Implementation details of the power calculations via simulations for scaled ABE

Version 0.03 (14-May-2013)

Author: Detlew Labes

EMA method (`power.scABE()`)

Method description in a cook book manner:

- Evaluate all data (log-transformed) via an ANOVA equal to the classical cross-over design with treatment, period, sequence and subject within sequence.
  Get the point estimate (pe) for T-R and the mse from that ANOVA.
  The 90% confidence interval is obtained from pe and mse according to

$$[lL, uL] = pe \pm t_{(1-alpha),df} * \sqrt{mse * b_{k(ni)} * \sum \frac{1}{n_i}}$$

  The term under the square root is s_d^2. The term $b_{k(ni)} * \sum \frac{1}{n_i}$ is named C2.
- Evaluate the data (log-transformed) for the reference only via an ANOVA with period, sequence and subject within sequence. The mse of that evaluation is s_{wR}^2 (within-subject variance for the reference). It has df(RR) degrees of freedom associated.
- If $CV_{wR} = sqrt(\exp(s_{wR}^2) - 1)$ is greater 0.3 calculate the widened acceptance limits (in the log domain) according to

$$[lABEL, uABEL] = \pm 0.76 * s_{wR}$$

  If CV_{wR} is ≤ 0.3 use $[-\log(1.25), \log(1.25)]$.
  If CV_{wR} is > 0.5 use the acceptance limits for $CV_{wR} = 0.5$ (cap on widening).
- Decide BE if the 90% confidence interval is contained in the scaled (widened) acceptance limits.

The covered replicate crossover designs have the following characteristics (N=Σn$_i$):

Design	df	$b_{k(ni)}$	b_k	df (RR)	E(mse)
2x3x3 (partial replicate)	2*N - 3	1/6	1.5	N-2	$(\sigma_{wT}^2 + 2 * \sigma_{wR}^2)/3$
2x2x4 (full replicate)	3*N - 4	1/4	1	N-2	$(\sigma_{wT}^2 + \sigma_{wR}^2)/2$

b_k is the design constant assuming $n_i = N/seqs$.
E(mse) is the expectation of the mean squared error from a model without subject by treatment interaction composed from the intra-subject variabilities of Test and Reference, respectively.

Simulation implementation

Instead of simulating subject data we are simulating the needed statistics via their associated distributions. A first attempt (implemented in PowerTOST V1.1-00, V1.1-02)

- pe is normal distributed with mean=log(GMR) and sd=sqrt($E(mse) * C2$)
  GMR is the true (assumed) ratio for the population.
- s_d^2*df/(E(mse)*C2) is chi-squared distributed and simulated via
  $s_d^2 = E(mse) * C2 * rchi(nsims, df)/df$
- $s_{wR}^2 * df_{RR}/\sigma_{wR}^2$ is chi-squared distributed and simulated via
  $s_{wR}^2 = \sigma_{wR}^2 * rchi(nsims, df_{RR})/df_{RR}$

With the so simulated statistics the above described method for the BE decision is performed. The cases of BE=TRUE will counted and pBE = count(BE=TRUE)/nsims is calculated.

The above described simulation attempt proved as too naïve.
The agreement of the power so calculated with values obtained via the 'classical' way of simulating subject data was unsatisfactory. See Appendix.

If the simulations via subject data are correct the conclusion could only be that the simulation of mse and s_{wR}^2 via *independent* chi-square distributions is not appropriate. One consequence of this attempt is that studies are simulated in which s_{wT}^2 if calculated via the relations given in the Table above becomes negative.

To avoid this it was simulated as following:

- $s_{wR}^2 * df_{RR}/\sigma_{wR}^2$ is chi-squared distributed and simulated via $s_{wR}^2 = \sigma_{wR}^2 * rchi(nsims, df_{RR})/df_{RR}$
- $s_{wT}^2 * df_{TT}/\sigma_{wT}^2$ is chi-squared distributed and simulated via $s_{wT}^2 = \sigma_{wT}^2 * rchi(nsims, df_{TT})/df_{TT}$
- mse is calculated from the constituents s_{wR}^2 and s_{wT}^2 according to the relations given in the Table above and from that $s_d^2 = mse * C2$.

This approach however has the flaw that we are not able to give the df_{TT} in case of the 2x3x3 design. It was choosen equal to df_{RR}. So this approach is more or less empirical for the 2x3x3 design and only justified but the better numeric agreement of the power values compared to those obatined via subject data simulations.

<u>Open questions, understanding problems</u>:

1. Is there a better way to handle the simulations of mse and s_{wR}^2 via *dependent* chi-square distributions?

2. The E(mse) for the 2x3x3 (partial replicate) design was decided from subject data simulations. How can we derive this theoretically?

3. Is working with different variabilities within the EMA method reasonable at all? Or does the model used only allow equal variabilities?
   An indication for that is the observation that the EMA method and the FDA method via ISC lead to different expected variances of the mean of T-R :
   EMA: $sqrt(\frac{1}{3}(\sigma_{wT}^2 + 2 * \sigma_{wR}^2) * \frac{1}{6} * \sum\frac{1}{n_i})$ = $sqrt((\sigma_{wT}^2/2 + \sigma_{wR}^2) * \frac{1}{9} * \sum\frac{1}{n_i}$
   FDA: $sqrt((\sigma_{wT}^2 + \sigma_{wR}^2/2) * \sum\frac{1}{n_i} * \frac{1}{9})$

4. How can we incorporate a subject by treatment interaction in the E(mse)? Can we? The ANOVA model we have to use doesn't incorporate such a term.

FDA method (`power.RSABE()`)

Method description in a cook book manner:

- Calculate the intra-subject contrasts **T-R** (of the log-transformed PK metrics) and analyze them via an ANOVA(1) with sequence as soley effect. The intercept of this ANOVA gives the point estimator (pe) of μ_T - μ_R.
  The std error associated with the pe is
  $$s_d = sqrt(mse_1 * \frac{1}{seqs^2} * \sum \frac{1}{n_i})$$
  The associated degrees of freedom are df=N-seqs. The term $\frac{1}{seqs^2} * \sum \frac{1}{n_i}$ is named C3. In case of equal number of subjects in sequence groups n_i=N/seqs the term C3 reduces to 1/N.
- Calculate the intra-subject contrasts **R-R** (of the log-transformed PK metrics) and analyze them via an ANOVA(2) with sequence as soley effect. The intra-subject variance for the reference is $s_{wR}^2 = mse_2/2$. The associated degrees of freedom are also df(RR)=N-seqs.
- In case of the full replicate design (2x2x4) the previous step can be repeated for **T-T** to obtain s_{wT}^2. But this value isn't used further down. It's only a nice to have.
- If s_{wR} > 0.2935604 calculate the linearized reference scaled ABE criterion
  $$crit = pe^2 - s_d^2 - theta^2 * s_{wR}^2$$
  where theta = log(1.25)/0.25 = 0.8925742.
  Calculate a 95% upper confidence interval of this criterion via Howe[1] approximation according to
  $$E_m = pe^2 - s_d^2$$
  $$C_m = (abs(pe) + t_{(1-alpha),df} * s_d)^2$$
  $$E_s = theta^2 * s_{wR}^2$$
  $$C_s = \mathrm{E}_s * df_{RR}/Chi_{(1-alpha),dfRR}$$
  $$bound = E_m - E_s + sqrt((C_m - E_m)^2 + (C_s - E_s)^2)$$
  If the upper bound is lower than zero decide BE
- If s_{wR} ≤ 0.2935604 (CV_{wR}≤ 0.3) then perform ABE evaluation, i.e. calculate 90% confidence intervals and decide BE if these are contained in the acceptance range $[-\log(1.25, \log(1.25)]$. The FDA demands to use the Proc MIXED code for this evaluation, regardless of the design.

Simulation implementation

Instead of simulating via subject data we are simulating the needed statistics via their associated distributions:

- pe is normal distributed with mean=log(GMR) and sd=sqrt($E(mse_1) * C3$)
  GMR is the true (assumed) ratio for the population.
- s_d^2*df/(E(mse_1)*C3) is chi-squared distributed and simulated via
  $s_d^2 = E(mse_1) * C3 * rchi(nsims, df)/df$
- $s_{wR}^2 * df_{RR}/\sigma_{wR}^2$ is chi-squared distributed and simulated via
  $s_{wR}^2 = \sigma_{wR}^2 * rchi(nsims, df_{RR})/df_{RR}$

With the so simulated statistics the above described method for the BE decision is performed. The cases of BE=TRUE are counted and pBE = count(BE=TRUE)/nsims is calculated.

The expectation of the mse1 are taken from the literature about IBE as:

Design	E(mse_1)
2x3x3 (partial replicate) [2,3]	$\sigma_D^2 + \sigma_{wT}^2 + \sigma_{wR}^2/2$
2x2x4 (full replicate) [4, 5]	$\sigma_D^2 + (\sigma_{wT}^2 + \sigma_{wR}^2)/2$

The subject by formulation interaction term σ_D^2 is assumed to be zero. It is only present for eventually enhancement in future.

Open questions, understanding problems:

1. The ABE evaluation (90% CI's) in case of $s_{wR} \leq 0.2935604$ ($CV_{wR} \leq 0.3$) is done via the results from the ANOVA(1), i.e. we calculate the 90%CI with pe and s_d from that step. How does this affect the results? How could we test this?
   If there is a considerable effect, how can we then simulate the ABE decision?

2. The "unknown x", i.e. the term $-s_d^2$ in E_m (taken from the SAS code of the progesterone guidance[6]): Where did it came from? Have the two Laszlo's used it in their simulations? Their earlier papers do not contain this term.
   Mueller-Cohrs[7] notes that pe^2 is only approximately unbiased for $(\mu T-\mu R)^2$ and a user in the BEBA forum (http://forum.bebac.at/mix_entry.php?id=5943) gave the hint that it may be the bias correction is done by subtraction of s_d^2.

FDA method for NTID's (`power.NTIDFDA()`)

Method description[8] in a cook book manner, design 2x2x4:

- Calculate the intra-subject contrasts **T-R** (of the log-transformed PK metrics) and analyze them via an ANOVA(1) with sequence as soley effect. The intercept of this ANOVA gives the point estimator (pe) of μ_T - μ_R.
  The std error associated with the pe is
  $$s_d = sqrt(mse_1 * \frac{1}{4} * \sum \frac{1}{n_i})$$
  The associated degrees of freedom are df = N-2. The term $\frac{1}{4} * \sum \frac{1}{n_i}$ is named C3. In case of equal number of subjects in sequence groups n_i=N/2 the term C3 reduces to 1/N.
- Calculate the intra-subject contrasts **R-R** (of the log-transformed PK metrics) and analyze them via an ANOVA(2) with sequence as soley effect. The intra-subject variance for the reference is $s^2_{wR} = mse_2/2$. The associated degrees of freedom are dfRR = N-2.
- Repeated the previous step for **T-T** to obtain s^2_{wT}. This variance has the associated degrees of freedom dfTT = N-2 = dfRR.
- Calculate the linearized reference scaled ABE criterion according to
  $$crit = pe^2 - s_d^2 - theta^2 * s^2_{wR}$$
  where theta = -log(0.9)/0.1 = 1.053605157.
  Calculate a 95% upper confidence interval of this criterion via Howe[1] approximation according to
  $$E_m = pe^2 - s_d^2$$
  $$C_m = (abs(pe) + t_{(1-alpha),df} * s_d)^2$$
  $$E_s = theta^2 * s^2_{wR}$$
  $$C_s = \mathrm{E}_s * df_{RR}/Chi_{(1-alpha),dfRR}$$
  $$bound = E_m - E_s + sqrt((C_m - E_m)^2 + (C_s - E_s)^2)$$
- Decide BE if the upper confidence limit of the linearized reference scaled ABE criterion is ≤0 and if the conventional ABE test (90% CI of T versus R within ABE acceptance range) shows BE. The latter is similar to placing a cap at CV_{wR}=0.2142 ($s_{wR} = \log(1.25)/theta = 0.2117905$) on the widening of the implied acceptance limits.
  Additionally the ratio of s_{wT}/s_{wR} should be ≤2.5. This is tested by calculating an upper confidence interval of this ratio via
  $$UL = \frac{s_{wT}/s_{wR}}{\sqrt{F_{1-alpha/2,dfTT,dfRR}}} \leq 2.5$$
  where $F_{1-alpha/2,dfTT,dfRR}$ is the value of the F-distribution with ν_1=dfTT and ν_2=dfRR degrees of freedom that has probability 1-alpha/2 to its **right** (see [8]).
  In R: `Fval=qf(1-alpha/2,dfTT,dfRR,lower.tail=FALSE))`.
  Alpha is set =0.1.

Simulation implementation

Instead of simulating via subject data we are simulating the needed statistics via their associated distributions:

- pe is normal distributed with mean=log(GMR) and sd=sqrt($E(mse_1) * C3$)
  GMR is the true (assumed) ratio for the population.
- $s_d^2 * df/(E(mse_1) * C3)$ is chi-squared distributed and simulated via
  $s_d^2 = E(mse_1) * C3 * rchi(nsims, df)/df$
- $s_{wR}^2 * df_{RR}/\sigma_{wR}^2$ is chi-squared distributed and simulated via
  $s_{wR}^2 = \sigma_{wR}^2 * rchi(nsims, df_{RR})/df_{RR}$
- $s_{wT}^2 * df_{TT}/\sigma_{wT}^2$ is chi-squared distributed and simulated via
  $s_{wT}^2 = \sigma_{wT}^2 * rchi(nsims, df_{TT})/df_{TT}$

E(mse1) is taken as $\sigma_D^2 + (\sigma_{wT}^2 + \sigma_{wR}^2)/2$. See above and reference[3]. The subject by formulation interaction term σ_D^2 is assumed to be zero. It is only present for an eventually enhancement in future. The σ_{xY}^2 are the population values for the respective variances.

With the so simulated statistics the above described method for the BE decision is performed. The cases of BE=TRUE are counted (implies BE(ABE) =TRUE, BE(scABE)=TRUE and ratio $s_{wT}/s_{wR} \leq 2.5$). From the counts pBE = count(BE=TRUE)/nsims is calculated as 'empirical' power.

Open questions, understanding problems:

- The ABE evaluation (90% CI's) is done via the results from the ANOVA(1), i.e. we calculate the 90%CI with pe and s_d from that step. The FDA on the other hand recommends to do that by the Proc MIXED code for SAS. This is scarcely implementable in R.
  How does this affect the results? How could we test this?
  If there is a considerable effect, how can we then simulate the ABE decision?

Appendix: Preliminary results of simulations via subject data

EMA method, GMR=0.95, 1E+5 sims if not otherwise given

CVwT	CVwR	n	sims	pBE	power.scABEL V1.1-02corr	Diff.	power.scABEL V1.1-03	Diff.
Design 2x3x3								
0.2	0.2	12		0.7538	0.7519	0.0020	0.7538	0.0000
		24		0.9616	0.9620	0.0004	0.9616	0.0000
0.3	0.3	12		0.4050	0.3974	0.0076	0.4120	-0.0070
		12	1E+6	0.4067	0.3958	**0.0109**	0.4112	-0.0045
		24		0.7794	0.7716	0.0079	0.7815	**-0.0020**
		48		0.9630	0.9604	0.0026	0.9635	-0.0004
0.40898	0.40898	12		0.2814	0.2941	**-0.0127**	0.2822	-0.0009
		12	1E+6	0.2825	0.2937	**-0.0112**	0.2816	0.0009
		24		0.7389	0.7223	**0.0166**	0.7453	-0.0064
		48		0.9618	0.9548	0.0070	0.9619	-0.0001
0.5	0.5	12		0.1940	0.2209	**-0.0269**	0.1910	0.0031
		24		0.7050	0.6953	0.0097	0.7091	-0.0041
		48		0.9627	0.9581	0.0046	0.9634	-0.0007
0.3	0.5	12		0.3741	0.3502	**0.0238**	0.3500	**0.0240**
		24		0.8628	0.8035	**0.0593**	0.8206	**0.0422**
		48		0.9937	0.9811	**0.0126**	0.9856	0.0081
0.5	0.3	12		0.1440	0.1642	**-0.0202**	0.1514	-0.0073
		24		0.5175	0.5592	**-0.0416**	0.5662	**-0.0487**
		48		0.8283	0.8661	**-0.0378**	0.8688	**-0.0405**
Design 2x2x4								
0.2	0.2	12		0.9023	0.9018	0.0004	0.9014	0.0009
		24		0.9947	0.9949	-0.0002	0.9947	0.0000
0.3	0.3	12		0.6570	0.6452	**0.0118**	0.6625	-0.0055
		24		0.9135	0.9072	0.0063	0.9138	-0.0003
		48		0.9942	0.9941	0.0001	0.9946	-0.0004
0.40898	0.40898	12		0.5493	0.5344	**0.0149**	0.5606	**-0.0113**
		24		0.8885	0.8781	**0.0104**	0.8921	-0.0037
		48		0.9920	0.9907	0.0014	0.9928	-0.0007
0.5	0.5	12		0.4704	0.4670	0.0034	0.4776	-0.0072
		24		0.8788	0.8720	0.0069	0.8805	-0.0017
		48		0.9914	0.9904	0.0010	0.9921	-0.0007
0.3	0.5	12		0.6951	0.6773	**0.0178**	0.7078	**-0.0126**
		24		0.9604	0.9528	0.0076	0.9641	-0.0037
		48		0.9984	0.9985	-0.0001	0.9989	-0.0005
0.5	0.3	12		0.3029	0.2990	0.0039	0.3130	**-0.0101**
		24		0.6969	0.6963	0.0006	0.7004	-0.0035
		48		0.9336	0.9306	0.0030	0.9319	0.0017

Red: abs(diff)>0.002

Agreement not perfect but – except the calculations with CVwT ≠CVwR – to some degree satisfactory for me. The best what I could do in the moment.

FDA method, GMR=0.95, 1E+5 sims if not otherwise given

CVwT	CVwR	n	sims	pBE	power.RSABE	Diff
Design 2x3x3						
0.2	0.2	12		0.7106	0.7108	-0.0002
		24		0.9560	0.9561	-0.0001
0.3	0.3	12		0.4123	0.4132	-0.0009
		24		0.7980	0.7990	-0.0010
		48		0.9700	0.9691	0.0009
0.40898	0.40898	12		0.3808	0.3801	0.0006
		24		0.8089	0.8104	-0.0016
		48		0.9831	0.9827	0.0004
0.5	0.5	12		0.3795	0.3779	0.0017
		24		0.8132	0.8153	**-0.0020**
		48		0.9763	0.9765	-0.0003
0.3	0.5	12		0.6296	0.6289	0.0006
		24		0.9406	0.9416	-0.0009
		48		0.9962	0.9961	0.0001
Design 2x2x4						
0.2	0.2	12		0.8737	0.8744	0.0007
		24		0.9931	0.9933	-0.0002
0.3	0.3	12		0.6374	0.6321	0.0054
		12	1E6	0.6355	0.6348	0.0007
		24		0.9172	0.9165	0.0006
		48		0.9948	0.9948	0.0000
0.40898	0.40898	12		0.5933	0.5913	**0.0020**
		24		0.9234	0.9231	0.0003
		48		0.9968	0.9971	-0.0003
0.5	0.5	12		0.5912	0.5903	0.0009
		24		0.9238	0.9235	0.0003
		48		0.9935	0.9938	-0.0002
0.3	0.5	12		0.7491	0.7483	0.0008
		24		0.9709	0.9710	-0.0002
		48		0.9986	0.9990	-0.0004
0.5	0.3	12		0.3263	0.3264	-0.0002
		24		0.7264	0.7244	**0.0020**
		48		0.9457	0.9444	0.0014

Red: abs(diff)>0.002

Agreement totally satisfactory for me.

FDA method for NTID's, design 2x2x4, 1E+5 sims if not otherwise given

CVwT	CVwR	n	sims	pBE	power.NTIDFDA	Diff
GMR=0.95						
0.05	0.05	12		0.0564	0.0583	-0.0019
		24		0.0644	0.0633	0.0011
0.075	0.075	12		0.2492	0.2505	-0.0014
		24		0.4266	0.4283	-0.0017
		48		0.6738	0.6707	0.0030
0.1	0.1	12		0.4037	0.4029	0.0008
		24		0.6871	0.6865	0.0006
		48		0.9222	0.9198	0.0023
0.125	0.125	12		0.4982	0.4968	0.0014
		24		0.8134	0.8123	0.0012
		48		0.9774	0.9752	0.0021
0.15	0.15	12		0.5597	0.5568	0.0029
		24		0.8762	0.8749	0.0013
		48		0.9914	0.9911	0.0003
0.175	0.175	12		0.5954	0.5939	0.0014
		24				
		48				
0.125	0.175	12		0.7231	0.7209	0.0022
		24				
		48				
0.175	0.125	12		0.2861	0.2872	-0.0011
		24				
		48				

Red: abs(diff)>0.002

Agreement not perfect but satisfactory for me.

References

[1] Howe W.G.
"Approximate confidence limits on the mean of X+Y where X and Y are two tabled independent random variables ”
J. of the American Statistical Association 1974, 69(347): 789-794

[2] McNally R.J.
"Tests for Individual and Population Bioequivalence Using 3-Period Crossover Designs"
http://www.stat.colostate.edu/statresearch/stattechreports/Technical%20Reports/2002/02-7%20McNally.pdf

[3] Shein-Chung Chow, Jun Shao, and Hansheng Wang
"Individual Bioequivalence testing under 2×3 designs"
Statist. Med. 2002; 21:629–648

[4] McNally R.J, Iyer H., Mathew T.
“Tests for Individual and Population Bioequivalence Based on Generalized p-Values”
Stat Med. 2003 Jan 15;22(1):31-53.

[5] Hansheng Wang and Shein-Chung Chow
"On statistical power for average Bioequivalence testing under Replicated crossover designs"
J. of Biopharmaceutical Statistics Vol. 12, No. 3, pp. 295–309, 2002

[6] FDA "Draft Guidance on Progesterone"
Recommended Apr 2010; Revised Feb 2011
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM209294.pdf

[7] Mueller-Cohrs J.
"Analysis of a three-period two-treatment pharmacokinetic study to assess scaled average bioequivalence "
PHUSE 2009, Paper SP04
www.phusewiki.org/docs/2009%20PAPERS/SP04.pdf

[8] FDA "Draft Guidance on Warfarine Sodium"
Recommended Dec 2012
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201283.pdf