

Design and Analysis of Alcohol/Benzo Interaction Studies*

Daowen Zhang

NC State University
DEPARTMENT OF STATISTICS



zhang@stat.ncsu.edu

<http://www4.ncsu.edu/~dzhang2>

* Joint work with Zhaoling Meng, Hui Huan at Sanofi-Aventis

Outline

1. Motivation
2. Study Design: Cross-over and Parallel Design
3. Cross-over Design
 - 3.1. Multivariate Linear Mixed Models
 - 3.2. Non-inferiority Approach
 - 3.3. Superiority Approach
 - 3.4. Detecting Individual Interaction Effect
4. Parallel Design
5. Implementation of the Analysis
6. Discussion

1. Motivation

- In many phase I clinical trials, need to examine the effect on some important PD parameters of the combined use of the new drug and alcohol/benzo over new drug alone: **interaction**.
- # of PD parameters is large.
- How to design such studies?
- How to analyze the data?

2. Study Design: Cross-over and Parallel Design

- When wash-out of the drug effect is reasonable, prefer cross-over designs
- When wash-out of the drug is hard to achieve, use parallel designs

3. Cross-over Design

- 4 treatments:

Treatment A	compound placebo + alcohol placebo
Treatment B	compound + alcohol placebo
Treatment C	compound placebo + alcohol
Treatment D	compound + alcohol

- 4 sequences:

1. sequence 1: A – B – C – D
2. sequence 2: B – D – A – C
3. sequence 3: C – A – D – B
4. sequence 4: D – C – B – A

3.1. Multivariate Linear Mixed Models

- Data: $K \times 1$ vector Y for K PD parameters measured at each period with a period-specific baseline measurement.
- Re-orient Y so that large value of Y is not desirable.
- Multivariate linear mixed model for Y .
 1. Model for sequence 1:

$$Y_{ij1} = \mu_1 + \tau_j I(j \geq 2) + \beta_1 I(j = 2) + \beta_2 I(j = 4) + \beta_3 I(j = 6) + \beta_4 I(j = 8) + b_{i1} + \epsilon_{ij1}$$

2. Model for sequence 2:

$$Y_{ij2} = \mu_2 + \tau_j I(j \geq 2) + \beta_2 I(j = 2) + \beta_4 I(j = 4) \\ + \beta_1 I(j = 6) + \beta_3 I(j = 8) + b_{i2} + \epsilon_{ij2}$$

3. Model for sequence 3:

$$Y_{ij3} = \mu_3 + \tau_j I(j \geq 2) + \beta_3 I(j = 2) + \beta_1 I(j = 4) \\ + \beta_4 I(j = 6) + \beta_2 I(j = 8) + b_{i3} + \epsilon_{ij3}$$

4. Model for sequence 4:

$$Y_{ij4} = \mu_4 + \tau_j I(j \geq 2) + \beta_4 I(j = 2) + \beta_3 I(j = 4) \\ + \beta_2 I(j = 6) + \beta_1 I(j = 8) + b_{i4} + \epsilon_{ij4}$$

$$b_{il} \stackrel{\text{i.i.d}}{\sim} \text{N}(0, D), \quad \epsilon_{ijl} \stackrel{\text{i.i.d}}{\sim} \text{N}(0, \Sigma),$$

- D is unstructured and Σ is usually taken to be

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_K^2 \end{bmatrix},$$

where σ_k^2 is the pure measurement error variance, or

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots & \rho\sigma_1\sigma_K \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho\sigma_2\sigma_K \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma_K\sigma_1 & \rho\sigma_K\sigma_2 & \cdots & \sigma_K^2 \end{bmatrix}.$$

- $\gamma = \beta_4 - \beta_3$ characterizes the *worsening* effect of the combined use of alcohol and the drug over the alcohol alone.
- MLE of γ :

$$\hat{\gamma} = \frac{1}{4} \{ (\bar{Y}_{81} - \bar{Y}_{61}) + (\bar{Y}_{42} - \bar{Y}_{82}) + (\bar{Y}_{63} - \bar{Y}_{23}) + (\bar{Y}_{24} - \bar{Y}_{44}) \}$$

has the following distribution

$$\hat{\gamma} \sim N \left(\gamma, \frac{2\Sigma}{n} \right).$$

- Can be used for sample size calculation for a cross-over design.

3.2. Non-inferiority Approach

- Tolerable limits: b_1, b_2, \dots, b_K (agreed by regulatory agencies)
- $\gamma_k \leq b_k$: an undesirable worsening effect (if exists) for the k th PD endpoint can be tolerated.
- Null hypothesis we would like to reject

$$H_0 : \gamma_k > b_k \text{ for at least one } k \in \{1, 2, \dots, K\}.$$

- Alternative is

$$H_a : \gamma_k \leq b_k \text{ for all } k \in \{1, 2, \dots, K\}$$

- Reject $H_0 \Rightarrow$ No worsening interaction effect at any PD endpoint

Testing procedure using CI

- From the viewpoint of a regulatory agency, we don't want to falsely claim no worsening interaction effect when there is for at least one endpoint. Want to control type I error rate of any testing procedure.
- For a given $\alpha \in (0, 1)$, a $(1 - \alpha)$ confidence interval for γ_k

$$(-\infty, \hat{\gamma}_k + z_\alpha \text{SE}(\hat{\gamma}_k)] = (-\infty, \hat{\gamma}_k + z_\alpha \sigma_k \sqrt{2/n}]$$

- Suggest testing procedure:

$$\text{reject } H_0 \iff \hat{\gamma}_k + z_\alpha \sigma_k \sqrt{2/n} \leq b_k \quad \text{for all } k \in \{1, 2, \dots, K\}.$$

- Maximum type I error rate:

$$\max P[\hat{\gamma}_k + z_\alpha \sigma_k \sqrt{2/n} \leq b_k \quad \text{for all } k \in \{1, 2, \dots, K\} | H_0] = \alpha.$$

- Power of the testing procedure (evaluated at no interaction):

$$P[\hat{\gamma}_k + z_\alpha \sigma_k \sqrt{2/n} \leq b_k, k = 1, 2, \dots, K | \gamma_1 = \gamma_2 = \dots = \gamma_K = 0].$$

- For given type I error rate α , power $1 - \beta$ to claim no interaction when there is truly no worsening interaction effect, need to solve for the total sample size n from

$$P[\hat{\gamma}_k + z_\alpha \sigma_k \sqrt{2/n} \leq b_k, k = 1, 2, \dots, K | \gamma = 0] = 1 - \beta.$$

- For a general Σ , power calculation usually requires K -dimensional numerical integration, or Monte Carlo method can be used.
- When Σ has the same correlation ρ , the power calculation can be reduced to a one-dimensional numerical integration.

Table 1. Required sample size and achieved power for declaring no interaction between the compound and alcohol using non-inferiority testing procedure ($\alpha = 0.05, b_k/\sigma_k = 0.5$)

K	$\rho = 0$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$		$\rho = 0.9$	
	n	power	n	power	n	power	n	power	n	power
1	50	0.800	50	0.800	50	0.800	50	0.800	50	0.800
2	68	0.804	66	0.801	64	0.804	61	0.802	57	0.801
3	78	0.802	76	0.802	72	0.801	66	0.802	61	0.802
4	85	0.800	82	0.804	77	0.802	70	0.801	63	0.803
5	91	0.801	87	0.803	82	0.803	73	0.802	65	0.804
6	96	0.804	91	0.801	85	0.801	76	0.803	66	0.803
7	99	0.804	95	0.803	88	0.803	77	0.802	67	0.804

Table 2. Required sample size and achieved power for declaring no interaction between the compound and alcohol using non-inferiority testing procedure ($\alpha = 0.05, b_k/\sigma_k = 1$)

K	$\rho = 0$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$		$\rho = 0.9$	
	n	power	n	power	n	power	n	power	n	power
1	13	0.807	13	0.807	13	0.807	13	0.807	13	0.807
2	17	0.807	17	0.816	17	0.814	16	0.805	15	0.812
3	20	0.805	19	0.804	19	0.811	17	0.800	16	0.805
4	22	0.807	21	0.810	20	0.808	18	0.807	16	0.805
5	23	0.803	22	0.810	21	0.813	19	0.804	17	0.810
6	24	0.807	24	0.816	22	0.808	19	0.805	17	0.801
7	26	0.815	24	0.810	22	0.806	20	0.808	17	0.809

- **Remark:**

1. For given data, if we failed to reject H_0 , we may not be able to conclude that worsening interaction effects exist in some endpoints.
2. If Σ used for design is different from what is estimated, there might not be enough power to detect H_a .

- If we failed to reject H_0 , we may want to identify the endpoints on which the no interaction effect exists.
- Define p-value for the k th endpoint:

$$P_k = P[Z \leq (\hat{\gamma}_k - b_k)/\text{SE}(\hat{\gamma}_k)], \quad Z \sim N(0, 1)$$

- Sort P_1, P_2, \dots, P_K in an ascending order as $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(K)}$.
- If $P_{(K-1)} \leq \alpha/2$, conclude no interaction effect for the endpoint corresponding to $P_{(1)}, P_{(2)}, \dots, P_{(K-1)}$; Otherwise if $P_{(K-2)} \leq \alpha/3$, conclude no worsening interaction effect exists in endpoints corresponding to $P_{(1)}, P_{(2)}, \dots, P_{(K-2)}$, etc.

3.3. Superiority Approach

- Alternatively, we may consider testing

$$H_0 : \gamma = 0 \quad v.s. \quad H_a : \gamma \neq 0.$$

- A test statistic would be

$$T = \frac{n}{2} \hat{\gamma}^T \Sigma^{-1} \hat{\gamma}.$$

- Under H_0 ,

$$T \sim \chi_K^2.$$

- Reject H_0 at level α if $T \geq \chi_{K;\alpha}^2$.

- Under $H_a : \gamma \neq 0$, T is non-central χ_K^2 with non-centrality parameter

$$NC = \frac{n}{2} \gamma^T \Sigma^{-1} \gamma = \frac{n}{2} \delta^T R^{-1} \delta,$$

where R is the correlation matrix of Σ , $\delta = \gamma/\sigma$ is the effect size of the interaction effect.

- For given type I error rate α , power $1 - \beta$, denote by $\phi^2(\alpha, \beta, K)$ the non-centrality parameter of a χ^2 distribution with K degrees of freedom such that

$$P[\chi^2(K, \phi^2(\alpha, \beta, K)) > \chi_{\alpha, K}^2] = 1 - \beta.$$

- The test will have the desired power $1 - \beta$ if

$$NC = \frac{n}{2} \delta^T R^{-1} \delta \geq \phi^2(\alpha, \beta, K) \implies n \geq \frac{2\phi^2(\alpha, \beta, K)}{\delta^T R^{-1} \delta}.$$

- Need to specify R and δ for calculating sample size n .
- Suppose a_k is the pre-defined clinically important interaction effect.
- Would like to claim with power $1 - \beta$ that interaction exists in at least one endpoint when in fact it exists in $K^* (\leq K)$ endpoints.
- Effect size:

$$\Delta_k = a_k / \sigma_k, \quad k = 1, 2, \dots, K^*,$$

- If $R \approx I$, then

$$n = \frac{2\phi^2(\alpha, \gamma, K)}{\Delta_1^2 + \Delta_2^2 + \dots + \Delta_{K^*}^2}.$$

- In practice, K^* is unknown. A conservative way is to choose $K^* = 1$.

Table 3. Required sample size and achieved power for detecting interaction effects using superiority testing procedure ($\alpha = 0.05, a_k/\sigma_k = 0.5$)

K^*	$K = 3$		$K = 4$		$K = 5$		$K = 6$		$K = 7$	
	n	power	n	power	n	power	n	power	n	power
1	88	0.804	96	0.802	103	0.802	109	0.800	115	0.801
2	44	0.804	48	0.802	52	0.806	55	0.804	58	0.805
3	30	0.814	32	0.802	35	0.810	37	0.808	39	0.809
4			24	0.802	26	0.806	28	0.813	29	0.805
5					21	0.810	22	0.804	23	0.801
6							19	0.821	20	0.821
7									17	0.817

Table 4. Required sample size and achieved power for detecting interaction effects using superiority testing procedure ($\alpha = 0.05, a_k/\sigma_k = 1$)

K^*	$K = 3$		$K = 4$		$K = 5$		$K = 6$		$K = 7$	
	n	power	n	power	n	power	n	power	n	power
1	22	0.804	24	0.802	26	0.806	28	0.813	29	0.805
2	11	0.804	12	0.802	13	0.806	14	0.813	15	0.821
3	8	0.840	8	0.802	9	0.823	10	0.843	10	0.821
4			6	0.802	7	0.839	7	0.813	8	0.849
5					6	0.866	6	0.843	6	0.821
6							5	0.843	5	0.821
7									5	0.884

Some Remarks:

- Similar to the non-inferiority approach, clinically meaningful bounds a_k 's have to be specified.
- The type II error rate here is similar to the type I error rate for the non-inferiority approach. Therefore, the power has to be specified at a very high level to make this approach reasonable.
- Given data, the estimate of Σ may be very different from the one used in the design; the study then may not have enough power to detect interaction effect, even if it may exist;
- On the other hand, we may reject $H_0 : \gamma = 0$, but the interaction effect effects may not be worsening effects or may be within the tolerance limits.

3.4. Detecting individual (worsening) interaction effect using Hochberg's approach

- P_k : one-sided p-value from testing

$$H_k : \gamma_k = 0 \quad vs. \quad H_k^a : \gamma_k > 0.$$

1. Order $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(K)}$
2. If $P_{(K)} \leq \alpha$, then all H_k 's are rejected.
3. If $P_{(K)} > \alpha$, then do not reject $H_{(K)}$ and compare $P_{(K-1)}$ to $\alpha/2$.
4. If $P_{(K-1)} \leq \alpha/2$, then $H_{(1)}, H_{(2)}, \dots, H_{(K-1)}$ are rejected.
5. If $P_{(K-1)} > \alpha/2$, then do not reject $H_{(K-1)}$ and compare $P_{(K-2)}$ to $\alpha/3$.
6. If $P_{(K-2)} \leq \alpha/3$, then $H_{(1)}, H_{(2)}, \dots, H_{(K-2)}$ are rejected.
7. The procedure keeps iterating.

Table 5. Required sample size and achieved power for identifying at least one out of K^* endpoints in which an interaction effect exists using Hochberg's approach ($K = 7$)

K^*	$\rho = 0$				$\rho = 0.25$				$\rho = 0.5$			
	$a_k/\sigma_k = 0.5$		$a_k/\sigma_k = 1$		$a_k/\sigma_k = 0.5$		$a_k/\sigma_k = 1$		$a_k/\sigma_k = 0.5$		$a_k/\sigma_k = 1$	
	n	power	n	power	n	power	n	power	n	power	n	power
1	101	0.802	26	0.805	101	0.802	26	0.805	100	0.801	26	0.803
2	64	0.802	17	0.816	70	0.805	18	0.811	76	0.801	19	0.802
3	50	0.805	13	0.814	57	0.803	15	0.816	65	0.802	17	0.809
4	41	0.803	11	0.801	49	0.805	13	0.816	59	0.805	15	0.804
5	36	0.803	9	0.808	44	0.805	12	0.822	54	0.803	14	0.817
6	32	0.808	9	0.831	40	0.802	11	0.810	51	0.806	13	0.805
7	29	0.808	7	0.803	37	0.803	10	0.815	48	0.803	12	0.808

4. Parallel Design

- 2 groups:
 1. group 1: placebo + alcohol
 2. group 2: compound + alcohol
- Model for the data:

group1: placebo + alcohol

$$Y_{ij1} = \beta_0 + \beta_1 I(j \geq 2) + \tau_j I(j \geq 2) + b_{i1} + \epsilon_{ij1}$$

group 2: compound + alcohol

$$Y_{ij2} = \tilde{\beta}_0 + \beta_2 I(j \geq 2) + \tau_j I(j \geq 2) + b_{i2} + \epsilon_{ij2},$$

$$b_{i1}, b_{i2} \stackrel{\text{i.i.d}}{\sim} N(0, D), \quad \epsilon_{ij1}, \epsilon_{ij2} \stackrel{\text{i.i.d}}{\sim} N(0, \Sigma).$$

- $\gamma = \beta_2 - \beta_1$ is the interaction effect of the combined use of the compound and alcohol over alcohol alone.
- Without assuming $\beta_0 = \tilde{\beta}_0$,

$$\hat{\gamma} = \frac{2}{n} \left\{ \left(\frac{1}{m} \sum_{j=2}^{m+1} Y_{+j2} - Y_{+02} \right) - \left(\frac{1}{m} \sum_{j=2}^{m+1} Y_{+j1} - Y_{+01} \right) \right\},$$

where $Y_{+j\ell} = \sum_{i=1}^{n/2} Y_{ij\ell}$ for $j = 1, 2, \dots, m + 1$ and $\ell = 1, 2$.

- $\hat{\gamma}$ has the following distribution

$$\hat{\gamma} \sim N \left(\gamma, \frac{4(m+1)\Sigma}{nm} \right)$$

- Can be used for sample size calculation similar to the cross-over design.

5. Implementation of the Analysis

Cross-over design:

- Data: Simulated data set cross contains data from 40 subjects
 - ★ 3 PD variables: y_1 , y_2 , y_3 ; coded by a common variable y , identified by endpt (1, 2, 3)
 - ★ 8 time points: time (1, 2, ..., 8)
 - ★ 4 sequences: seq1 , seq2 , seq3 , seq4
 - ★ 4 treatment dummies: trt1 , trt2 , trt3 , trt4
 - ★ Patients' ID: id
- Interested in comparing treatments 3 and 4: $\gamma = \beta_4 - \beta_3$.
- Define $\text{trt34} = \text{trt3} + \text{trt4}$
- Read estimates of trt4

SAS program for the multivariate mixed model:

```
data y1; set cross;
  y = y1;
  endpt = 1;
  drop y1;
run;
```

```
data y2; set cross;
  y = y2;
  endpt = 2;
  drop y2;
run;
```

```
data y3; set cross;
  y = y3;
  endpt = 3;
  drop y3;
run;
```

```
data main; set y1 y2 y3;
  z1 = (endpt=1);
  z2 = (endpt=2);
  z3 = (endpt=3);
  trt34 = trt3+trt4;
run;

proc sort data=main;
  by id time;
run;

title "Comparison between treatment 4 and treatment 3";
proc mixed data=main;
  class id endpt time;
  model y = endpt*seq1 endpt*seq2 endpt*seq3 endpt*seq4
          endpt*time3 endpt*time4 endpt*time5
          endpt*time6 endpt*time7 endpt*time8
          endpt*trt1 endpt*trt2 endpt*trt34 endpt*trt4
          / noint s ddfm=satterth;
  random z1-z3 / subject=id type=un g;
  repeated / subject=id(time) type=un r;
run;
```

Part of the output:

Comparison between treatment 4 and treatment 3

Estimated R Matrix for id(time) 1 1

Row	Col1	Col2	Col3
1	1.1674	0.2761	0.2251
2	0.2761	1.3301	0.3592
3	0.2251	0.3592	1.2436

Estimated G Matrix

Row	Effect	id	Col1	Col2	Col3
1	z1	1	1.6568	0.6957	0.8264
2	z2	1	0.6957	1.8754	0.9758
3	z3	1	0.8264	0.9758	1.9928

Solution for Fixed Effects

Effect	endpt	Estimate	Standard Error	DF	t Value
trt1*endpt	1	0.6077	0.2833	270	2.15
trt1*endpt	2	1.0913	0.3024	270	3.61
trt1*endpt	3	0.9522	0.2924	270	3.26
trt2*endpt	1	2.0664	0.2833	270	7.29
trt2*endpt	2	2.3243	0.3024	270	7.69
trt2*endpt	3	2.3458	0.2924	270	8.02
trt34*endpt	1	1.9294	0.2833	270	6.81
trt34*endpt	2	1.2666	0.3024	270	4.19
trt34*endpt	3	2.0609	0.2924	270	7.05
trt4*endpt	1	-0.2322	0.2416	270	-0.96
trt4*endpt	2	0.03800	0.2579	270	0.15
trt4*endpt	3	0.6242	0.2494	270	2.50

- Assume tolerance limits for 3 endpoints: 0.5, 0.5 and 1 and set $\alpha = 0.05$.
- Upper bounds $\hat{\gamma} + z_{\alpha}\text{SE}(\hat{\gamma})$ are 0.165, 0.462 and 1.034
 \Rightarrow
Cannot reject H_0 : in favor of H_a : No worsening interaction effect in all three endpoints.
- May want to identify endpoints on which no worsening interaction effect exists.

- P-values for 3 endpoints:

endpoint 1: $P_1 = P[Z \leq (-0.2322 - 0.5)/0.2416] = 0.0012$

endpoint 2: $P_2 = P[Z \leq (0.038 - 0.5)/0.2579] = 0.0366$

endpoint 3: $P_3 = P[Z \leq (0.6242 - 1)/0.2494] = 0.0659$

★ $P_3 > \alpha \Rightarrow$ cannot conclude no worsening effect for endpoint 3.

★ $P_2 > \alpha/2 = 0.025 \Rightarrow$ cannot conclude no worsening effect for endpoint 2.

★ $P_1 < \alpha/3 = 0.0167 \Rightarrow$ no worsening effect for endpoint 1.

6. Discussion

- Present cross-over and parallel designs to investigate alcohol/benzo and drug interaction.
- Challenge: large number of PD endpoints.
- Multivariate linear mixed model for the data.
- Consider several approach for testing the interaction; Non-inferiority approach may be favored by regulatory agencies; Need to specify the tolerance limits that are agreed by regulatory agencies and industry.
- Analysis has to be consistent with the design; Multiple adjustment?