Using Mata to accelerate the familywise error rate calculation for multi-arm multi-stage clinical trial designs

Daniel Bratton

MRC Clinical Trials Unit at UCL

12th September 2014

Overview

- Multi-arm multi-stage (MAMS) designs
- Familywise error rate (FWER)
- How to calculate FWER using simulation
- Speed of simulation using Stata only
- How to perform the calculation in Mata — pointers
- Comparison of speed of calculations
- Summary

Multi-arm multi-stage (MAMS) designs

- MAMS designs are aimed at accelerating the evaluation of new therapies over more conventional approaches.
- **Multi-arm** — evaluate multiple new treatments in a single trial against a common control arm.
- **Multi-stage** — evaluate each arm at a series of interim analyses, ceasing recruitment to poorly performing arms.
- Interim assessments can be made on an intermediate outcome (I) which is on the causal pathway to the primary outcome (D) of the trial.
- e.g. in cancer, I = progression-free survival (PFS) and D = overall survival (OS).

Example — 3-arm 2-stage $I = D$ trial

```
nstage, nstage(2) alpha(0.5 0.025) omega(0.95 0.9) hr0(1 1) hr1(0.75 0.75) ///
        t(1 1) accrue(250 250) arms(3 3)
```

Operating characteristics

	Alpha(1S)	Power	HR|H0	HR|H1	Crit.HR	Length*	Time*
Stage 1	0.5000	0.950	1.000	0.750	1.000	1.927	1.927
Stage 2	0.0250	0.901	1.000	0.750	0.842	2.584	4.511
Pairwise	0.0230	0.871				4.511	

* Length (duration of each stage) is expressed in one year periods and assumes survival times are exponentially distributed

Sample size and number of events

	Stage 1			Stage 2		
	Overall	Control	Exper.	Overall	Control	Exper.
Arms	3	1	2	3	1	2
Acc. rate	250	83	167	250	83	167
Patients*	482	161	321	1128	376	752
Events**	192	72	120	725	261	464

* Patients are cumulative across stages

** Events are cumulative across stages, but are only displayed for those arms to which patients are still being recruited

** Events are for the same outcome at stages 1 and 2

Familywise error rate (FWER)

- The FWER of a MAMS trial is the probability of recommending at least one ineffective treatment at the end of the trial.
- This often has to be controlled at some conventional level (e.g. 5%) especially in a confirmatory trial.
- FWER is maximised under the global null hypothesis, H_G (i.e. when H_0 is true for all experimental arms).
- Calculating the FWER under H_G is therefore of prime interest.
- The FWER of a MAMS design can be calculated using simulation.

Notation

- K experimental arms, J stages ($(K+1)$-arm J-stage trial).
- Denote stages by $j = 1, \ldots, J$ and experimental treatment arms by $k = 1, \ldots, K$.
- Denote the standardised test statistic (e.g. z-statistic for the log hazard ratio) for the kth arm in stage j by Z_{jk}.
- Ignoring stopping rules, $Z_{jk} \sim N(0, 1)$ under H_0.

Requirements for calculation

$I = D$ time-to-event outcomes

- Need to simulate $Z_{jk} \sim N(0, 1)$ such that:

$$\text{corr}(Z_{jk}, Z_{j'k}) = \sqrt{\frac{e_j}{e_{j'}}}\ ,\ j' > j \tag{1}$$

$$\text{corr}(Z_{jk}, Z_{jk'}) = \frac{A}{A+1}\ ,\ k \neq k' \tag{2}$$

  where e_j is the number of control arm events in the jth stage and A is the ratio between the number of events observed in the experimental and control arms.

- This can be done using a generalisation of a simulation procedure by Wason and Jaki (2012) for $I = D$ MAMS designs with equal group sizes.

FWER calculation

$I = D$ time-to-event outcomes

- Generate $x_{jk} \sim N(0,1)$ ($j = 1, \ldots, J$; $k = 0, \ldots, K$) such that

$$\text{corr}(x_{jk}, x_{j'k}) = \sqrt{\frac{e_j}{e_{j'}}}$$

- Can be achieved using the `drawnorm` command in Stata, e.g.

```
drawnorm x11 x21, corr(S) sd(‘sd’) n(250000) double
```

- 250,000 replicates provides a precise estimate of the FWER and takes only a few seconds to run (depending on J and K).

FWER calculation

- Then use the following formula to simulate Z_{jk} ($j = 1, \ldots, J$; $k = 1, \ldots, K$):

$$Z_{jk} = \sqrt{\frac{A}{A+1}} x_{j0} + \sqrt{\frac{1}{A+1}} x_{jk}$$

- The FWER is the proportion of replicates which, for any k:

$$Z_{jk} < z_{\alpha_j} \text{ for all } j = 1, \ldots, J$$

where α_j is the significance level for each comparison in the jth analysis.

FWER calculation — output

Stagewise significance levels: $\alpha_1 = 0.50, \; \alpha_2 = 0.025$

z11	z21	z12	z22	pass11	pass21	pass12	pass22
.57464456	.28498412	.63879568	.44834618	0	0	0	0
.82476848	.73010319	.32077323	.13767086	0	0	1	0
.83623639	.5763814	.61637843	.57662708	0	0	0	0
.96220493	.8850676	.92685639	.63945521	0	0	0	0
.34649736	.16833927	.54588476	.11634171	1	0	0	0
.9857904	.68945553	.57506626	.82511308	0	0	0	0
.59947405	.30840953	.69953476	.78644393	0	0	0	0
.06774324	.00540695	.35532399	.05308241	1	1	1	0
.66492092	.64917445	.72497049	.2583492	0	0	0	0
.80596521	.49308509	.82937441	.86934286	0	0	0	0

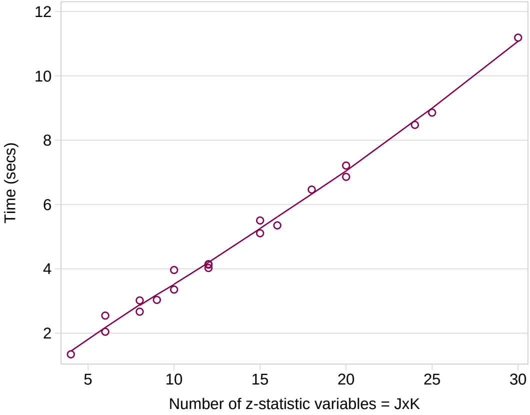
Type I error!

Implementation

- This calculation estimates the FWER to be 4.1% (SE 0.04%) for the 3-arm 2-stage design shown earlier.
- The methods are applicable to any outcome analysed using a normally distributed test statistic.
- A subroutine for calculating the FWER of a MAMS design has been added to `nstage` which now estimates FWER by default.
- Other useful things can be calculated such as expected number of events or expected sample size (a measure of the 'efficiency' of the design).

Speed of `nstage` with FWER calculation

Below is the time taken for `nstage` to output the design of MAMS designs with $K = 2 - 6$ experimental arms and $J = 2 - 5$ stages (20 designs in total) using Intel Core2 Duo 3GHz processor and 2GB RAM.



Accelerating the calculation using Mata

- Although the calculation is quite quick we often have to search through lots of designs to find the most efficient or suitable one to use in practice.
- Thus the total computing time will be very long!
- This could be considerably reduced by performing the calculation using Mata.
- Instead of generating a dataset, we can generate a $J \times K \times R$ matrix containing the simulated Z_{jk} where R is the total # of replicates.
- This involves the use of ‘pointers’ to generate 3-dimensional matrices in Mata.

Pointers

- A pointer contains the address of another variable or matrix.
- Thus each element of a 2D matrix of pointers could point to a vector of numbers, so it is effectively a 3D matrix.
- They use two operators: the reference (`&`) and the dereference (`*`).
- `&` instructs what the pointer, `p`, should point to, e.g. `p=&x`
- `*` is then used ask what the pointer is pointing to, e.g. `*p`
- Much of what I learnt about pointers is from http://www.ssc.wisc.edu/sscc/pubs/4-26.htm and Stata's Mata manual.

Generating a 3D matrix in Mata

To generate a $3 \times 3 \times 3$ matrix of zeroes, P, we do the following

```
P = J(3,3,NULL)

for (i=1; i<=rows(P); i++) {
    for (j=1; j<=cols(P); j++) {

        P[i,j] = &J(3,1,0)

    }
}
```

Using 3D matrices in Mata

To work with an element of the pointer, `P[i,j]`, we use the following syntax:

```
*(P[i,j])
```

To work with an element in the 3D matrix, `P[i,j,k]`, we use

```
(*(P[i,j]))[k]
```

Using 3D matrices in Mata

```
. mata:
------------------------------------------------- mata (type end to exit) ------
:
: P = J(3,3,NULL)

: for (i=1; i<=rows(P); i++) {
>         for (j=1; j<=cols(P); j++) {
>                 P[i,j] = &J(3,1,0)
>         }
> }

:
: *(P[1,1])
            1
     +-----+
   1 |  0  |
   2 |  0  |
   3 |  0  |
     +-----+

:
: (*(P[1,1]))[2]
  0

:
: end
--------------------------------------------------------------------------------
```

FWER calculation in Mata - generating the x_{jk}

- First generate $J \times (K+1)$ matrix of pointers X each of which point to a $R \times 1$ vector of zeroes
- Then do the following

```
Sc = cholesky(S)
for (r=1 ; r<=R ; r++) {
    for (k=1 ; k<=K+1 ; k++) {

        y = invnormal(runiform(J,1))
        x = Sc*y

        for (j=1 ; j<=J ; j++) {
            (*(X[j,k]))[r] = x[j]
        }
    }
}
```

FWER calculation in Mata - generating the z_{jk}

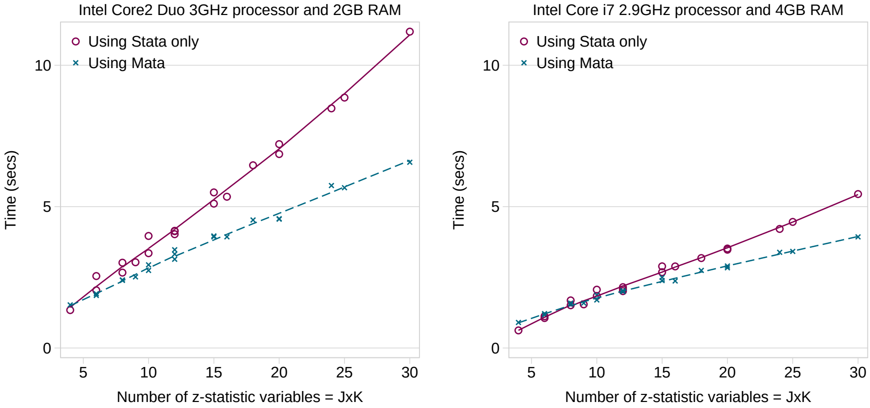
- Generate $J \times K$ matrix of pointers Z each of which point to a $R \times 1$ vector of zeroes
- Then do the following

```
for (j=1 ; j<=J ; j++) {
    for (k=1 ; k<=K ; k++) {

        *(Z[j,k]) = (sqrt(A/(A+1))**(X[j,1])
                        + sqrt(1/(A+1))**(X[j,k+1])
                          :> -invnormal(a[j]))

        if (j>1) *(Z[j,k]) =
                      *(Z[j,k]):**(Z[j-1,k])
    }
}
```

Comparison of speed of FWER calculations



Summary

- We have developed a calculation for the FWER of MAMS designs and implemented it into the `nstage` program for MAMS trials with time-to-event outcomes.
- `nstage` now outputs the FWER by default
- The calculation works by simulating the joint distribution of the z-test statistics for each arm at each stage.
- Calculating the FWER of a MAMS design using Stata alone is relatively quick for small J and K.
- However, a faster calculation would make searching over multiple designs to find the most suitable one more practical

Summary

- Computing time can be considerably reduced by performing the calculation in Mata, particularly for large J and K
- This can be accomplished through the use of pointers to generate 3-dimensional matrices of the simulated z_{jk}.
- However, there is little difference in speed for fast machines and for smaller J and K the Mata calculation may even be slower!
- It's very probable that better programming is what's needed.

References

Barthel, F. M.-S., Royston, P., & Parmar, M. K. B. (2009). A menu-driven facility for sample-size calculation in novel multi-arm, multi-stage randomized controlled trials with a time-to-event outcome. *Stata Journal*, *9*(4), 505-523(19).

Bratton, D. J., Choodari-Oskooei, B., & Royston, P. (2014). A menu-driven facility for sample size calculation in multi-arm multi-stage randomised controlled trials with time-to-event outcomes: Update. *Stata Journal*. (In press)

Bratton, D. J., Phillips, P. P. J., & Parmar, M. K. B. (2013). A multi-arm multi-stage clinical trial design for binary outcomes with application to tuberculosis. *Medical Research Methodology*, *13*, 139.

Royston, P., Barthel, F. M., Parmar, M. K., Choodari-Oskooei, B., & Isham, V. (2011). Designs for clinical trials with time-to-event outcomes based on stopping guidelines for lack of benefit. *Trials*, *12*, 81.

Royston, P., Parmar, M. K., & Qian, W. (2003). Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. *Stat Med*, *22*(14), 2239–56.

Wason, J. M., & Jaki, T. (2012). Optimal design of multi-arm multi-stage trials. *Stat Med*, *31*(30), 4269-79.