Balaam's design revisited

Julia Singer, PhD, Chief Scientific Officer, Accelsiors LTD

Joëlle Monnet Gaud, Director, Biostatistics, Fresenius-Kabi SwissBioSim

Disclaimer

The views and opinions expressed in the following presentation are those of the individual presenter and should not be attributed to, or considered as reflecting the position of Fresenius-Kabi or its management.

STRUCTURE OF THE PRESENTATION

- Background
- Design
- Models fitted in different publications
- Power considerations
- Possible ways to increase efficiency
- Conclusions

1. BACKGROUND

- Biologic drug inter-subject variability can be relatively large
  - some subject baseline characteristics can influence the PK profile and therefore impact inter-subject variability.
- Cross over designs could be more appropriate
  - intra-subject variability is not impacted by subject baseline characteristics

➡ Would some cross-over designs help when variability is large and half life not defined as short?

- Efficiency of «less standard» study designs was investigated, in particular the Balaam design
  - By estimating carry-over and treatment effects, could we propose a study design with a reasonable washout-period, while still benefiting from an increased power/lower sample size compared to parallel study design?

2. DESIGN (*N.L.Balaam: A Two-period Design with t^2 Experimental Units, Biometrics*; t=nr. of treatments)

Sequence	Period 1	Period 2
1	A	A
2	B	B
3	A	B
4	B	A

- Sequences 3 and 4 – cross-over part
- Sequences 1 and 2 – parallel part, enable the estimation of a carry-over effect and also a treatment-specific within-subject variability

Primary objective: to estimate the difference (CI) between treatment means. For PK studies on a log scale, CI level =90%.

3. BASE MODEL (questionnable terms highlighted)

$$Y_{ijk} = \mu + G_k + F_{(j.k)} + P_j + C_{(j-1,k)} + S_{ik} + \varepsilon_{ijk}$$

Y_{ijk}=outcome measured for subject *i* from sequence *k* in period *j*
μ=intercept
G_k=effect of sequence *k*
$F_{(j,k)}$=effect of the treatment received in sequence *k*, period *j*
P_j=effect of period *j*
$C_{(j-1,k)}$=carryover effect in period *j* (resulting from period *j-1*) in sequence *k*
S_{ik}= random subject effect for subject *i* from sequence *k*
ε_{ijk} = residual error for subject *i* from sequence *k* in period *j*
$S_{ik} \sim N(0,s_b^2)$ – s_b^2 corresponding to the between-subject variance
$\varepsilon_{iik} \sim N(0,s_w^2)$ – s_w^2 corresponding to the within-subject variance
$i=1,\ldots,n_k$
j=1,2
k=1. 2, 3, 4
S_{ik} and ε_{ijk} are independent

4. MODELS IN DIFFERENT PUBLICATIONS

Model 1 (Chow, SC and Liu, JP [2], R package [4])

- S_{ik}= fixed subject effect for subject *i* from sequence *k*

Model 2 (R package)

- S_{ik}= random subject effect for subject *i* from sequence *k*

Model 3 (Jones & Kenward [3], Brown & Prescott [5])

- G_k= effect of sequence *k* deleted from the model
- S_{ik}= fixed subject effect for subject *i* from sequence *k*

Model 4 (Brown & Prescott [5])

- G_k=effect of sequence *k* deleted from the model
- S_{ik}= random subject effect for subject *i* from sequence *k*

5. ARE THESE MODELS REALLY DIFFERENT?

Model	Estimate
1 – fixed subject	
2 – random subject	
3 – fixed subject – NO sequence	
4 - random subject – NO sequence	

Models (1)-(3) lead to exactly the same estimate

- Jones & Kenward: in the Appendix of their book they prove that the model with or without sequence effect leads to the same estimator if a fixed subject effect is included in the model (Models 2 and 3).
- The R package offers two, apparently different choices but there is no difference in fact.
- In Chow, SC and Liu, JP 's book it is emphasized that it is important to have sequence effect in the model to test the success of the randomization. Although they also mention that if sequence effect is included in the model then the treatment effect is only estimated from the cross-over part (i.e. sequences 3 and 4).

6. EFFICIENCY IN TERMS OF POWER

If sequence effect is in the model: treatment effect is only estimated from the cross-over part (sequences 3 and 4). On the other hand, Model 3 is equivalent to Model 2.

Therefore, the most information about treatment difference is obtained when no sequence effect is included in the model and subject effect is considered as random (Model 4).

Example (simulated, 10000 cases)

PK study (assuming a log-normal distribution), within-subject CV=20%, total CV=45%, GMR between treatment arms 1.05, adding (on the log scale) small effects for sequence, period and carry-over. Equivalence margins for the 90% CI of GMR (0.8, 1.25).
N=12 subjects per sequence (provides a 90% power for the 2x2 cross-over).

Power to conclude equivalence for models (1-3):	64%
Power to conclude equivalence for model 4:	88%
Power to detect difference between the carry-over terms for models (1-3):	39%
Power to detect difference between the carry-over terms for model 4:	31%

EFFICIENCY IN TERMS OF POWER (continued)

"If between-patient variation is smaller, the existence of the AA and BB treatment groups would lead us to expect greater benefits from the mixed models approach." (Brown&Prescott)

Example (continued): all the assumptions as before, total CV=45%, within-subject CV in (15%, 20%, 25% -> between-subject variance decreasing)

Power to conclude equivalence

Within-subject CV	Model 1	Model 4	Diff (4 vs 1)
15%	88%	97%	11%
20%	*64%*	*88%*	24%
25%	37%	79%	42%

EFFICIENCY IN TERMS OF POWER (continued)

Benefits from the mixed models approach can be even larger if drop-out cases occur between the two periods (i.e. subjects with missing data contribute to the estimation in Period 1).

Example (continued): all the assumptions as before, a 10% drop-out rate (MCAR) was introduced in period 2.

Power to conclude equivalence

Within-subject CV	Model 1 No drop-out	Model 4 No drop-out	Model 1 With drop-out	Model 4 With drop-out
15%	88%	97%	79%	93%
20%	*64%*	*88%*	51%	82%
25%	37%	79%	24%	71%

7. CONCLUSIONS

- Apparently different models may lead to the same estimators :
  - ✓ If subject is inherently nested into sequence (by study design) and subject is a fixed effect in the model then it does not matter whether sequence effect is included in the model or not, the estimate of treatment effect is the same
- The efficiency of a model in terms of power does not depend only on its goodness of fit but also on the null and alternative hypothesis of interest
  - ✓ i.e. in this case models were nested, and still the model with less parameters was more powerful for this specific null hypothesis (for equivalence of treatment means).
  - ✓ Residual degrees of freedom have little influence on power compared to the influence of variances contributing to the estimation of SE (as pointed out by Stephen Senn [6])

REFERENCES

1. Balaam N.L.: A Two-period Design with t^2 Experimental Units, *Biometrics*, Vol. 24, pp. 61-73, 1968;
2. Chow S.C., Liu, J.P.: Design and Analysis of Bioequivalence Studies. CRC Press, Third Edition, 2009.
3. Jones B, Kenward MG: Design and Analysis of Cross-Over Trials. CRC Press, Third Edition, 2015
4. Schütz, H: replicateBE: Average Bioequivalence with Expanding Limits (ABEL). Version 1.0.11, R-project.
5. Brown H, Prescott R: Applied Mixed Models in Medicine, Second Edition, Wiley, 2006.
6. Stephen Senn: Book Review: Design and Analysis of Bioequivalence Studies. Biometrics, Vol. 56, pp. 959-960, 2000
7. Lu Y, Chow SC and Zhang ZZ: Statistical Designs for Assessing Interchangeability of Biosimilar Products. Drug Des: 2013