Optimal multi-arm multi-stage (*MAMS*) platform randomized trials using Stata: Rationale, design, and implementation

Babak Choodari-Oskooei, PhD
Principal Research Fellow
b.choodari-Oskooei@ucl.ac.uk

MRC Clinical Trials Unit
Institute of Clinical Trials and Methodology
UCL
London, UK

22 February 2024 | Stata Biostatistics and Epidemiology Virtual Symposium

Smarter Studies
Global Impact
Better Health

Outline

- Introduction to *MAMS* designs: Rationale and advantages
- Examples: Cancer and surgery
- Sample sizes calculations using `nstage` Stata commands
- How can we make *MAMS* designs more efficient?
- Summary

Multi-Arm Multi-Stage (*MAMS*) designs

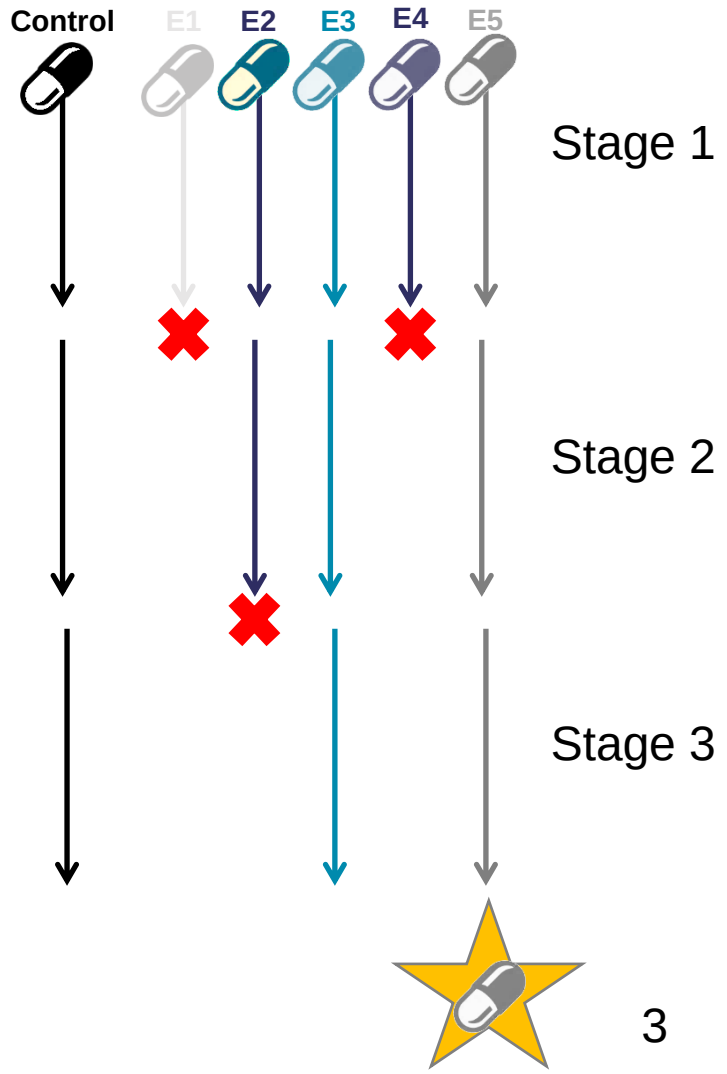
- Methods by **Royston & Parmar et al**. (Statistics in Medicine, 2003)
  - For time-to-event outcomes

- Context: Randomised clinical trials
  - Efficacy and safety of new interventions in a defined population

- Control of operating characteristics are important
  - Probability of false positive (Type I error)
    - Of interest to regulators and reviewers
  - Probability of true positive (Power)
    - Of interest to funders

- Multiple research arms vs a common control arm (or standard-of-care)

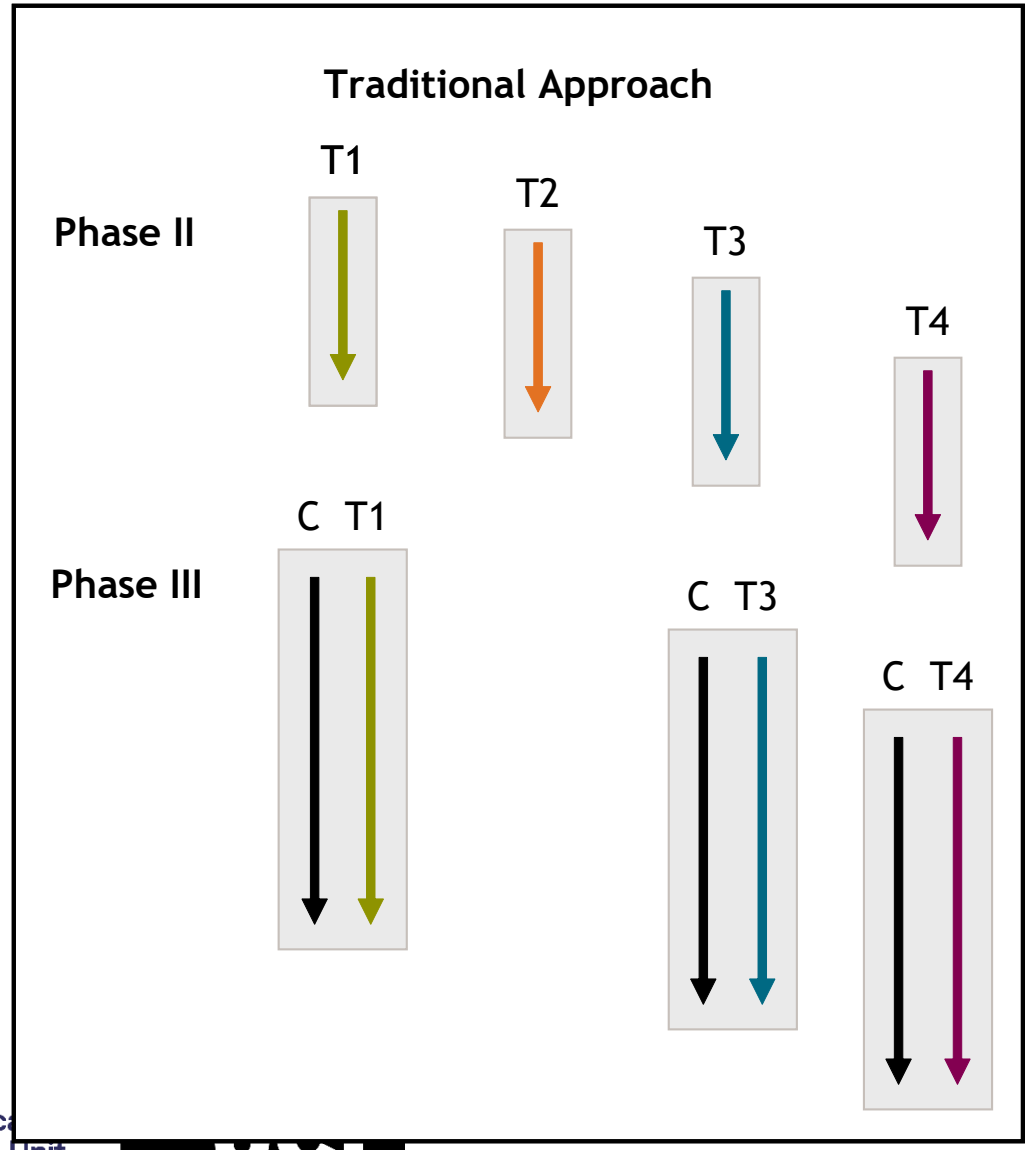
- *MAMS* design has several advantages:
  - One of which is the use of an *intermediate* (I) outcome

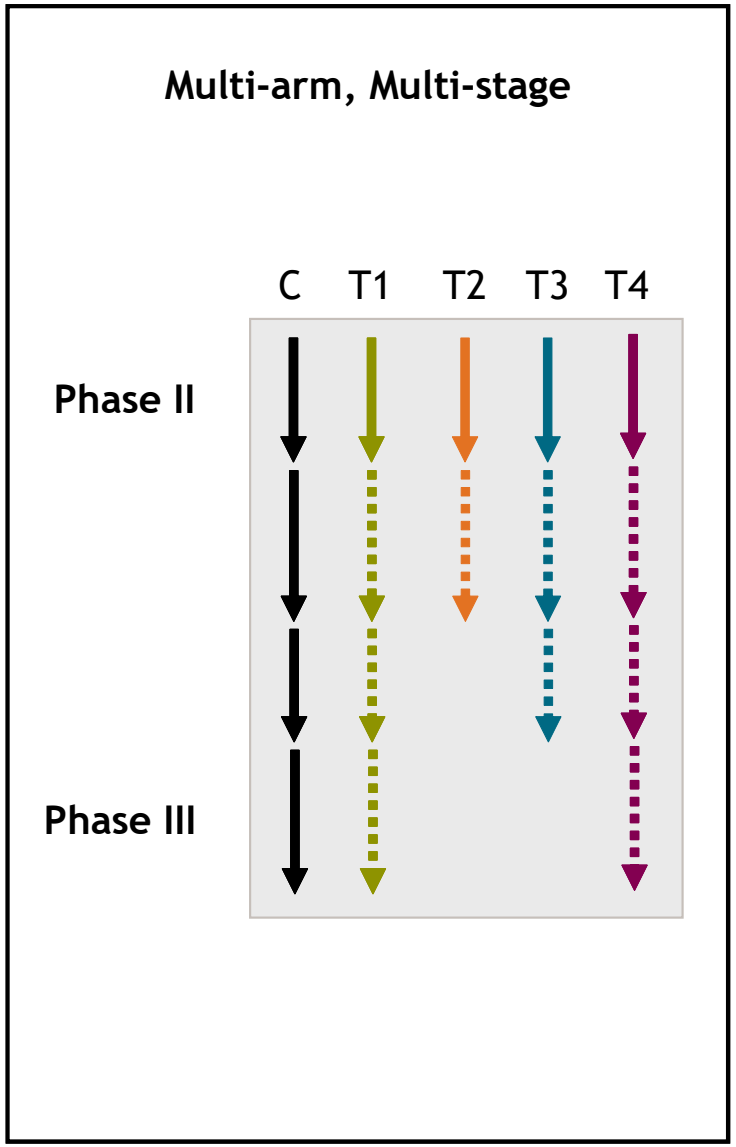


MRC Clinical Trials Unit

UCL

Traditional approach to testing





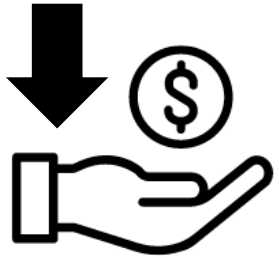
MRC Clinical Trials Unit

MAMS platform trial

- Have a single **master** protocol,
- Address **multiple** research questions **over time**,
- Can **add** new research arms as well as dropping the existing one(s)

Advantages of *MAMS* platform trials

Faster
Multiple treatments tested at the same time



Cost
No need to set up a new trial for each treatment

Facilitate recruitment
Fewer patients required overall

Flexibility
Drop and add treatments

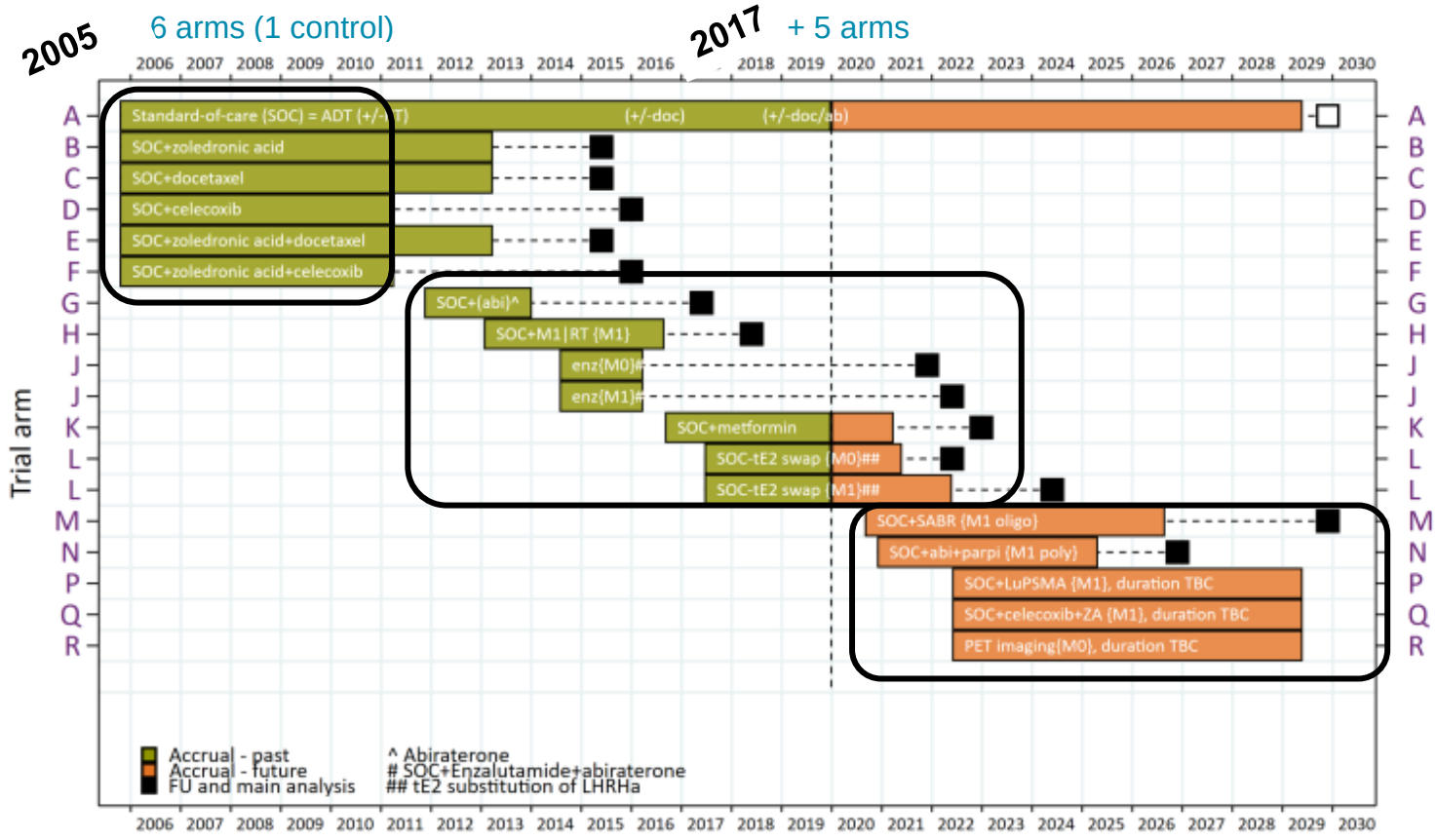
Outline

- Introduction to MAMS designs: Rationale and advantages
- **Examples: Cancer and surgery**
- Sample sizes calculations using `nstage` Stata commands
- How can we make MAMS designs more efficient?
- Summary & discussion

Example 1:

Flagship STAMPEDE trial: advanced prostate cancer

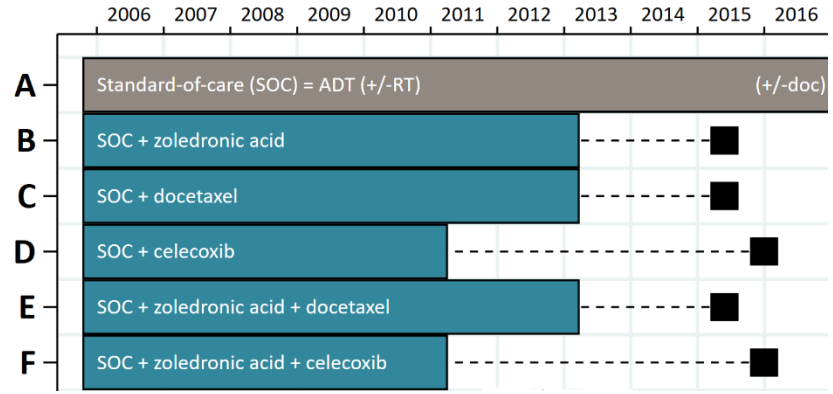
STAMPEDE trial: Advanced prostate cancer



MRC Clinical Trials Unit

UCL

STAMPEDE trial



Original design:

- STAMPEDE started with 6 arms.
  - i.e., 5 pairwise comparison.

- Each comparison posed a distinct research question.

- It tested a distinct hypothesis in each pairwise comparison.

- Therefore, probability of false positive (type I error rate) was controlled for each pairwise comparison – PWER = 0.025, one-sided.

MRC Clinical Trials Unit

UCL

Original STAMPEDE trial: Comparisons

Design comparisons:

No	Design	Type I error rate Control (value)	Final stage Sig. level (one-sided)	Pairwise Power	Total events*
1	5 two-arm trials	Pairwise (0.025)	0.025	0.83	2580
2	6-arm MAMS trial	Pairwise (0.025)	0.025	0.83	1336

- Reduction in effective sample size: 48%!

*) Total events: required number of events for primary analysis across all arms.

MRC Clinical Trials Unit

UCL

Original STAMPEDE trial: Comparisons

Design comparisons:

No	Design	Type I error rate Control (value)	Final stage Sig. level (one-sided)	Pairwise Power	Total events*
1	5 two-arm trials	Pairwise (0.025)***	0.025	0.83	2580
2	6-arm MAMS trial	Pairwise (0.025)	0.025	0.83	1336
3	6-arm MAMS trial	*Familywise* (0.025)	**0.0055****	0.83	1857

*) Total events: required number of events for primary analysis across all arms.
**) Multiplicity-adjusted final stage significance level, Dunnett's correction.
***) The *Familywise* error rate of 5 two-arm trials is about 12%!

STAMPEDE: Design specification

Design parameters: one arm vs control

Stage/ analysis	Outcome	Hazard Ratio	Design Power	One-sided α	
1: LOB	FFS	0.75	95%	0.50	
2: LOB	FFS	0.75	95%	0.25	
3: LOB	FFS	0.75	95%	0.10	
4: Efficacy	OS	0.75	90%	0.025	

Primary Outcome

LOB: lack-of-benefit analysis

MRC Clinical Trials Unit

UCL

STAMPEDE: Design specification

Design parameters: one arm vs control

Stage/ analysis	Outcome	Hazard Ratio	Design Power	One-sided α	Control arm events
1: LOB	FFS	0.75	95%	0.50	113
2: LOB	FFS	0.75	95%	0.25	223
3: LOB	FFS	0.75	95%	0.10	350
4: Efficacy	OS	0.75	90%	0.025	436

Primary Outcome

LOB: lack-of-benefit analysis

MRC Clinical Trials Unit

UCL

STAMPEDE: Design specification

Interim Stages

- Interim outcome*
- High power to avoid dropping an effective treatment
- Significance becomes stricter over time

Design parameters: one arm vs control

Stage/ analysis	Outcome	Hazard Ratio	Design Power	One-sided α	Control arm events
1: LOB	FFS	0.75	95%	0.50	113
2: LOB	FFS	0.75	95%	0.25	223
3: LOB	FFS	0.75	95%	0.10	350
4: Efficacy	OS	0.75	90%	0.025	436

LOB: lack-of-benefit analysis

*PSA-failure, local progression, nodal progression, progression of metastases or new metastases or death from prostate ca.

Chosen on assumption that any trt which shows an advantage in OS will probably show an advantage in FFS first and unlikely to be an OS advantage if no FFS advantage.

STAMPEDE: Design specification

Design parameters: one arm vs control

Stage/ analysis	Outcome	Hazard Ratio	Design Power	One-sided α	Control events
1: LOB	FFS	0.75	95%	0.50	113
2: LOB	FFS	0.75	95%	0.25	223
3: LOB	FFS	0.75	95%	0.10	350
4: Efficacy	OS	0.75	90%	0.025	436

At stage II, if P-value > 0.25, IDMC likely to recommend stopping treatment arm for lack-of-benefit

MRC Clinical Trials Unit

UCL

Appropriate intermediate (I) outcome

- Increases efficiency:
    - speeds up the weeding out of the insufficiently promising treatments.

Key assumptions:

1. “Information” on I-outcome accrues at the same rate or faster rate than that of D-outcome
2. The I-outcome is on the pathway between the treatments and D-outcome.
3. If the null hypothesis is true for the I-outcome, it must also hold for D-outcome.

- I-outcome *does not* have to be a perfect surrogate for D-outcome in *Prentice* sense.

Key example in cancer: I = event for progression-free survival; D = death

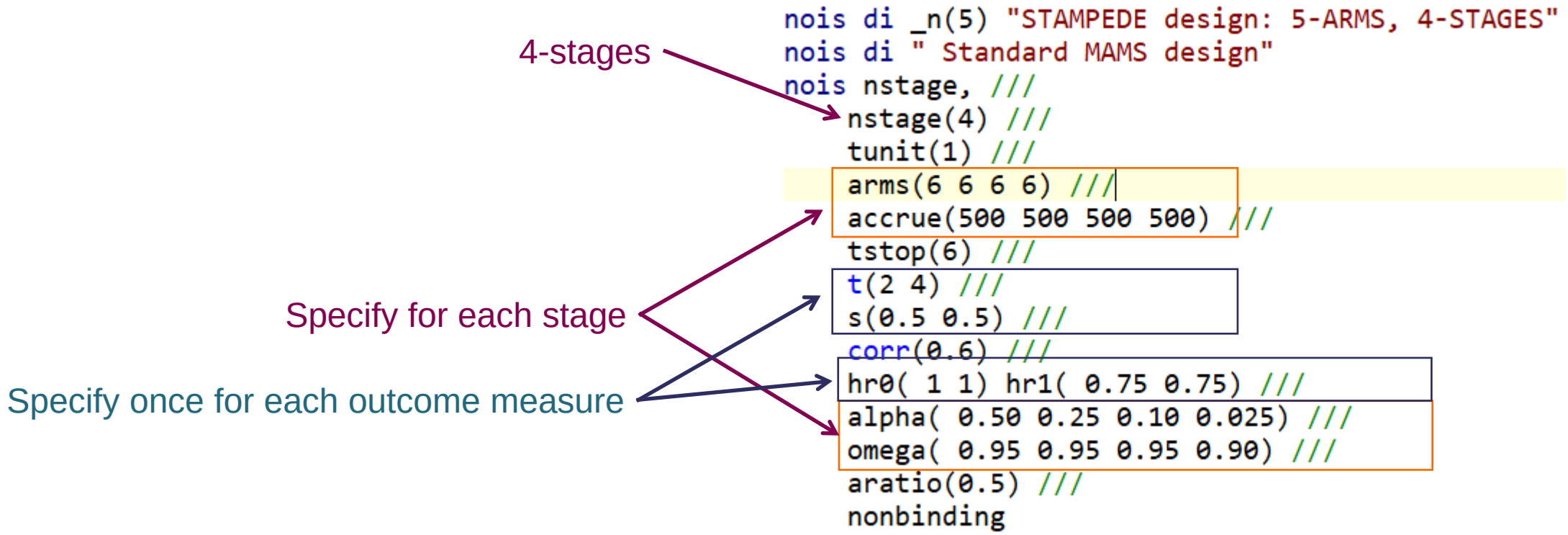
Outline

- Introduction to MAMS designs: Rationale and advantages
- Examples: Cancer and surgery
- **Sample sizes calculations using `nstage` Stata commands**
- How can we make MAMS designs more efficient?
  - Optimal MAMS designs
- Summary & discussion

MRC Clinical Trials Unit

UCL

`nstage` Stata command: STAMPEDE design



`nstage` Stata command: STAMPEDE design

Time units = 1 (years)

6 arms: 5 research and 1 control arms

500 pts/yr

Stop accrual at 6 yr

Target HR for interim & final: HR=0.75

Allocation (C:E) 2:1
(Pts on research arm for each pts on control arm)

```
nois di _n(5) "STAMPEDE design: 5-ARMS, 4-STAGES"
nois di " Standard MAMS design"
nois nstage, ///
  nstage(4) ///
  tunit(1) ///
  arms(6 6 6 6) ///
  accrue(500 500 500 500) ///
  tstop(6) ///
  t(2 4) ///
  s(0.5 0.5) ///
  corr(0.6) ///
  hr0( 1 1) hr1( 0.75 0.75) ///
  alpha( 0.50 0.25 0.10 0.025) ///
  omega( 0.95 0.95 0.95 0.90) ///
  aratio(0.5) ///
  nonbinding
```

`nstage` Output 1: STAMPEDE

```
Median survival time (I-outcome): 2 time units
Median survival time (D-outcome): 4 time units


Operating characteristics
----------------------------------------------------------------------------
Stage  Alpha(LOB)*    Power    HR|H0    HR|H1  Crit.HR Length**   Time**
----------------------------------------------------------------------------
1          0.5000    0.950    1.000    0.750    1.000    2.436    2.436
2          0.2500    0.950    1.000    0.750    0.920    1.189    3.625
3          0.1000    0.950    1.000    0.750    0.882    1.161    4.786
4          0.0250    0.901    1.000    0.750    0.841    2.572    7.359
----------------------------------------------------------------------------
Pairwise Error Rate                 0.0250          Pairwise Power   0.8995
Max. Familywise Error Rate (SE)     0.1033 (0.0006)
----------------------------------------------------------------------------
LOB = lack of benefit
Note: patient accrual stopped at time  6.000
 *   All alphas are one-sided
 **  Length (duration of each stage) is expressed in  periods and
assumes survival times are exponentially distributed.
    Time is expressed in cumulative periods.
```

```
nois di _n(5) "STAMPEDE design: 5-ARMS, 4-STAGES"
nois di " Standard MAMS design"
nois nstage, ///
    nstage(4) ///
    tunit(1) ///
    arms(6 6 6 6) ///
    accrue(500 500 500 500) ///
    tstop(6) ///
    t(2 4) ///
    s(0.5 0.5) ///
    corr(0.6) ///
    hr0( 1 1) hr1( 0.70 0.75) ///
    alpha( 0.450 0.200 0.050 0.025) ///
    omega( 0.95 0.95 0.95 0.90) ///
    aratio(0.5) ///
    nonbinding
```

`nstage` Output 2: STAMPEDE

Sample size and number of events

	Overall	Control	Exper.
---------Stage 1---------			
Arms	6	1	5
Acc. rate	500	143	357
Patients*	1218	348	870
Events**	343	113	230
---------Stage 2---------			
	Overall	Control	Exper.
Arms	6	1	5
Acc. rate	500	143	357
Patients*	1813	518	1295
Events**	683	223	460
---------Stage 3---------			
	Overall	Control	Exper.
Arms	6	1	5
Acc. rate	500	143	357
Patients*	2393	684	1709
Events**	1085	350	735
---------Stage 4---------			
	Overall	Control	Exper.
Arms	6	1	5
Acc. rate	500	143	357
Patients*	3000	857	2143
Events**	1336	436	900

```
nois di _n(5) "STAMPEDE design: 5-ARMS, 4-STAGES"
nois di " Standard MAMS design"
nois nstage, ///
    nstage(4) ///
    tunit(1) ///
    arms(6 6 6 6) ///
    accrue(500 500 500 500) ///
    tstop(6) ///
    t(2 4) ///
    s(0.5 0.5) ///
    corr(0.6) ///
    hr0( 1 1) hr1( 0.70 0.75) ///
    alpha( 0.450 0.200 0.050 0.025) ///
    omega( 0.95 0.95 0.95 0.90) ///
    aratio(0.5) ///
    nonbinding
```

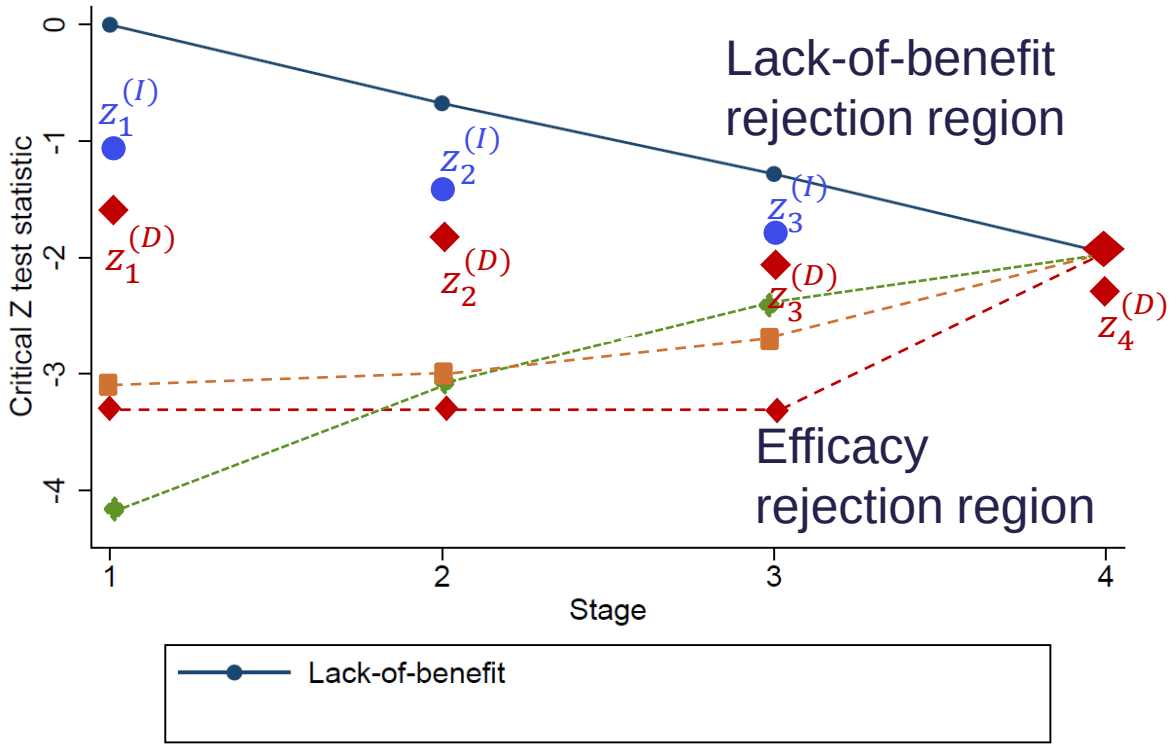
MRC Clinical Trials Unit

UCL

Interim efficacy stopping boundaries, I≠D

- **Haybittle-Peto (HP)**
  - Constant
- **O-Brien-Fleming type (OBF)**
  - Extreme early
- **Spending function**
  - Based on information time

- **When *I* & *D* are different outcomes:**
  - The LOB boundaries are on the ***I***-outcome



`nstage`: STAMPEDE with LOB and ESB

```
nois di _n(5) "STAMPEDE design: 5-ARMS, 4-STAGES"
nois di "MAMS with both LOB and ESB"
nois nstage, ///
    nstage(4) ///
    tunit(1) ///
    arms(6 6 6 6) ///
    accrue(500 500 500 500) ///
    tstop(6) ///
    t(2 4) ///
    s(0.5 0.5) ///
    corr(0.6) ///
    hr0( 1 1) hr1( 0.75 0.75) ///
    alpha( 0.50 0.25 0.10 0.025) ///
    omega( 0.95 0.95 0.95 0.90) ///
    aratio(0.5) ///
    nonbinding ///
    esb(hp=0.0005) ///
    fwer(0.025)
```

Interim stopping boundaries for efficacy

Familywise type I error rate – Dunnett's correction

MRC Clinical Trials Unit

UCL

`nstage` output 1: STAMPEDE with LOB and ESB

```
nois di _n(5) "STAMPEDE design: 5-ARMS, 4-STAGES"
nois di "MAMS with both LOB and ESB"
nois nstage, ///
    nstage(4) ///
    tunit(1) ///
    arms(6 6 6 6) ///
    accrue(500 500 500 500) ///
    tstop(6) ///
    t(2 4) ///
    s(0.5 0.5) ///
    corr(0.6) ///
    hr0( 1 1) hr1( 0.75 0.75) ///
    alpha( 0.50 0.25 0.10 0.025) ///
    omega( 0.95 0.95 0.95 0.90) ///
    aratio(0.5) ///
    nonbinding ///
    esb(hp=0.0005) ///
    fwer(0.025)
```

```
Median survival time (I-outcome): 2 time units

Median survival time (D-outcome): 4 time units


Operating characteristics
------------------------------------------------------------------------------
Stage    Alpha   Alpha   Power   HR|H0   HR|H1    Crit.HR   Crit.HR    Length**   Time**
        (LOB)*  (ESB)*                          (LOB)     (ESB)
------------------------------------------------------------------------------
1       0.5000  0.0005  0.950   1.000   0.750    1.000     0.439      2.436      2.436
2       0.2500  0.0005  0.950   1.000   0.750    0.920     0.512      1.189      3.625
3       0.1000  0.0005  0.950   1.000   0.750    0.882     0.553      1.161      4.786
4       0.0043       .  0.901   1.000   0.750    0.824         .      5.790     10.576
------------------------------------------------------------------------------

Max. Pairwise Error Rate                0.0054              Pairwise Power    0.9001
Max. Familywise Error Rate (SE)         0.0251 (0.0002)
------------------------------------------------------------------------------
```

MRC Clinical Trials Unit

UCL

`nstage` output 1: STAMPEDE with LOB and ESB

```
nois di _n(5) "STAMPEDE design: 5-ARMS, 4-STAGES"
nois di "MAMS with both LOB and ESB"
nois nstage, ///
    nstage(4) ///
    tunit(1) ///
    arms(6 6 6 6) ///
    accrue(500 500 500 500) ///
    tstop(6) ///
    t(2 4) ///
    s(0.5 0.5) ///
    corr(0.6) ///
    hr0( 1 1) hr1( 0.75 0.75) ///
    alpha( 0.50 0.25 0.10 0.025) ///
    omega( 0.95 0.95 0.95 0.90) ///
    aratio(0.5) ///
    nonbinding ///
    esb(hp=0.0005) ///
    fwer(0.025)
```

Median survival time (I-outcome): 2 time units

Median survival time (D-outcome): 4 time units

Operating characteristics

Stage	Alpha (LOB)*	Alpha (ESB)*	Power	HR|H0	HR|H1	Crit. HR (LOB)	Crit. HR (ESB)	Length*	Time**
1	0.5000	0.0005	0.950	1.000	0.750	1.000	0.439	2.436	2.436
2	0.2500	0.0005	0.950	1.000	0.750	0.920	0.512	1.189	3.625
3	0.1000	0.0005	0.950	1.000	0.750	0.882	0.553	1.161	4.786
4	0.0043	.	0.901	1.000	0.750	0.824	.	5.790	10.576

Max. Pairwise Error Rate 0.0054 Pairwise Power 0.9001

Max. Familywise Error Rate (SE) 0.0251 (0.0002)

Multiplicity-adjusted significance level for the primary analysis to strongly control FWER at 2.5% level

MRC Clinical Trials Unit

`nstage` output 2: STAMPEDE with LOB and ESB

```
Sample size and number of events
          ---------Stage 1---------
          Overall  Control   Exper.
Arms            6        1        5
Acc. rate     500      143      357
Patients*    1218      348      870
Events**      343      113      230
          ---------Stage 2---------
          Overall  Control   Exper.
Arms            6        1        5
Acc. rate     500      143      357
Patients*    1813      518     1295
Events**      683      223      460
          ---------Stage 3---------
          Overall  Control   Exper.
Arms            6        1        5
Acc. rate     500      143      357
Patients*    2393      684     1709
Events**     1085      350      735
          ---------Stage 4---------
          Overall  Control   Exper.
Arms            6        1        5
Acc. rate     500      143      357
Patients*    3000      857     2143
Events**     1941      616     1325
          ---------------------------
```

```
nois di _n(5) "STAMPEDE design: 5-ARMS, 4-STAGES"
nois di "MAMS with both LOB and ESB"
nois nstage, ///
    nstage(4) ///
    tunit(1) ///
    arms(6 6 6 6) ///
    accrue(500 500 500 500) ///
    tstop(6) ///
    t(2 4) ///
    s(0.5 0.5) ///
    corr(0.6) ///
    hr0( 1 1) hr1( 0.75 0.75) ///
    alpha( 0.50 0.25 0.10 0.025) ///
    omega( 0.95 0.95 0.95 0.90) ///
    aratio(0.5) ///
    nonbinding ///
    esb(hp=0.0005) ///
    fwer(0.025)
```

For more details and steps to design MAMS trials in the book chapter:

- *Choodari-Oskooei et al*, "Multi-arm multi-stage (MAMS) platform randomized clinical trials", in "Principles and Practice of Clinical Trials" by Springer (2022)
- Link: [bit.ly/3tmx0qT](bit.ly/3tmx0qT)

MRC Clinical Trials Unit

UCL

Outline

- Introduction to MAMS designs: Rationale and advantages
- Examples: Cancer and surgery
- Sample sizes calculations using `nstage` Stata commands
- **How can we make *MAMS* designs more efficient?**
  - **Optimal *MAMS* designs**
- Summary & discussion

MRC Clinical Trials Unit

UCL

Efficient (optimal) *MAMS* designs

- So far, chose stagewise design parameters then calculated overall operating characteristics.
- In the confirmatory setting the overall operating characteristics, type I and II error rates should be controlled as pre-specified level.
- In MAMS, several design options exist for a given overall type I and II error rates
- How can we choose the best design among these potential choices?
- This lends itself to the topic of efficient and optimal MAMS design
  - Here we are only exploring designs that are “mathematically” optimal.
  - “Clinically” optimality in also important, in terms of the practicality of the design and possible fragilities – personal communication with Patrick Royston.

Example 2:

ROSSINI-2 *MAMS* trial in surgical wound infection

Example 2: ROSSINI 2 surgical trial

8-arm 3-stage trial in surgery

- i.e. 7 pairwise comparisons

The overall type I error rate across all comparisons and stages (FWER) is controlled at 2.5% (one-sided)

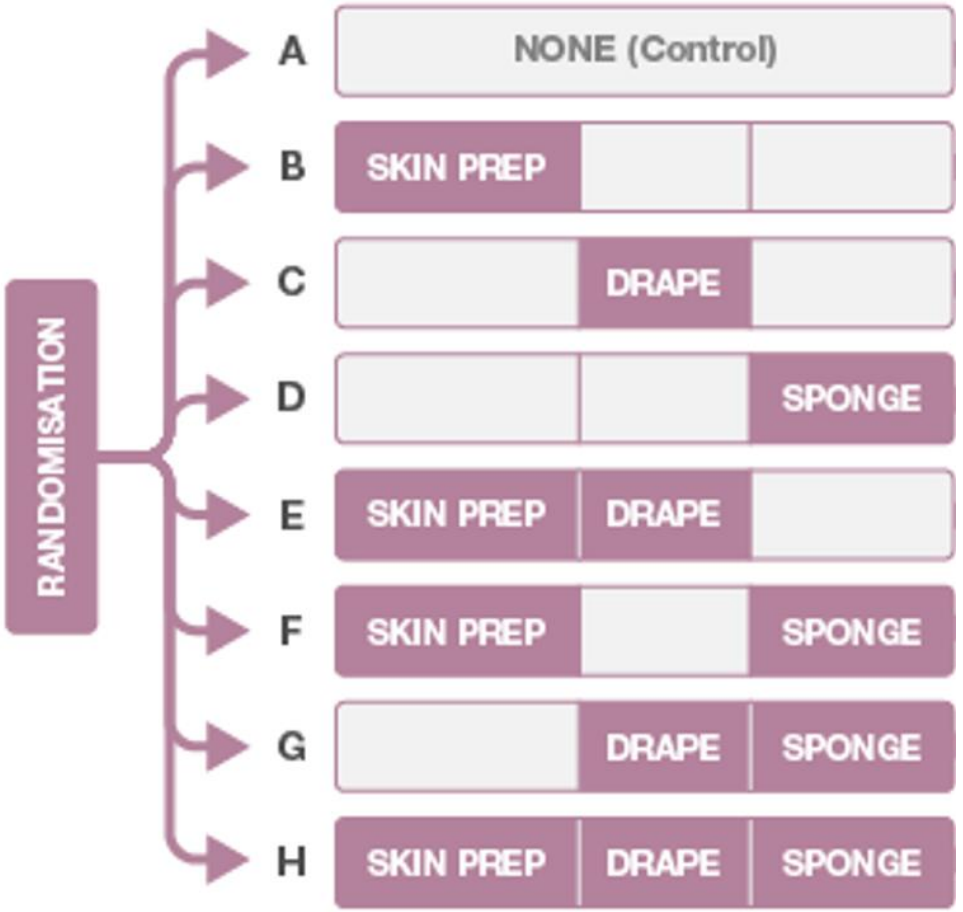
Overall pairwise power is controlled at 85%, i.e. for each

pairwise comparison of research arm against control

Clinical outcome composite binary,

- Surgical site infection (SSI) within 4 weeks
- Same outcome used in all stages

Target an effect size of 5% reduction from control arm risk of 15%



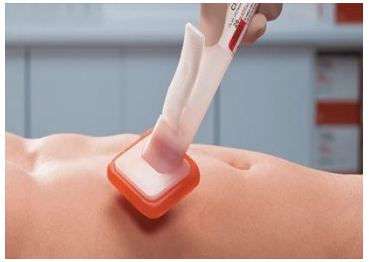
Example: ROSSINI-2 surgical trial

8-arm 3-stage surgical trial

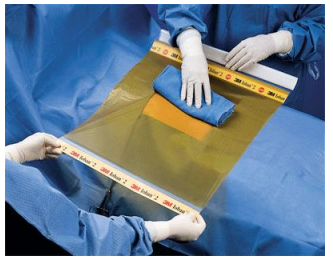
Clinical outcome: composite binary, surgical site infection (SSI);

Main interventions

A) Chlorhexidine 2% alcoholic <u>skin prep</u>
[versus any other standard wound prep agent of surgeon's choice]



B) Loban-impregnated <u>incise drapes</u>
[versus no drape]



C] Gentamicin-impregnated collagen <u>sponge</u>
[versus no sponge]

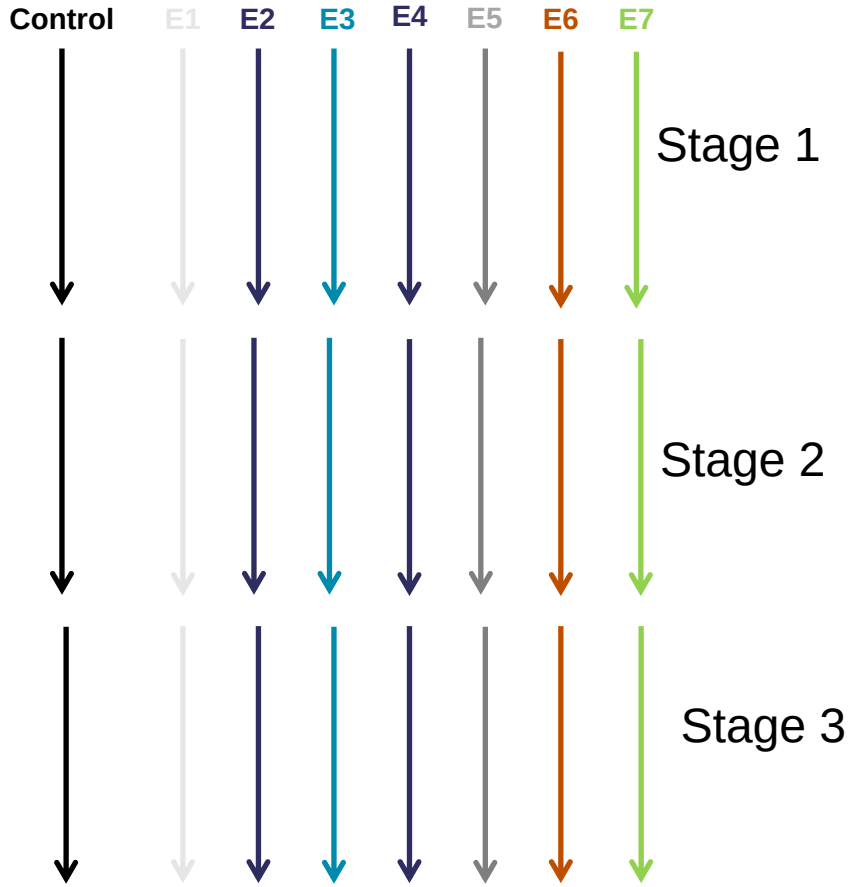
How can we embark on designing such a trial?

Step 1: Choose design significance levels and pairwise power at each stage such that overall type I and II error rates are controlled at prespecified levels.

- `nstagebinopt` Stata command

Step 2: Conduct sample size calculations and estimate the trial timelines, using the design parameters found in Step 1.

- `nstagebin` Stata command



Step 1 leads us to optimal *MAMS* designs

Biggest challenge is to find a “universal” optimality criterion.

Efficient designs minimise the expected sample size (ESS) under a certain scenario

Common choices are:

- minimax: designs with smallest ESS under the alternative hypothesis for all comparisons
- null-optimal: designs with smallest ESS under the null hypothesis for all comparisons

Admissible designs minimise a weighted sum of these two measures:

$$L = q.E(N|H_1) + (1-q).E(N|\ H_0) \qquad q \in [0,1], \text{ which is prespecified}$$

Loss function used in `nstagebinopt`

Back to ROSSINI-2 surgical trial

`nstagebinopt` command for admissible designs

Syntax:

```
nstagebinopt, alpha(0.025) power(0.85) fwer nstage(3) arms(8) theta0(0)
theta1(-0.05) ctrlp(0.15) ltfu(0.04) fu(4) accrate(118 248 248)
aratio(0.5) plot
```

Aim to control of the FWER
Remove if aim is to control
pairwise error rate (PWER)

`nstagebinopt` output

Admissible for wider range of q

n-stage (binary) trial design version 1.0.2, 09 June 2023

Admissible designs for a 8-arm 3-stage trial with binary outcome based on Choodari-Oskooei, Bratton, and Parmar (2023) Stata Journal 23(3).

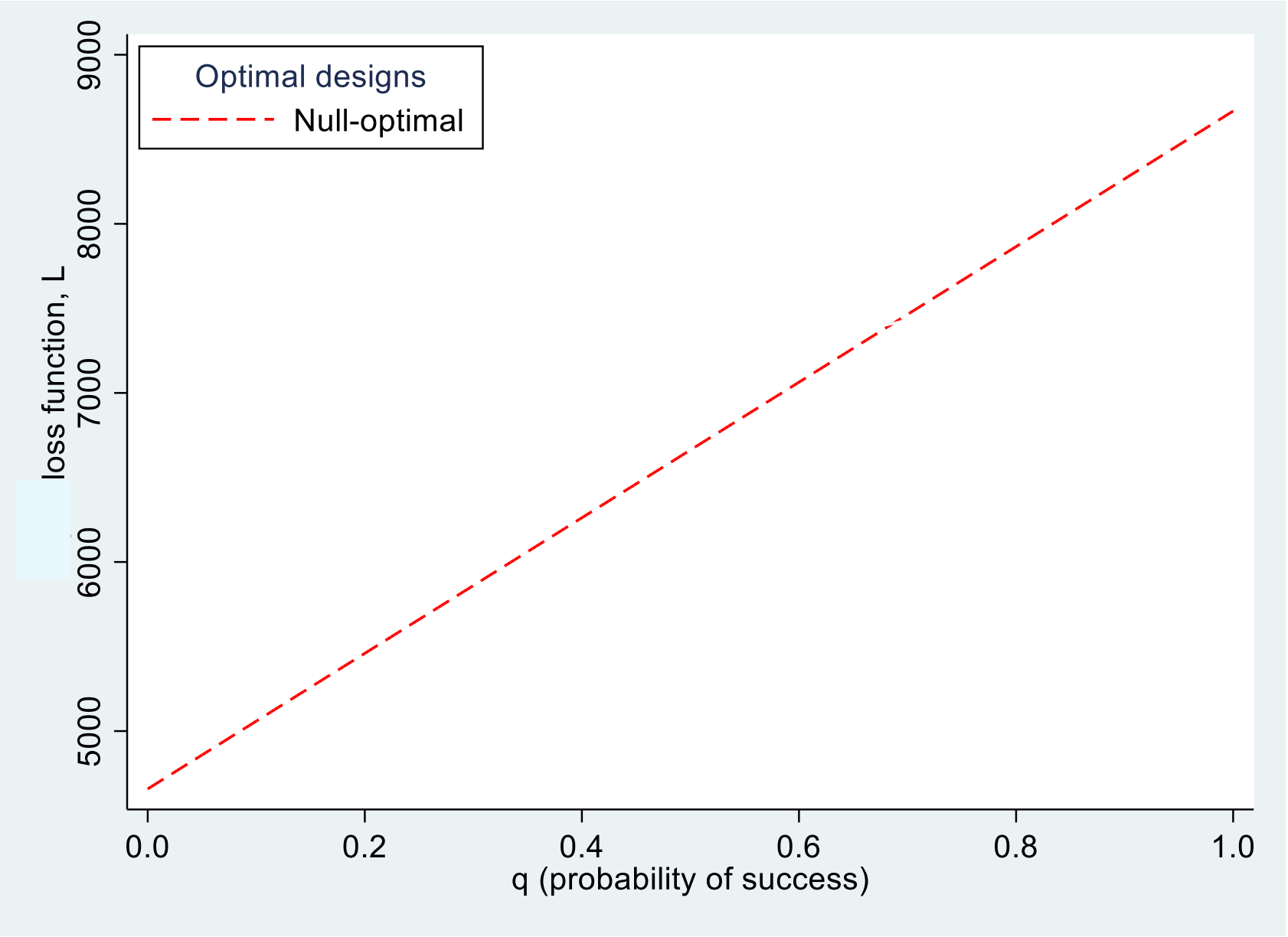
Design number	q-range	Stage	Sig. level	Power	Alloc. ratio	E(N|H0)	E(N|H1)	FWER (SE)
1	[0.00,0.09]	1	0.31	0.93	0.50	4658	8667	0.0249
		2	0.16	0.93				(0.0003)
		3	0.005	0.92				
2	[0.10,0.65]	1	0.40	0.94	0.50	4683	8437	0.0253
		2	0.14	0.94				(0.0003)
		3	0.005	0.91				
3	[0.66,0.77]	1	0.15	0.93	0.50	4989	8277	0.0254
		2	0.08	0.93				(0.0003)
		3	0.005	0.90				
4	[0.78,1.00]	1	0.27	0.99	0.50	6506	7824	0.0253
		2	0.14	0.99				(0.0003)
		3	0.004	0.85				

Note: each design minimises the loss function (1-q)E(N|H0)+qE(N|H1) for values of q specified in q_range. H1 is the hypothesis that all of the experimental arms are effective.

MRC Clinical Trials Unit

UCL

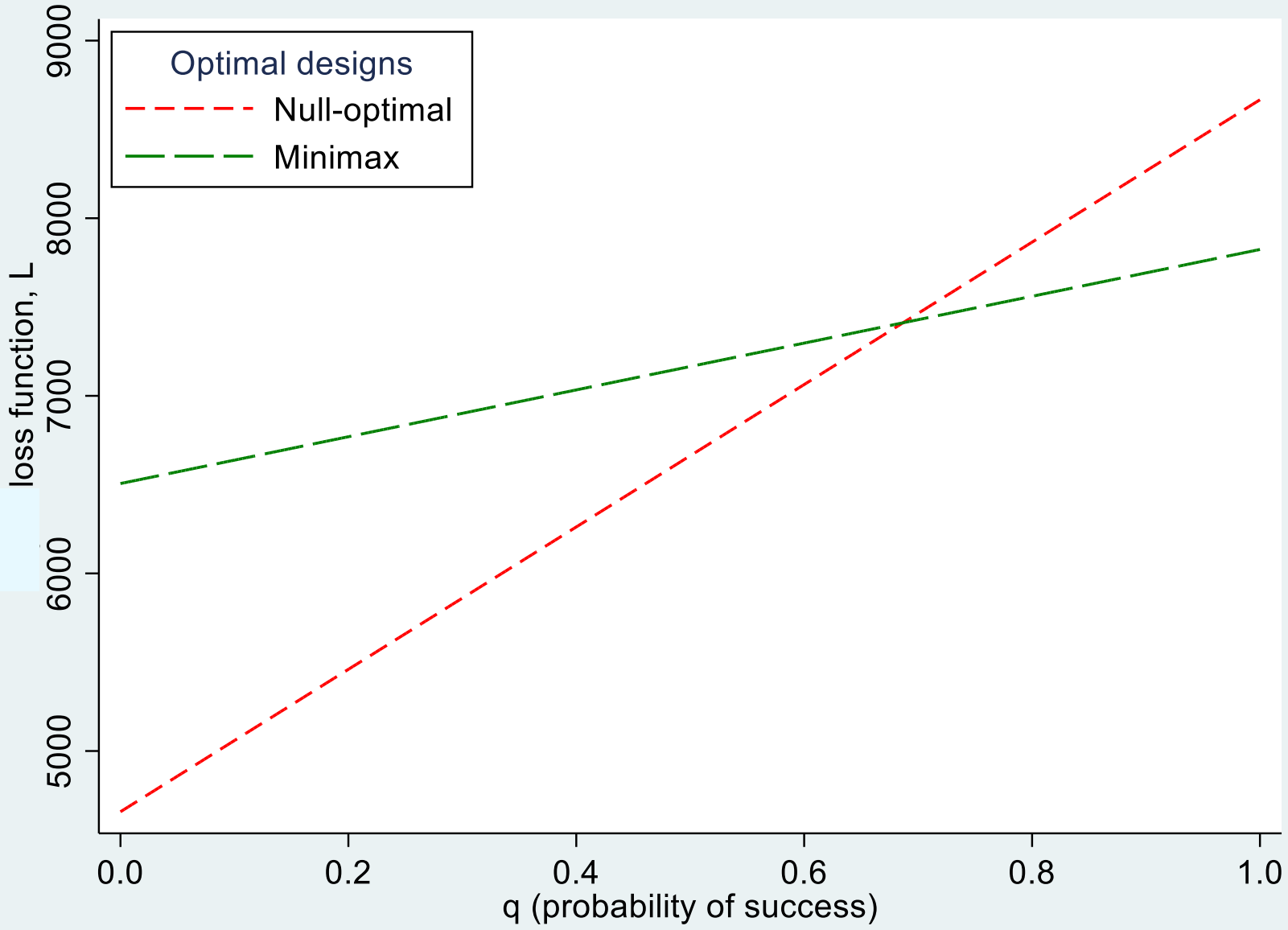
Loss function vs probability of success



$$L = q.E(N|H_1) + (1-q).E(N|\ H_0)$$

$$L = [E(N|H_1) - E(N|\ H_0)]\ .q + E(N|\ H_0)$$

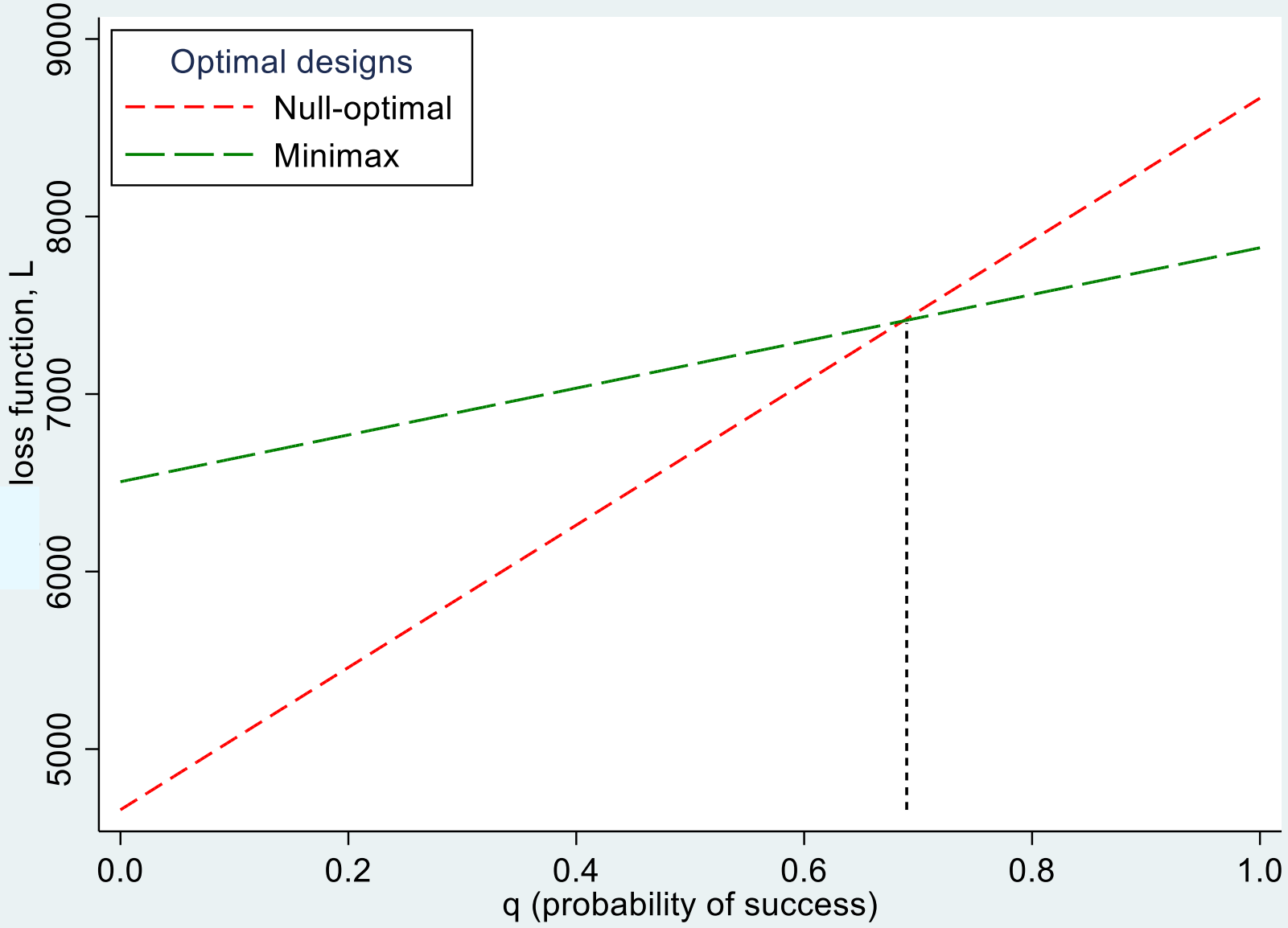
Loss function vs probability of success



$$L = q.E(N|H_1) + (1-q).E(N|\ H_0)$$

$$L = [E(N|H_1) - E(N|\ H_0)]\ .q + E(N|\ H_0)$$

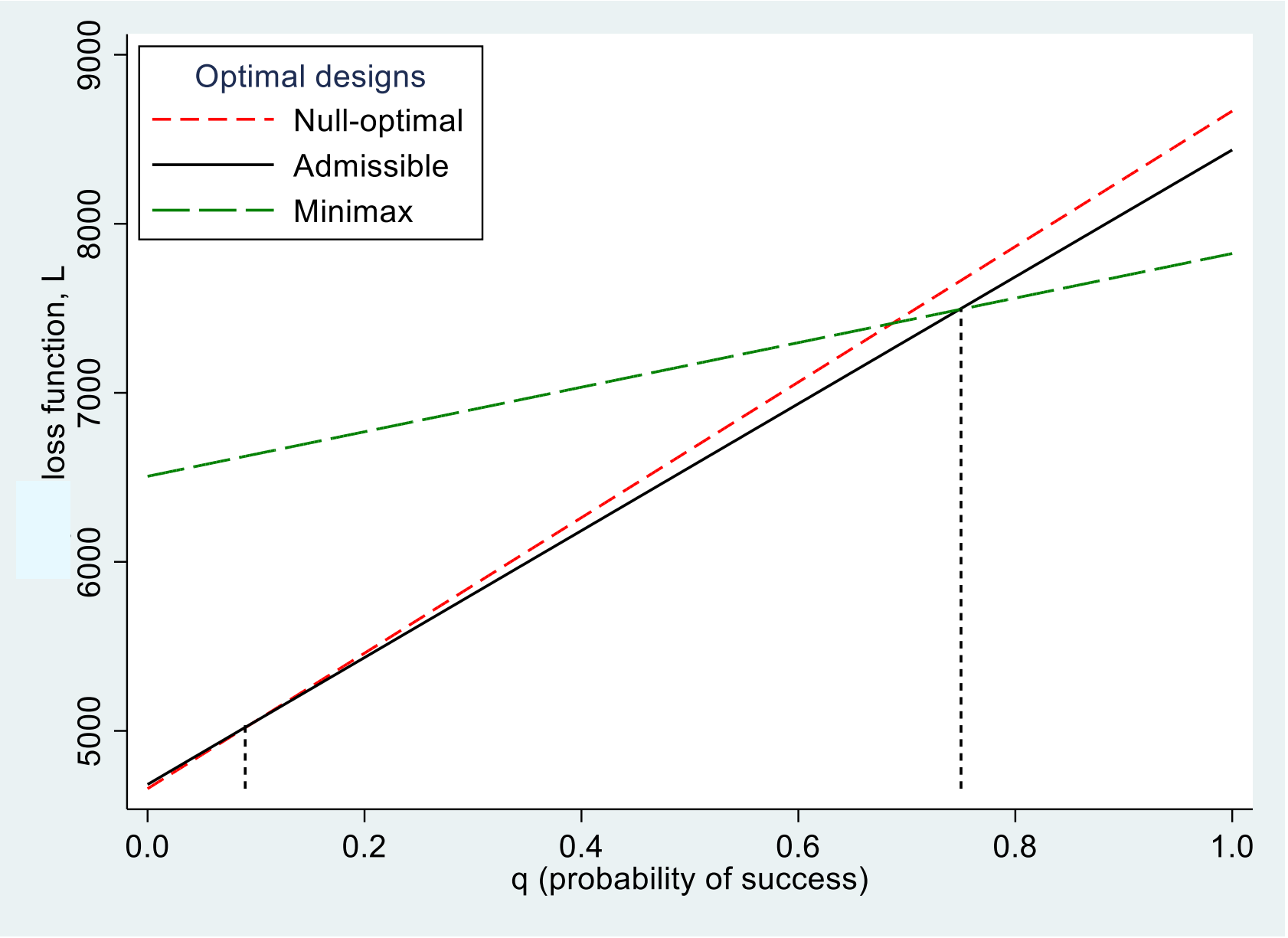
Loss function vs probability of success



$$L = q.E(N|H_1) + (1-q).E(N|\ H_0)$$

$$L = [E(N|H_1) - E(N|\ H_0)]\ .q + E(N|\ H_0)$$

Loss function vs probability of success



$$L = q.E(N|H_1) + (1-q).E(N|\ H_0)$$

$$L = [E(N|H_1) - E(N|\ H_0)]\ .q + E(N|\ H_0)$$

Step 2: `nstagebin` for sample size calculations

Syntax:

```
nstagebin, alpha(0.40 0.14 0.005) power(0.94 0.94 0.91) nstage(3) theta0(0)
theta1(-0.05) ctrlp(0.15) arms(8 6 4) ltfu(0.04) fu(4) accrate(118 248 248)
aratio(0.5) tunit(4)
```

MRC Clinical Trials Unit

UCL

Step 2: `nstagebin` output 1

Output has 2 main sections

Section 1

```
n-stage trial design - binary outcome       version 1.0.2, 09 June 2023
------------------------------------------------------------------------
Sample size for a 8-arm 3-stage trial with binary outcome based on
Bratton et al. (2013) BMC Med Res Meth 13:139 and Choodari-Oskooei,
Bratton, and Parmar (2023) Stata Journal 23(3).
------------------------------------------------------------------------

Control arm event rate = 0.15
Delay in observing outcome = 4 months
Attrition rate for outcome = 0.04

Operating characteristics
------------------------------------------------------------------------------
             Alpha(1S)      Power   theta|H0   theta|H1    Length*      Time*
------------------------------------------------------------------------------
Stage 1         0.4000      0.940      0.000     -0.050     19.979     19.979
Stage 2         0.1400      0.940      0.000     -0.050      9.165     29.144
Stage 3         0.0050      0.910      0.000     -0.050     11.994     41.138
Pairwise        0.0040      0.850                                    41.138
FWER(SE)**      0.0253   (0.0003)
------------------------------------------------------------------------------
 *  Length (duration of each stage) is expressed in month periods
 **  FWER is calculated using simulations with 250000 replications
```

MRC Clinical Trials Unit

UCL

Step 2: `nstagebin` output 2

Section 2

```
---------Stage 1---------
                        Overall  Control   Exper.
-------------------------------------------------
Number of active arms         8        1        7
Accrual rate*             118.0     26.2     91.8
Active arms
Patients for analysis      1809      402      201
Patients recruited**       2358      524      262
All arms
Patients recruited**       2358
-------------------------------------------------
                         ---------Stage 2---------
                        Overall  Control   Exper.
-------------------------------------------------
Number of active arms         6        1        5
Accrual rate*             248.0     70.9    177.1
Active arms
Patients for analysis      2989      854      427
Patients recruited**       4108     1173      587
All arms
Patients recruited**       4632
-------------------------------------------------
                         ---------Stage 3---------
                        Overall  Control   Exper.
-------------------------------------------------
Number of active arms         4        1        3
Accrual rate*             248.0     99.2    148.8
Active arms
Patients for analysis      4719     1887      944
Patients recruited**       4915     1966      983
All arms
Patients recruited**       6613
-------------------------------------------------
```

MRC Clinical Trials Unit

UCL

More on `nstage`, `nstagebin` & `nstagebinopt`

- They allow for intermediate outcome (I) observable before definitive outcome (D) for interim Lack-Of-Benefit (LOB) assessment.
  - e.g., culture status (I-outcome) & failure/relapse in tuberculosis
- They use Dunnett's probability to account for multiplicity.
  - Consider the underlying correlation structure between different tests
  - More efficient than Bonferroni or Sidak corrections
- Feasible/admissible designs found by **`nstagebinopt`** can be saved in a dataset.
  - For further inspection of their properties regarding trial timelines, sample sizes, etc
- Computationally, they are very efficient.
  - 2-arm 2-stage designs: both output the results in less than a second
  - 8-arm 3-stage design: **`nstagebinopt`** (95 seconds), **`nstagebin`** (5 seconds)

Validating/testing `nstage, nstagebin & nstagebinopt`

The FWER and overall power:

- Checked against analytical solutions where possible
- Used `mvnormal` Stata command (Grayling and Mander)

Sample size calculations:

- Compared against Cytel's EAST software and `artbin` Stata command
- Perfect agreement was achieved for a wide range of design types, taking into account differences in rounding

Re-ran the design do files of *MAMS* trials, compared the outputs/results, and checked for error messages and discrepancies.

The commands have been used to design *MAMS* trials in cancer, TB, maternal health, surgery, infections, vascular diseases, etc.

Resources: Articles and examples on `nstage`

Link to the latest article:
bit.ly/48fdcHq

The Stata Journal
Volume 23, Issue 3, September 2023, Pages 774-798

https://doi.org/10.1177/1536867X231196295

Sage Journals

Article and Columns

Facilities for optimizing and designing multiarm multistage (MAMS) randomized controlled trials with binary outcomes

Babak Choodari-Oskooei[1], Daniel J. Bratton[2], and Mahesh K. B. Parmar[3]

Abstract

We introduce two commands, `nstagebin` and `nstagebinopt`, that can be used to facilitate the design of multiarm multistage (MAMS) trials with binary outcomes. MAMS designs are a class of efficient and adaptive randomized clinical trials that have successfully been used in many disease areas, including cancer, tuberculosis, maternal health, COVID-19, and surgery. The `nstagebinopt` command finds a class of efficient "admissible" designs based on an optimality criterion using a systematic search procedure. The `nstagebin` command calculates the stagewise sample sizes, trial timelines, and overall operating characteristics of MAMS designs with binary outcomes. Both commands allow the use of Dunnett's correction to account for multiple testing. We also use the ROSSINI 2 MAMS design, an ongoing MAMS trial in surgical wound infection, to illustrate the capabilities of both commands. The new commands facilitate the design of MAMS trials with binary outcomes where more than one research question can be addressed under one protocol.

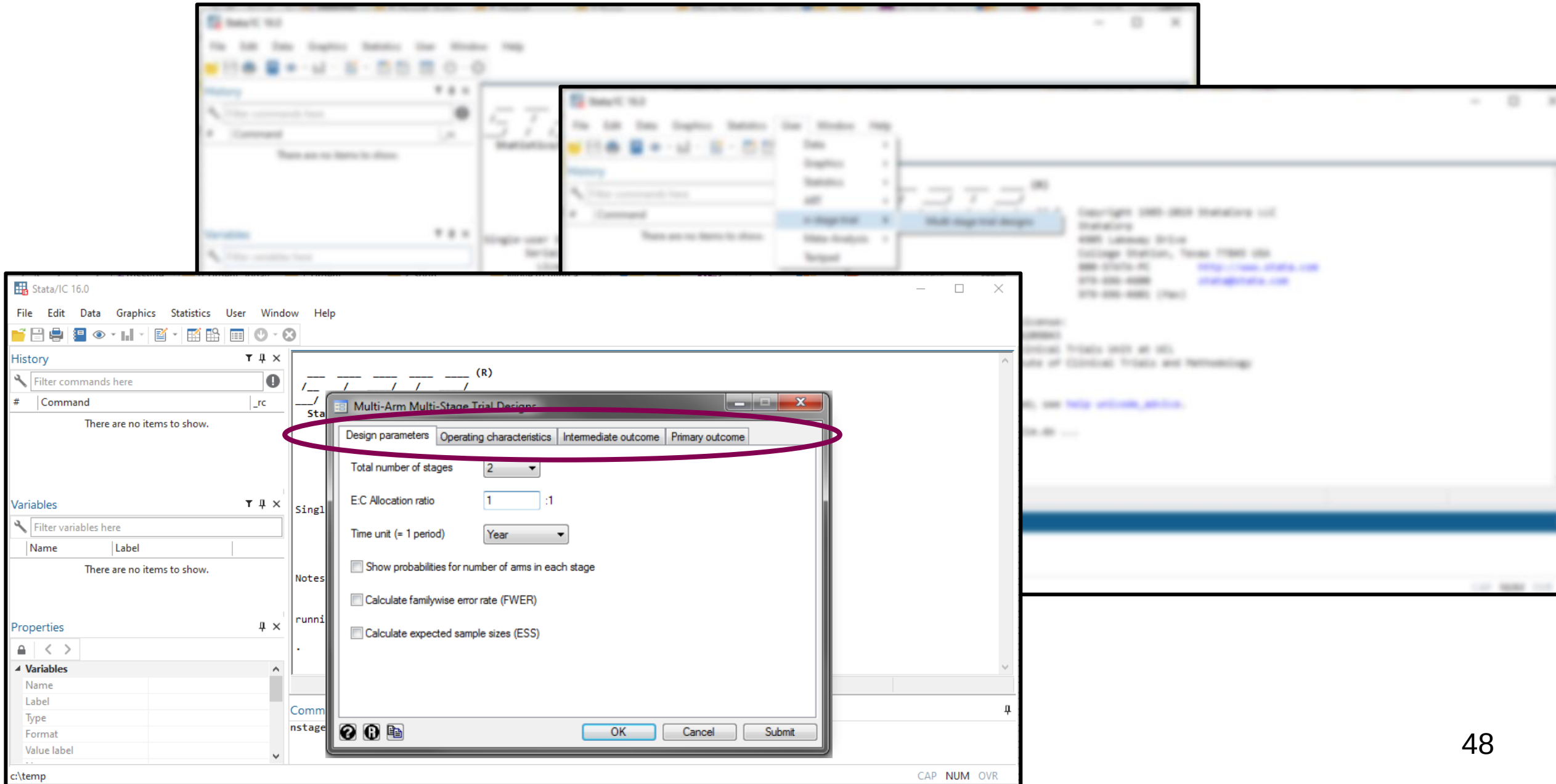
Keywords

st0728, nstagebin, nstagebinopt, multiarm multistage, MAMS, familywise type I error rate, FWER, α functions, adaptive designs

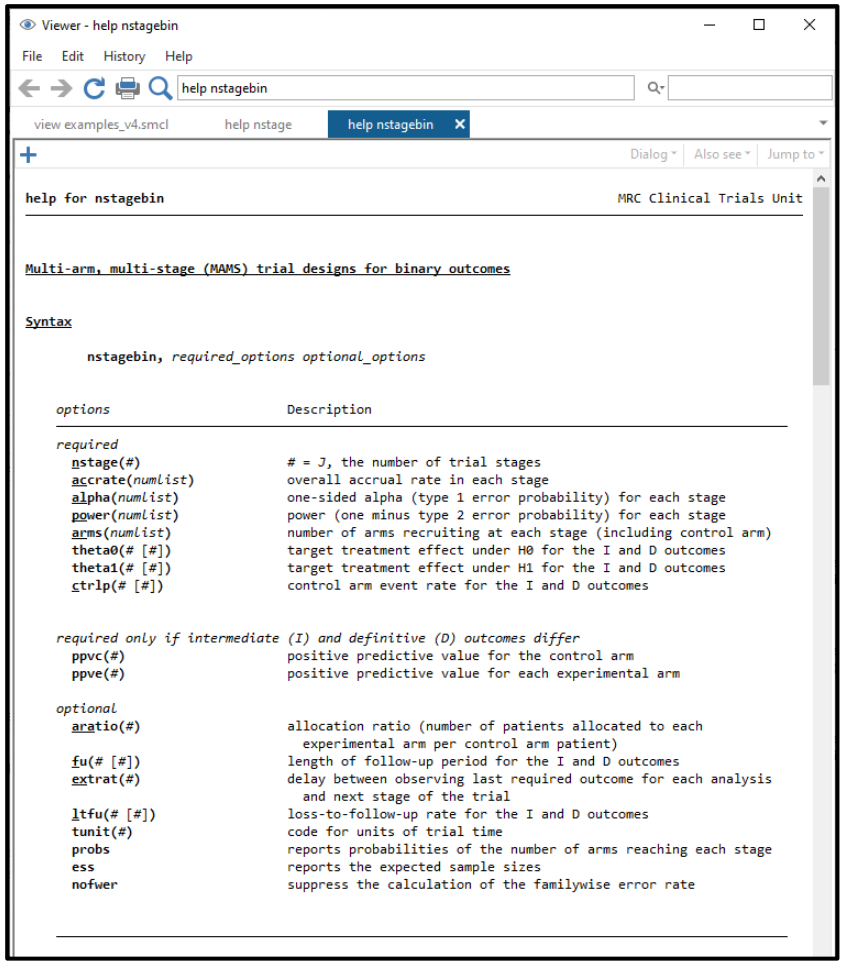
MRC Clinical Trials Unit

UCL

`nstage` Stata suite for *MAMS* designs



`nstage` Stata suite for *MAMS* designs



Viewer - help nstagebin

File Edit History Help

help nstagebin

view examples_v4.smcl | help nstage | help nstagebin

Dialog | Also see | Jump to

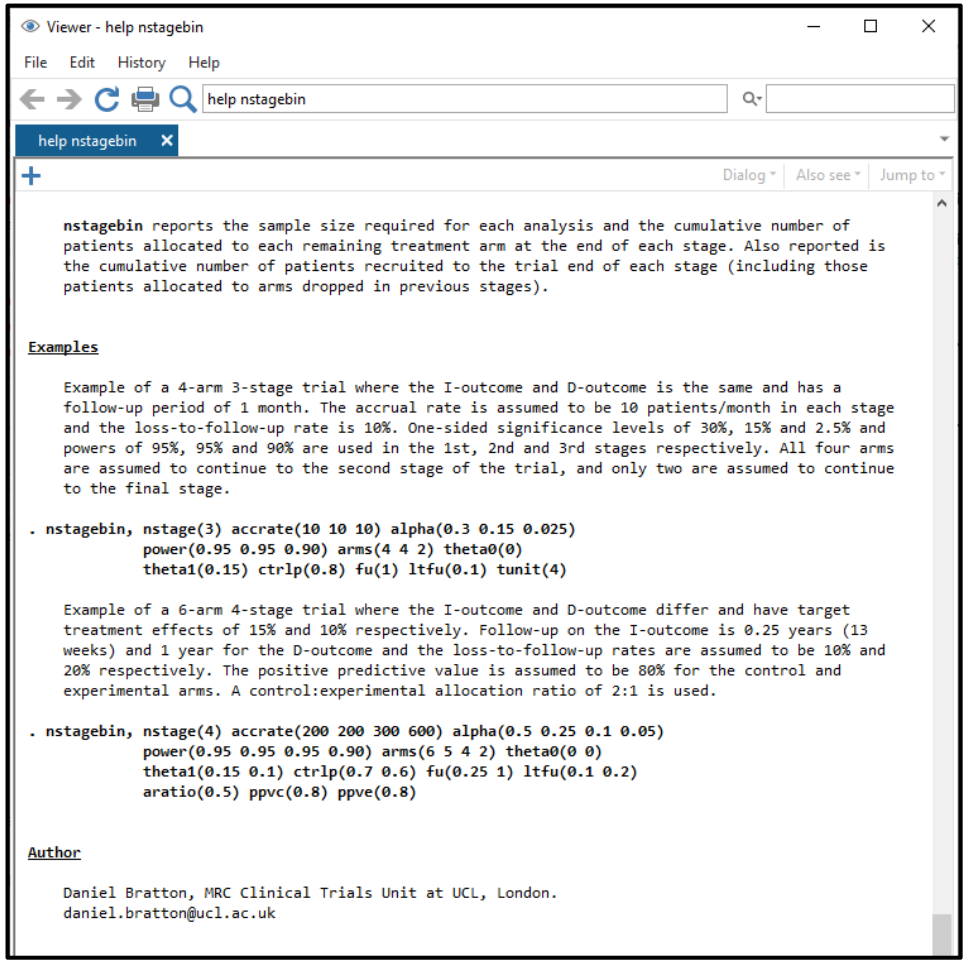
help for nstagebin MRC Clinical Trials Unit

Multi-arm, multi-stage (MAMS) trial designs for binary outcomes

Syntax

nstagebin, *required_options optional_options*

options	Description
required	
nstage(#)	# = J, the number of trial stages
accrate(*numlist***)**	overall accrual rate in each stage
alpha(*numlist***)**	one-sided alpha (type 1 error probability) for each stage
power(*numlist***)**	power (one minus type 2 error probability) for each stage
arms(*numlist***)**	number of arms recruiting at each stage (including control arm)
theta0(# [#])	target treatment effect under H0 for the I and D outcomes
theta1(# [#])	target treatment effect under H1 for the I and D outcomes
ctrlp(# [#])	control arm event rate for the I and D outcomes
required only if intermediate (I) and definitive (D) outcomes differ	
ppvc(#)	positive predictive value for the control arm
ppve(#)	positive predictive value for each experimental arm
optional	
aratio(#)	allocation ratio (number of patients allocated to each experimental arm per control arm patient)
fu(# [#])	length of follow-up period for the I and D outcomes
extrat(#)	delay between observing last required outcome for each analysis and next stage of the trial
ltfu(# [#])	loss-to-follow-up rate for the I and D outcomes
tunit(#)	code for units of trial time
probs	reports probabilities of the number of arms reaching each stage
ess	reports the expected sample sizes
nofwer	suppress the calculation of the familywise error rate



Viewer - help nstagebin

File Edit History Help

help nstagebin

help nstagebin

Dialog | Also see | Jump to

nstagebin reports the sample size required for each analysis and the cumulative number of patients allocated to each remaining treatment arm at the end of each stage. Also reported is the cumulative number of patients recruited to the trial end of each stage (including those patients allocated to arms dropped in previous stages).

Examples

Example of a 4-arm 3-stage trial where the I-outcome and D-outcome is the same and has a follow-up period of 1 month. The accrual rate is assumed to be 10 patients/month in each stage and the loss-to-follow-up rate is 10%. One-sided significance levels of 30%, 15% and 2.5% and powers of 95%, 95% and 90% are used in the 1st, 2nd and 3rd stages respectively. All four arms are assumed to continue to the second stage of the trial, and only two are assumed to continue to the final stage.

```
. nstagebin, nstage(3) accrate(10 10 10) alpha(0.3 0.15 0.025)
             power(0.95 0.95 0.90) arms(4 4 2) theta0(0)
             theta1(0.15) ctrlp(0.8) fu(1) ltfu(0.1) tunit(4)
```

Example of a 6-arm 4-stage trial where the I-outcome and D-outcome differ and have target treatment effects of 15% and 10% respectively. Follow-up on the I-outcome is 0.25 years (13 weeks) and 1 year for the D-outcome and the loss-to-follow-up rates are assumed to be 10% and 20% respectively. The positive predictive value is assumed to be 80% for the control and experimental arms. A control:experimental allocation ratio of 2:1 is used.

```
. nstagebin, nstage(4) accrate(200 200 300 600) alpha(0.5 0.25 0.1 0.05)
             power(0.95 0.95 0.95 0.90) arms(6 5 4 2) theta0(0 0)
             theta1(0.15 0.1) ctrlp(0.7 0.6) fu(0.25 1) ltfu(0.1 0.2)
             aratio(0.5) ppvc(0.8) ppve(0.8)
```

Author

Daniel Bratton, MRC Clinical Trials Unit at UCL, London.
daniel.bratton@ucl.ac.uk

Summary

- `nstage` suite of Stata commands can be used to design efficient *MAMS* trials with a given accrual pattern - available from the SSC.

- Use simulations (*MAMS*) and analytical derivations (two-arm setting) to calculate the operating characteristics.

- Validated against numerous other software and published sample sizes.

- The associated Stata Journal article are available with example trials and codes.

- Further work will allow use of any combination of outcomes (e.g. continuous I outcome, binary D outcome) and incorporating treatment selection at early stages.

Key references

Design

- Choodari-Oskooei, et al. (2023) Treatment selection in multi-arm multi-stage designs: With application to a postpartum haemorrhage trial. *Clinical Trials*. 20(1):71-80. doi:10.1177/17407745221136527

- Parmar, et al. (2017) Testing many treatments within a single protocol over 10 years at MRC Clinical Trials Unit at UCL: Multi-arm, multi-stage platform, umbrella and basket protocols. *Clinical Trials*.14(5):451-461. doi:10.1177/1740774517725697

- Royston, et al. (2003) Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. *Statistics in Medicine*. 22:2239–2256. doi: 10.1002/sim.143

Key references

Analysis

- Choodari-Oskooei, et al. (2013) Impact of lack-of-benefit stopping rules on treatment effect estimates of two-arm multi-stage (TAMS) trials with time to event outcome. *Trials* 14, 23. https://doi.org/10.1186/1745-6215-14-23
- Barthel, et al. (2009) How do multi-stage, multi-arm trials compare to the traditional two-arm parallel group design – a reanalysis of 4 trials. *Trials* 10, 21. https://doi.org/10.1186/1745-6215-10-21

Conduct

- Schiavone, et al. (2019) This is a platform alteration: a trial management perspective on the operational aspects of adaptive and platform and umbrella protocols. *Trials* 20, 264. https://doi.org/10.1186/s13063-019-3216-8
- Morrell, et al. (2019) Mind the gap? The platform trial as a working environment. *Trials* 20, 297. https://doi.org/10.1186/s13063-019-3377-5
- Sydes, et al. (2012) Flexible trial design in practice - stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: a multi-arm multi-stage randomized controlled trial. *Trials* 13, 168. https://doi.org/10.1186/1745-6215-13-168

Key references

Software:

- Choodari-Oskooei, et al. (2023) Facilities for optimizing and designing multi-arm multi-stage (MAMS) randomized controlled trials with binary outcomes. *The Stata Journal*; 23(3), 774–798. doi: https://doi.org/10.1177/1536867X231196295
- Bratton, et al. (2015) A menu-driven facility for sample-size calculation in multi-arm, multi-stage randomized controlled trials with time-to-event outcomes: Update. *The Stata Journal*.15(2):350-368. doi: https://doi.org/10.1177/1536867X1501500202

Book chapter:

- Choodari-Oskooei, et al. (2022) Multi-arm multi-stage (MAMS) platform randomized clinical trials. In: Principles and Practice of Clinical Trials. Springer. bit.ly/3tmx0qT

Video tutorial:

- Tutorial on the `nstage` suite of commands: bit.ly/3Mxpzal
- Tutorial on MAMS designs: bit.ly/3SEPEGh; bit.ly/3X1Hg5q

MRC Clinical Trials Unit

UCL

Acknowledgements

Thanks to:

- Colleagues & PhD students at the MRC CTU at UCL: Past and present
  - Alexandra Blenkinsop (Imperial College London)
  - Daniel Bratton (GSK)
  - Ian White (UCL)
  - Matthew Sydes (UCL)
  - Max Parmar (UCL)
  - Patrick Royston (UCL)
  - Sophie Barthel (PRA Health Sciences)
- All the collaborators: particularly, those involved in STAMPEDE and ROSSINI-2 trials
- Stata Corps and the organisers of this fantastic symposium

MRC Clinical Trials Unit

UCL

Thank you for your attention

Happy to take questions!
Send it to: b.choodari-Oskooei@ucl.ac.uk

MRC Clinical Trials Unit

UCL