Bringing patient population size into clinical trial design using response-adaptive randomisation

Sofía S. Villar

Biometrika Postdoctoral Fellow

† MRC Biostatistics Unit, Cambridge

Clinical trials in small populations workshop:
Forwards-Looking Forum

MRC | Hubs for Trials Methodology Research

MRC | Biostatistics Unit

1st December

Outline

Choosing sample sizes

Error control vs. Patient benefit

Two different criteria to choose CTs sample sizes n:

(1) controlling for error probabilities (traditional type I and II) and (2) maximising of the expected number of successfully treated patients in a patient population (N).

- Resulting sample sizes are similar only for "*large*" values of N
- Cheng et al (2003) using decision analysis showed that the optimal sample size of a 2-armed RCT (optimal in terms of patient benefit) is $\propto \sqrt{N}$.
- Thus, *large* N means $>>$ 1-4 million patients which results in trials of size 1000-2000 (typical Phase III trial size).
- The European Medicines Agency defines a rare disease as a condition affecting < 5 in 10,000 people in the EU.

Choosing sample sizes: how much research?

Ultrarare: Rare within Rare

- The suboptimality gap (in terms of patient benefit) of the error control (EC) sample size approach is larger as $N \to 0$.
- If $N \approx 400000$, $n^* \approx 600$ while for $N \approx 1000$, $n^* \approx 32$!
- So within *rare*, the optimal research-practice balance is not the same (Annular pancreas vs. Hutchinson-Gilford progeria)
- EC approach does not consider population size to determine n.
- For rare conditions we have to include N to "optimally" draw the line between research and practice.
- The *sample size* with the highest possible expected patient benefit given the information available results from a fully sequential Bayesian optimal decision model. (Bandit models)

Choosing sample sizes through *online learning*

Bandit Models: idealism vs. pragmatism

- Bandit models find n^* through online learning.
- The *right level* of experimentation (research) depends on:
  * what the data suggest might be the underlying treatment difference among the arms
  * how many more patients stand to benefit from that difference
- However, it's not all roses in the bandit world...
  * Bandits' solution is computationally expensive, is not easily summarised and does not impose type I and II error rates.
  * Patients' allocations to treatments during the learning phase are most of the time not randomised.

How to implement? $\rightarrow$ *Bayesian* response-adaptive randomisation

Outline

Balancing goals through Adaptive randomisation

Myopic vs. Forward-looking rules

- Response-adaptive randomisation (RAR) rules *adapt* allocation probabilities based on the outcome and allocation information of previous patients to meet certain goal.
- Bayesian RAR aligns allocation probabilities with treatment efficacy as information about it accumulates.
- Simulation results indicating the advantage of Bayesian BAR over equal randomisation Eick and Berry (1995). CTs and RAR: Giles et al (2003). I-SPY2, BATTLE,...
- These RAR offer the possibility of balancing dual (sometimes conflicting) goals. Yet, they are based on past information only (i.e. they are myopic procedures).
- Non-myopic rules offer further benefits Villar et al (2015).

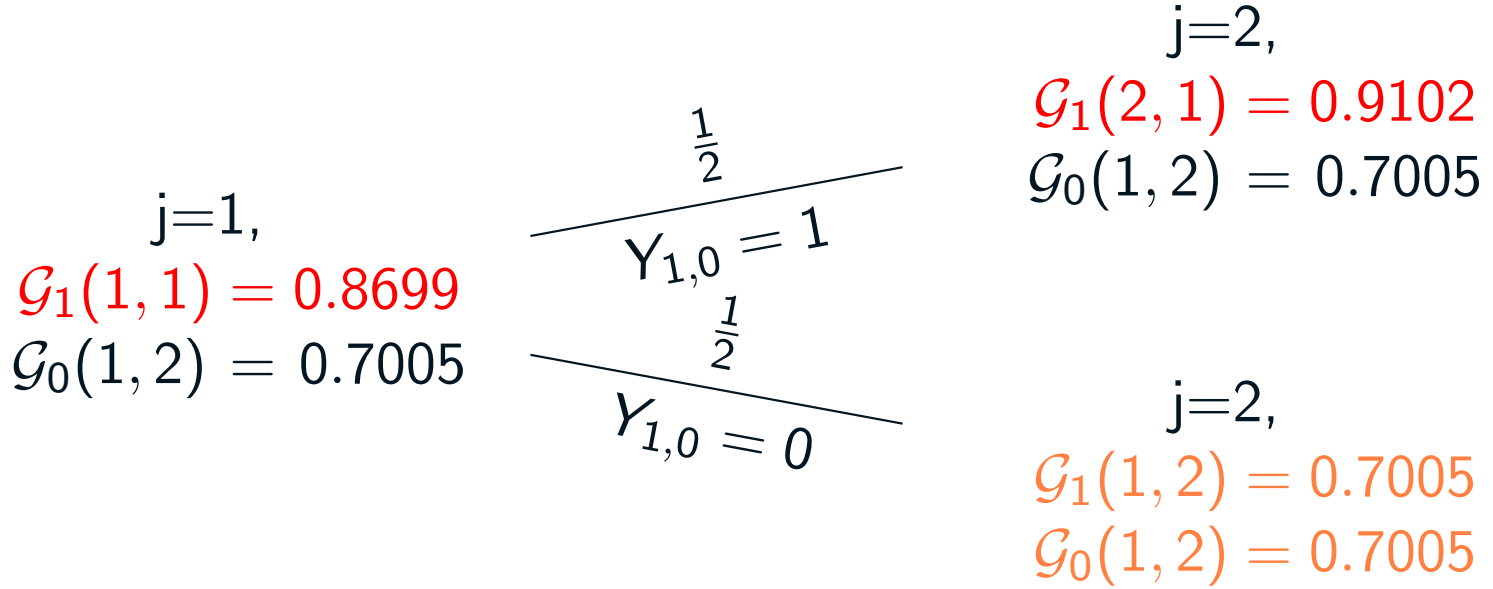
Using both Posterior and Predictive probabilities

How much research/practice given what we know and might happen

n patients enrolled sequentially in groups of size b over J stages. Villar et al (2015) defined group allocation probabilities as follows:

$\pi_{k,j}$: the probability of allocating arm k at stage j, which is common to all patients in block j, when using the Gittins index (Gittins and Jones, 1979) and given data observed up to block j-1.

Example: $b = 2$



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

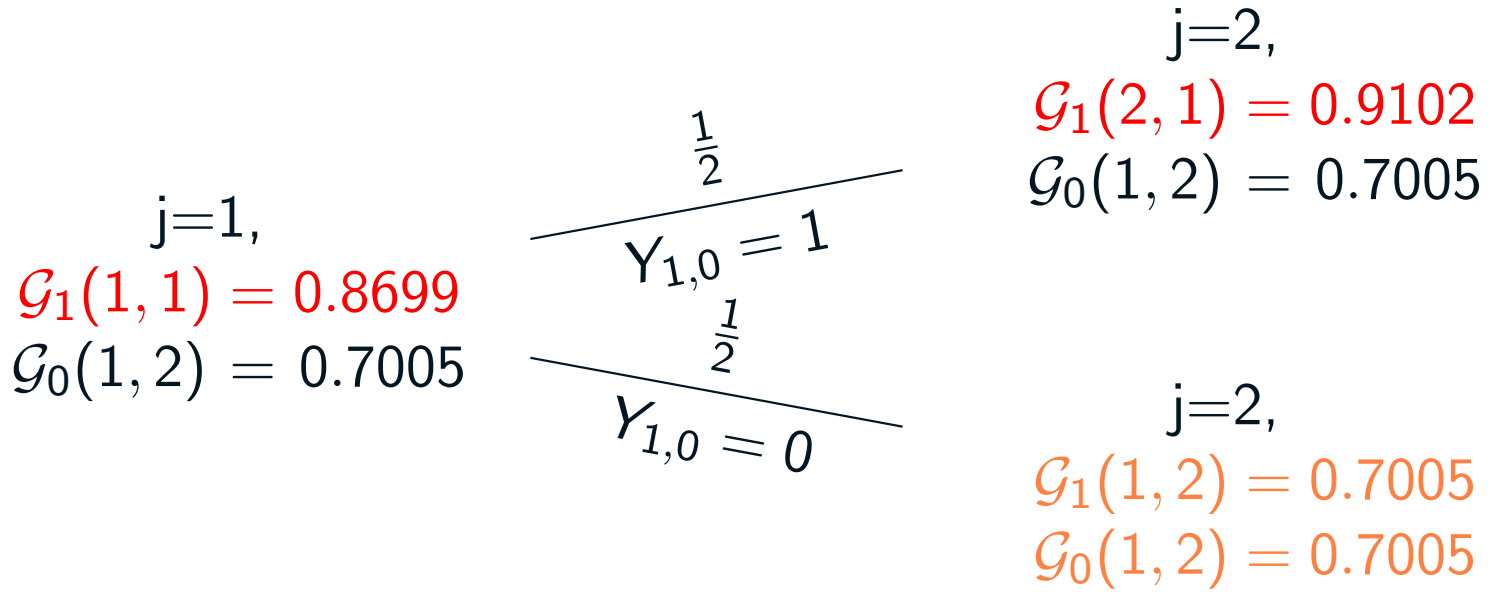
Using both Posterior and Predictive probabilities

How much research/practice given what we know and might happen

n patients enrolled sequentially in groups of size b over J stages. Villar et al (2015) defined group allocation probabilities as follows:

$\pi_{k,j}$: the probability of allocating arm k at stage j, which is common to all patients in block j, when using the Gittins index (Gittins and Jones, 1979) and given data observed up to block j-1.

Example: $b = 2$



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

The Forward Looking Gittins Index

Example: Redesigning a real trial

NeoSphere is a 4-arm FR trial in breast cancer with 417 patients. The response rates reported in group A, B, C and D respectively were 29.0%, 45.8%, 16.8% and 24.0%.

	$H_1 : \boldsymbol{p_1} = [0.29\ 0.458\ 0.168\ 0.24]$		
	$(1-\beta)$	p^* (s.e.)	*ENS* (s.e.)
FR	0.653	0.250 (0.02)	120.88 (9.34)
Trippa et al	0.895	0.451 (0.04)	150.98 (10.3)
C FLGI (block=9)	0.816	0.665 (0.06)	166.40 (11.9)
Thompson Sampling	0.782	0.585 (0.10)	155.93 (13.4)
FLGI (block=9)	0.177	0.804 (0.09)	174.11 (13.3)
GI	0.140	0.840 (0.10)	177.97 (13.0)
UB		1	190.99 (0.00)

with the $\pi_{k,j}$ probabilities computed via Montecarlo simulation.

Outline

Using both Posterior and Predictive probabilities

- For common conditions the paradigm based on constraining on error probabilities determines sample sizes which provide substantial overall patient benefit.
- For rare conditions, patient population is the constraint and unless we introduce it in the design this cannot provide sufficient overall patient benefit.
- Forward looking response-adaptive randomisation provide a practical way of implementing this. Another example of implementable rules: poster by Williamson et al (2015).
- Further improvements when we add the 'B' for Bayesian to the response-adaptive rule by using historical data appropriately in poster by Bennett et al (2015).

References I

Questions & Comments

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

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

Adaptive assignment versus balanced randomization in clinical trials: a decision analysis (1995) Berry, Donald A and Eick, Stephen G *Statistics in medicine* **14** (3)231–246

Cheng, Yi and Su, Fusheng and Berry, Donald A (2003) Choosing sample size for a clinical trial using decision analysis *Biometrika* **90** (4)923–936

Giles, F. J., Kantarjian, H. M, Cortes, J. E, Garcia-Manero, G., Verstovsek, S. et al Adaptive randomized study of Idarubicin and Cytarabine versus Troxacitabine and Cytarabine versus Troxacitabine and Idarubicin in untreated patients 50 years or older with adverse Karyotype Acute Myeloid Leukemia. *Journal of Clinical Oncology*, 21 (9) 1722-1727 American Society of Clinical Oncology

References II

Questions & Comments

Hu, F., and Rosenberger, W. F. (2006). *The theory of response-adaptive randomization in clinical trials*. John Wiley & Sons.

Villar, S., Bowden, J. and Wason, J. (2015) Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges. *Statistical Science* **30**, *No. 2, 199215*.

Villar, S., Wason, J. and Bowden, J. (2015) Response-adaptive Randomization for Multi-arm Clinical Trials using the Forward Looking Gittins Index rule *To appear in Biometrics*.

Trippa, L., Lee, E., Wen, P., Batchelor, T., Cloughesy, T., Parmigiani, G. et al (2012) Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. *Journal of Clinical Oncology*, **30**(26), 3258–3263.

Berry, Donald A (2011) Adaptive clinical trials: the promise and the caution *J*ournal of Clinical Oncology,**29** (6) 606–609. American Society of Clinical Oncology.

Questions & Comments

Useful vs. Indisputable

Thanks for the attention! ☺