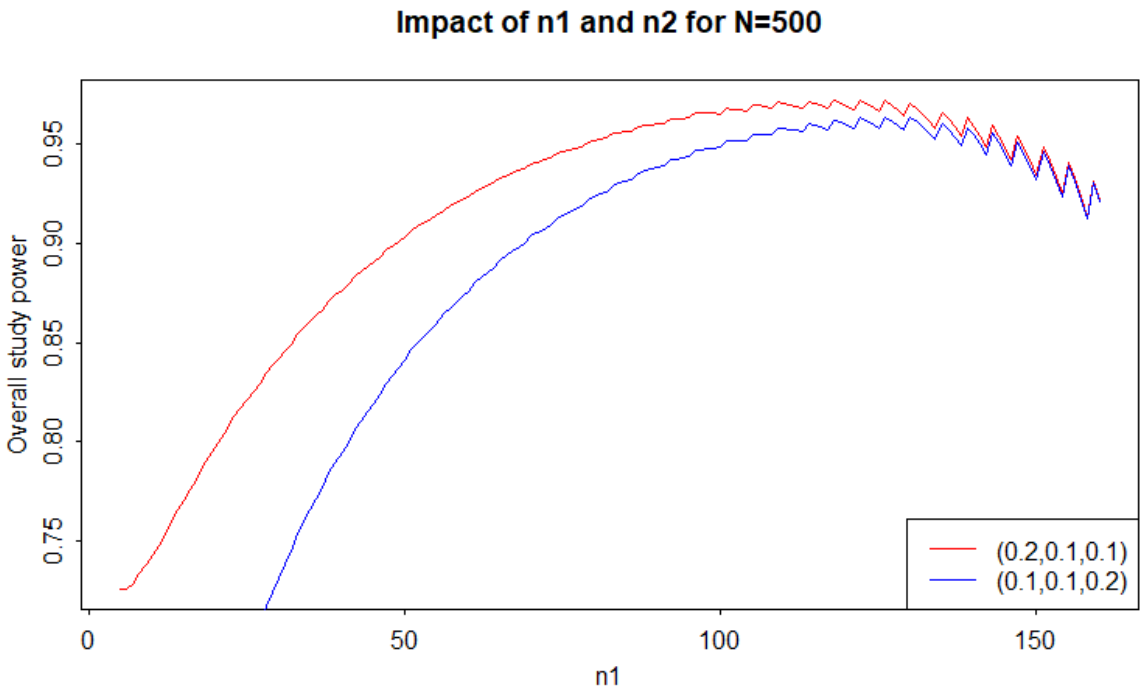
SIMULATION RESULTS: HIGHER SAMPLE SIZES

(π_1, π_2, π_3)	Total sample size	$n_1; n_2$	Probability of rejecting any dose (%)			Probability of rejecting the best dose (%)		
			Posch	S&S	UEL	Posch	S&S	UEL
(0.3,0.1,0.1)	80	20;20	74.6	76.6	78.6	74.4	76.7	78.3
	100	25;25	83.5	86.8	90.4	83.4	86.8	89.8
(0.1,0.1,0.3)	80	20;20	72.7	69.1	75.2	72.3	68.9	74.8
	100	25;25	82.2	81.6	87.8	82.0	81.4	86.6
(0.3,0.2,0.2)	80	20;20	78.7	61.9	81.0	62.0	54.1	63.9
	100	25;25	85.8	73.3	92.1	69.2	64.1	71.7
(0.2,0.2,0.3)	80	20;20	74.6	54.0	76.4	52.2	42.6	52.9
	100	25;25	81.7	65.5	89.0	58.9	52.5	60.5
(0.3,0.2,0.1)	80	20;20	77.1	67.3	79.9	67.3	62.3	70.0
	100	25;25	84.9	78.0	91.4	75.2	72.2	79.2

SIMULATION RESULTS: ALTERNATIVE NUISANCE PARAMETER STRATEGIES

(π_1, π_2, π_3)	Coverage probability for 95% UEL CIs under various strategies for nuisance parameters (%)			
	True values ("oracle")	Observed rates (method in main manuscript)	10% less than observed rates	10% greater than observed rates
(0.1,0.1,0.1)	97.7	98.3	97.6	98.1
(0.3,0.3,0.3)	96.7	98.1	97.8	97.8
(0.4,0.1,0.1)	97.3	97.2	97.1	96.5
(0.1,0.1,0.4)	97.8	97.2	97.5	96.1

DESIGN CONSIDERATIONS: POWER FOR TOTAL N=500



IMPACT OF SELECTION RULE VIOLATIONS ON TEST AND CI

(π_1, π_2, π_3)	Type 1 error/ Probability to reject any treatment (%)			Coverage probability of 95% CI for π_τ (%)			Mean width of 95% CI for π_τ		
	No violation	Tie violation	10% violation	No violation	Tie violation	10% violation	No violation	Tie violation	10% violation
(0.1,0.1,0.1)	2.0	2.2	1.8	98.3	97.7	97.1	0.290	0.284	0.254
(0.3,0.3,0.3)	**84.3**	**84.3**	77.7	97.8	97.6	96.5	0.413	0.412	0.402
(0.4,0.1,0.1)	88.9	84.5	76.1	96.8	97.3	97.4	0.408	0.401	0.384
(0.1,0.1,0.4)	84.9	88.6	80.3	97.4	96.8	97.1	0.408	0.413	0.397
(0.4,0.3,0.3)	94.2	91.7	87.5	97.8	98.0	97.1	0.426	0.422	0.414
(0.3,0.3,0.4)	92.0	93.8	88.1	98.0	97.5	96.5	0.426	0.428	0.419
(0.4,0.3,0.2)	92.1	88.1	82.3	97.5	97.9	96.7	0.421	0.416	0.406

$\pi_0 = 0.1;\ n_1 = n_2 = 12;\ \alpha/2 = 0.025$

Tie violation: in case of a tie in Stage 1 responses, choose the second-most preferred treatment

IMPACT OF SELECTION RULE VIOLATIONS ON UMVUE

(π_1, π_2, π_3)	Bias of UMVUE			Empirical SD of UMVUE		
	No violation	Tie violation	10% violation	No violation	Tie violation	10% violation
(0.1,0.1,0.1)	0.002	-0.010	-0.042	0.071	0.076	0.076
(0.3,0.3,0.3)	0.001	-0.012	-0.054	0.112	0.123	0.140
(0.4,0.1,0.1)	0.000	0.003	-0.003	0.117	0.142	0.181
(0.1,0.1,0.4)	0.001	-0.007	-0.008	0.136	0.123	0.164
(0.4,0.3,0.3)	0.001	-0.008	-0.050	0.124	0.143	0.164
(0.3,0.3,0.4)	0.001	-0.015	-0.051	0.126	0.130	0.154
(0.4,0.3,0.2)	0.000	-0.005	-0.042	0.125	0.148	0.173

$\pi_0 = 0.1;\ \ n_1 = n_2 = 12;\ \ \alpha/2 = 0.025$

Tie violation: in case of a tie in Stage 1 responses, choose the second-most preferred treatment