

# クロスオーバー試験の計画および解析

富山大学大学院医学薬学研究部 バイオ統計学・臨床疫学教室  
折笠 秀樹

## はじめに

ランダム化比較試験 (RCT と略す) は, パラレル比較試験とクロスオーバー (crossover または cross-over) 試験に分かれる。ほとんどの RCT はパラレル比較試験であるが, ときにクロスオーバー試験を行うことがある。とくに, 生物学的同等性試験や薬物動態試験で使われることが多い。和訳では交差 (あるいは交叉) 試験とよぶ。

このクロスオーバー試験は古くから臨床応用されてきたようだが, Simpson は 1938 年にさまざまなクロスオーバーデザインを示した<sup>1)</sup>。著名な統計学者である Cochran らは牧畜業の分野で適用例を示した<sup>2)</sup>。Fieller も bioassay の分野で 2×2 デザインを述べた<sup>3)</sup>。当初は change-over デザインとよぶほうが多かった。このクロスオーバーデザインは Williams<sup>4)</sup> や Finney<sup>5)</sup> によって進展した。統計学的には, split-plot あるいは repeated-measures デザインを単純化し

た例といえる。平易な解説<sup>6)</sup>および成書<sup>7~9)</sup>も参照されたい。

## 数値データの 2×2 デザイン

クロスオーバーデザインではパラレルデザインと同様に, まず無作為に 2 群へ振り分ける。1 群 (Sequence 1) は第 I 期 (Period I) に治療 A を受け, 第 II 期 (Period II) は治療 B を受ける (図 1)。2 群 (Sequence 2) はその逆であり, 第 I 期に治療 B を受け, 第 II 期は治療 A を受ける。群のことは group とよぶこともあるが, sequence とよぶほうが多いだろう。Sequence を順 (序) と訳しても, 1 順というのはなじまないのので, 本稿では 1 群あるいは Sequence 1 とよぶ。2 群 2 期のクロスオーバーデザインのことを, 2×2 デザインまたは AB/BA デザインなどとよぶ。

2×2 デザインを定式化したのが図 2 である。全体

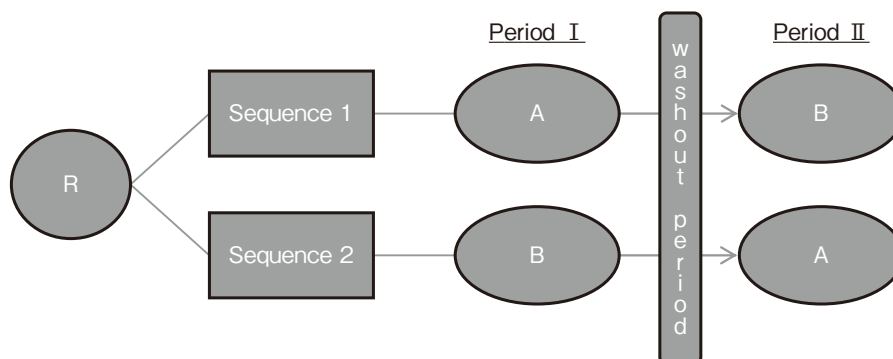


図 1 2×2 (AB/BA) クロスオーバー試験デザイン

R はランダム割付の略である。ランダム割付すれば, 二つの sequence のあいだで被験者背景は類似する。Sequence 1 では第 I 期 (Period I) に治療 A, 第 II 期 (Period II) に治療 B を施す。Sequence 2 ではその逆になる。Sequence のことを Group とよぶこともあり, ランダム割付された 2 sequences のことは 2 順 (序) とはいわず, 2 群というのが通常である。第 I 期の影響を取り除くために, 途中, ウォッシュアウト期間 (washout period) を設けることが多い。

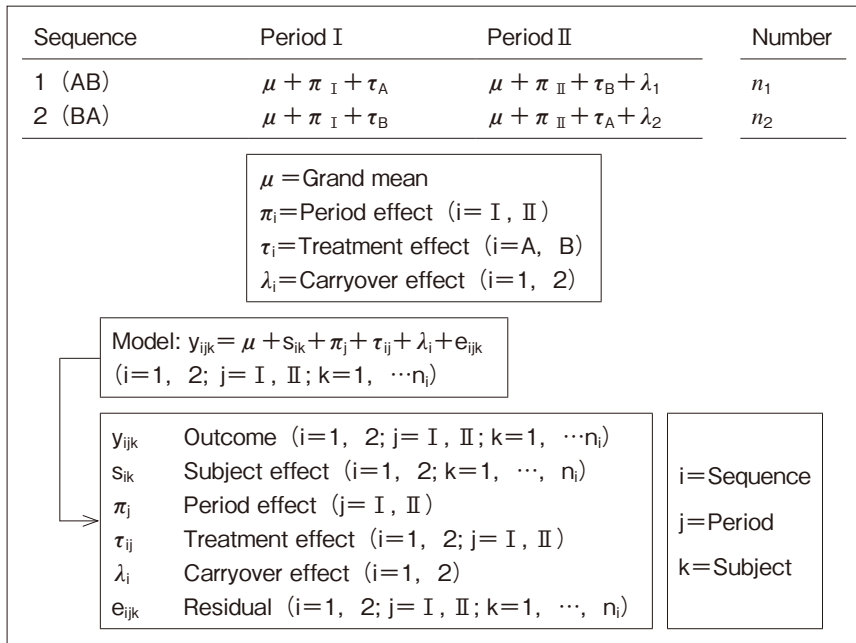


図 2 2群2期のクロスオーバーデザイン (AB/BA デザイン) の定式化

治療効果 ( $\tau_{ij}$ ) については, sequence・period が決まれば治療 A または B が決まる。持ち越し効果は時期 I の効果が持ち越されるので, 群 1, 2 ごとに定義される。仮に A が実薬で B がプラセボだとすると, プラセボには効果がないので  $\lambda_2=0$  と想定はされるが, 実際にはプラセボにも効果を示す人がいるため必ずしもそうはならない。 $\lambda_1=\lambda_2$ であれば carryover effect はないことを意味する。

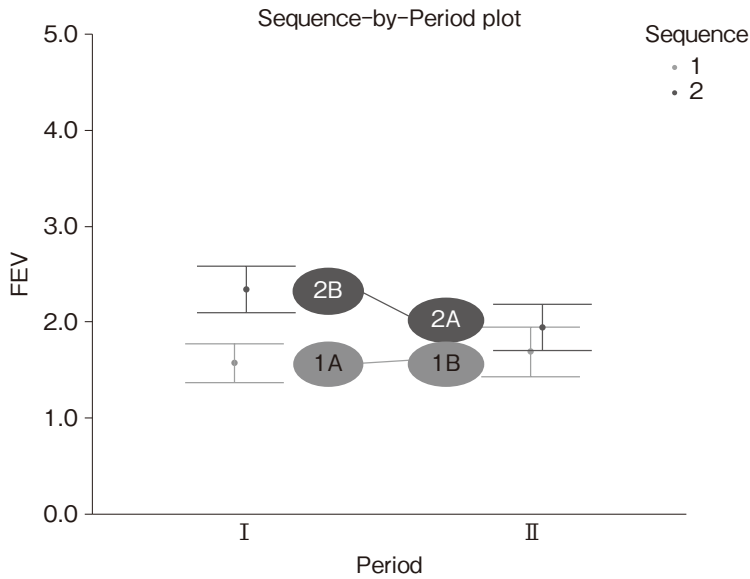
表 1 喘息患者を対象としたクロスオーバー試験における FEV<sub>1</sub> 値データ<sup>11)</sup>

Sequence	Subject	Period I	Period II
1 (AB)	1	1.28	1.33
1	2	1.6	2.21
1	3	2.46	2.43
1	4	1.41	1.81
1	5	1.4	0.85
1	6	1.12	1.2
1	7	0.9	0.9
1	8	2.41	2.79
2 (BA)	9	3.06	1.38
2	10	2.68	2.1
2	11	2.6	2.32
2	12	1.48	1.3
2	13	2.08	2.34
2	14	2.72	2.48
2	15	1.94	1.11
2	16	3.35	3.23
2	17	1.16	1.25

群 (Sequence) 1 では Period I に A, Period II に B が投与された。群 (Sequence) 2 ではその反対で, Period I に B, Period II に A が投与された。FEV<sub>1</sub> 値とはスパイロメトリー検査で最初の 1 秒間の努力呼気量 (単位 L) であり, 低いほど呼吸器機能が衰えていることを示す。群 1 には 8 名, 群 2 には 9 名が割り付けられた。

平均 (grand mean), 時期効果 (period effect, P のギリシャ文字  $\pi$  を使う), 治療効果 (treatment effect, T のギリシャ文字  $\tau$  を使う), 持ち越し効果 (carryover effect), そして被験者効果 (subject effect) という専門用語を用いる。持ち越し効果とは, 第 I 期の効果が第 II 期まで持ち越されることを表す。これを防ぐために, 図 1 に示したウォッシュアウト期間 (washout period, 休薬期間とよぶこともある) という期間を設けることが多い。米国 FDA 生物学的同等性試験ガイドラインでは, 消失半減期  $t_{1/2}$  の 5 倍以上設けることが勧められている<sup>10)</sup>。

喘息患者を対象にクロスオーバー試験が実施された<sup>11)</sup>。そのときのアウトカムは FEV<sub>1</sub> (スパイロメトリー検査で最初の 1 秒間の努力呼気量, 単位 L) であった。全部で 17 名の被験者に対するデータを時期ごとに示した (表 1)。このデータを統計解析するわけだが, 最初から分散分析に取り掛かるのはよろしくない。図 3 のような sequence-by-period プロットを描いてみよう。Treatment-by-period 交互作用, そして sequence effect がうかがわれる。このことから持ち越し効果が存在するかもしれないので, 後で正式な検討が望まれる。本題の治療効果については Period I では B のほうが優れるものの, Period II では多少だが A のほうが優れるようなので, 治療効果は正式な解析で検討してみないとわからない。



誤差バーは、平均から1標準誤差の範囲を表す

図 3 Sequence-by-period plot による検討

Period I では治療 B が高く、Period II では逆のため、treatment-by-period 交互作用がうかがわれる。Treatment-by-period 交互作用には carryover effect も含まれるので、carryover effect については心配しておいたほうがよいだろう。Sequence 1 のほうが 2 よりも低めなことからもそれがいえる。治療効果については、Period I では治療 B がかなり優れるが、Period II では逆に治療 A のほうが少しだけ優れる。全体でどうなのかは検定してみないとわからない。

表 2 Carryover, treatment, period effects を t 検定するために改変したデータ構造

Subject	Sequence	( I + II ) Carryover effect	( I - II )/2 Treatment effect	( I - II )/2 or ( II - I )/2 Period effect
1	1	2.61	-0.025	-0.025
2	1	3.81	-0.305	-0.305
3	1	4.89	0.015	0.015
4	1	3.22	-0.2	-0.2
5	1	2.25	0.275	0.275
6	1	2.32	-0.04	-0.04
7	1	1.8	0	0
8	1	5.2	-0.19	-0.19
9	2	4.44	0.84	-0.84
10	2	4.78	0.29	-0.29
11	2	4.92	0.14	-0.14
12	2	2.78	0.09	-0.09
13	2	4.42	-0.13	0.13
14	2	5.2	0.12	-0.12
15	2	3.05	0.415	-0.415
16	2	6.58	0.06	-0.06
17	2	2.41	-0.045	0.045

Period ( I + II ) で carryover ( or sequence ) effect が検証できる。Period ( I - II )/2 で treatment effect が検証できる。Sequence 1 では Period ( I - II )/2, Sequence 2 では Period ( II - I )/2 で period effect が検証できる。

このような見通しをつけるのが目的である。これ以外にも、sequence-by-treatment plot などを描いてみるとよいだろう。

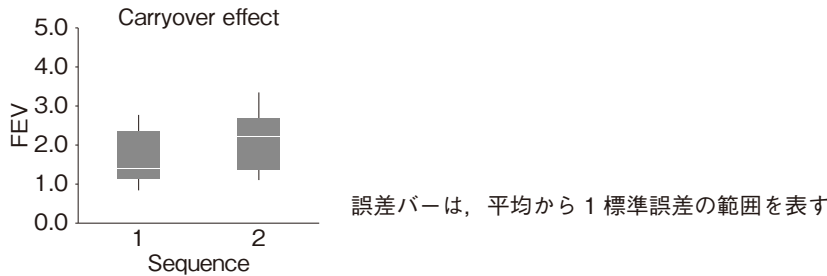
### 単純な t 検定による解析

持ち越し効果、治療効果、時期効果、それぞれについて単純な t 検定を実施してみよう。この単純な t 検定の結果は、後で示す分散分析の結果と実は同

$t_{1k} = y_{1Ik} + y_{1Iik}$  (Sequence 1 のなかでの Period I と II の和,  $k=1, \dots, n_1$ )  
 $t_{2k} = y_{2Ik} + y_{2Iik}$  (Sequence 2 のなかでの Period I と II の和,  $k=1, \dots, n_2$ )  
 $\bar{t}_1 = \sum t_{1k}/n_1$  (Sequence 1 の平均値),  $\bar{t}_2 = \sum t_{2k}/n_2$  (Sequence 2 の平均値)

Sequence 1 (AB)	$\mu + \pi_I + \tau_A$	$\mu + \pi_{II} + \tau_B + \lambda_1$
Sequence 2 (BA)	$\mu + \pi_I + \tau_B$	$\mu + \pi_{II} + \tau_A + \lambda_2$

→ Carryover effect  $\Leftrightarrow \lambda_1 - \lambda_2 = \hat{\lambda}_d = \bar{t}_1 - \bar{t}_2$ .



JMP による解析

Y: FEV, Period I + II for both groups

(Within-subject で時期和をとれば、持ち越し効果の検定が可能)

X: Sequence

差	-1.0242	t 値	-1.62347
差の標準誤差	0.6309	自由度	15
	P 値 (Prob >  t )		0.1253

→  $\hat{\lambda}_d = -1.0242$  ( $P=0.1253$ ) → Sequence 1 のほうが平均で 1.024 L 低い → B から A への持ち越しが大きい

#### 図 4 Carryover effect の検定

Carryover effect は、sequence 内での時期和に関する平均値の比較 ( $t$  検定) で検証できる。ランダム割付けにより両群は類似するが、両群で平均値が異なるのは carryover effect の影響と考えられる。そこで、sequence effect とよんでもよい。本検定では有意水準を 5% ではなく、10% にすることが勧められている。10% 水準でも carryover effect は非有意であり、これが好ましい結果である。

じになることに注意しよう。単純な  $t$  検定をするために、表 1 のデータを表 2 のような様式へ改変する。

Period I と II の和の平均値を sequence 間で比較すると、それは持ち越し効果 ( $\lambda_1 - \lambda_2$ ) の推定値となる。つまり、持ち越し効果が検討できることになる。JMP™ で解析した結果を図 4 に示した。P 値は 0.1253 であり非有意なので、持ち越し効果はないとして今後の解析を行ってもよいだろう。なお、持ち越し効果の有意水準は 5% ではなく、検出力が低いので 10% で行うのがよい。

次に治療効果であるが、ここでは Period I と II の差の平均値を sequence 間で比較すればよい。ここでは持ち越し効果はないと仮定している。時期効果は

存在することを前提にしている。JMP™ で解析した結果は図 5 に示したように、 $P=0.0472$  で治療法間には統計学的に有意な差が認められた。治療効果 (A-B) の推定値は  $-0.257$  L なので、治療 B のほうが FEV<sub>1</sub> 値は高いことがわかる。

時期効果については、Sequence 1 では時期 1 から 2 の差、Sequence 2 では時期 II から I の差をとる。そして、平均値を sequence 間で比較すればよい。ここでも持ち越し効果はないと仮定している。治療効果は存在することを前提にしている。JMP™ で解析した結果は図 6 に示したように、時期効果は認められなかった。時期効果 (I - II) の推定値は 0.139 L なので、時期 I のほうが FEV<sub>1</sub> 値はやや高いことが

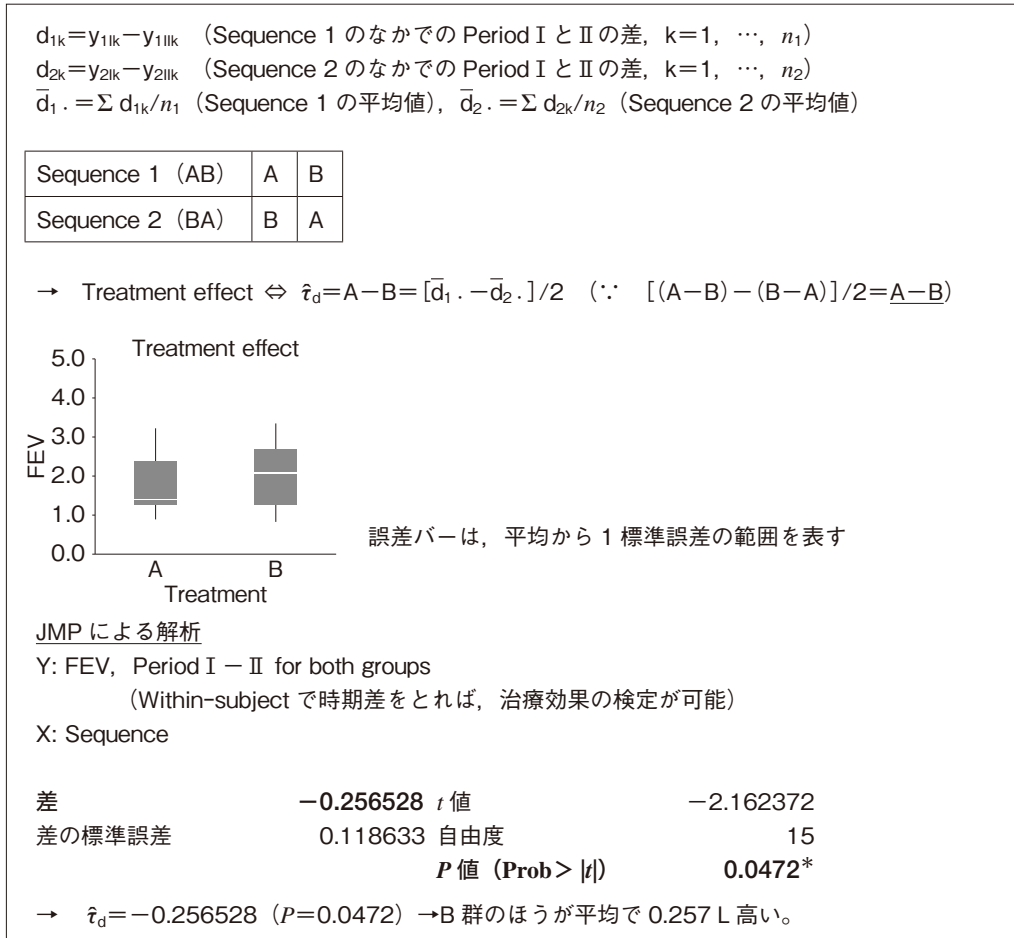


図 5 Treatment effect の検定

Treatment effect は、sequence 内での時期差に関する平均値の比較 (t 検定) で検証できる。Carryover effect はないと仮定している。Period effect は考慮したうえでの解析である。実際には unpaired t-test に帰着するが、基本は第 I 期と第 II 期の差に基づいているので、paired analysis の一種と考えられる。A 群のほうが 0.257 L だけ低く、統計学的には有意な差である。

わかる。

クロスオーバーデザインでは被験者内 (within subject) の差に注目して、図 7 のような対応のあるドットプロット (subject profiles plot ともよぶ) を示すのがよい。Sequence 1 と 2 とともに、治療 B のほうで FEV<sub>1</sub> 値は高いことがわかる。図 8 では pairwise difference (A-B) データに関して、paired t-test を実施した。治療効果の推定値は -0.265 であり、図 5 に示した解析結果と少しだけ異なっている。図 8 の解析では時期効果で調整していないが、図 5 の解析では時期効果で調整している違いである。t 検定のノンパラメトリック版である Wilcoxon 検定をすることも可能である。図 4~6 では t 検定のみ示した

が、そちらも Wilcoxon 検定が可能である。しかしながら、EMA 生物学的同等性ガイドラインでは、ノンパラメトリック検定では信頼区間が出せないのので、t 検定や次に述べる分散分析のほうを勧めている<sup>12)</sup>。

### 分散分析

それぞれの効果の一つずつ検定する手法は先に示したとおりであるが、それをいちどに実施するモデルが repeated measures ANOVA (反復測定分散分析) という手法である。これを実施するためには、表 3 のデータ構造へまず改変しておく必要がある。

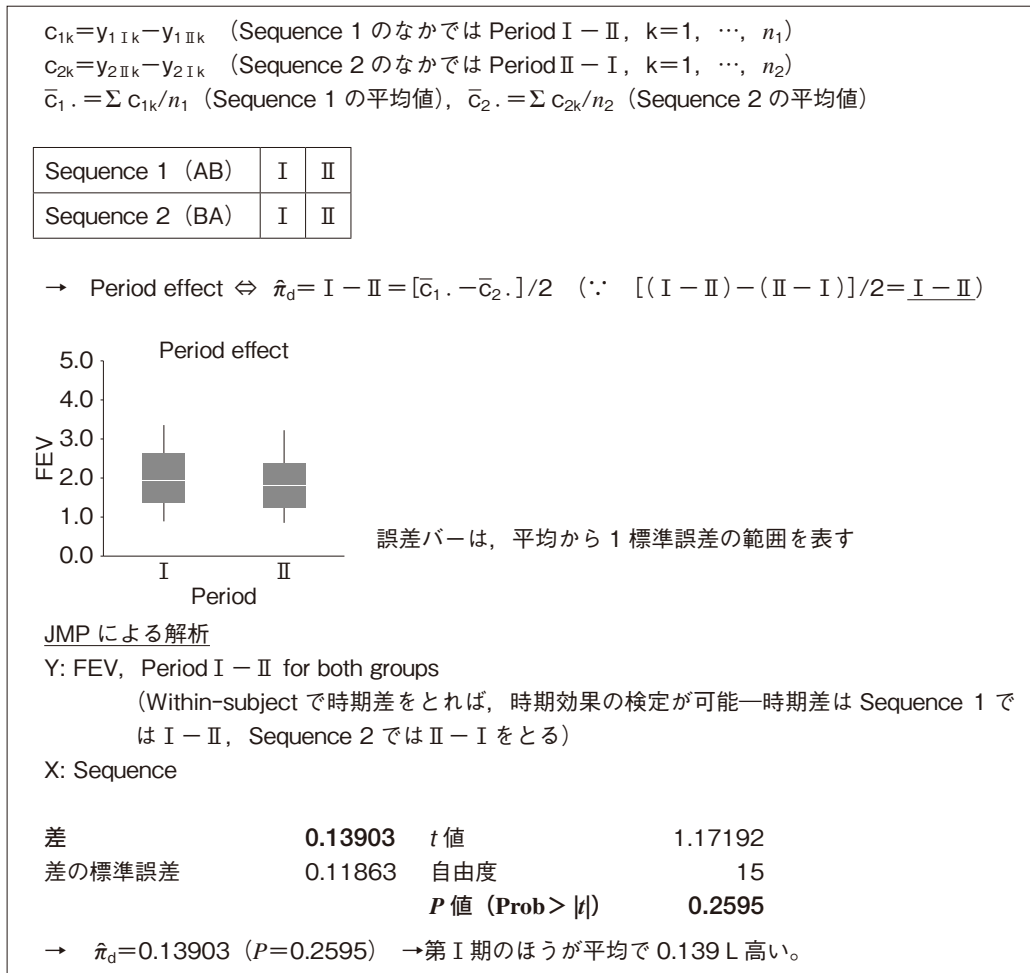


図 6 Period effect の検定

Period effect は、Sequence 1 では時期差 (I - II) の平均値、Sequence 2 では時期差 (II - I) の平均値の比較 (t 検定) で検証できる。Carryover effect はないと仮定している。Treatment effect は考慮したうえでの解析である。第 I 期のほうが 0.139 L 高いが、これは統計学的には非有意である。時期効果は非有意であることが望ましい。

図 9 のように、2×2 デザインでは平均値が全部で四つ得られる。全体平均があるので自由度の合計は 3 だが、sequence, treatment, period の三つに相当する。被験者間 (between subjects) 変動の解析により sequence effect, すなわち carryover effect が検定できる。被験者内 (within subject) 変動の解析により、治療効果と時期効果が検定できる。これを JMP™ で解析した結果を図 10 に示した。単純な t 検定での結果 (図 4~6) とすべて同じであることに注目されたい。JMP では平方和などは出力されないが、それは論文執筆などではあまり必要のない情報量である。なお、SAS™ を用いると平方和なども出力される。

図 10 には分散分析による検定結果を示したが、推定結果は別のモデルを使って得られる (図 11)。このような分散分析表での解析を提唱したのが Grizzle<sup>13)</sup> であり、後に Grieve<sup>14)</sup> が多少の訂正を加えた。

### 2×2 デザイン以外のデザイン

図 12 に、4×2 デザイン (Balaam デザイン<sup>15)</sup> とよばれる) を示した。4 群かつ 2 期のクロスオーバーデザインである。ここでは八つの平均値があるが、全体平均で自由度 1, 群効果で 3, 治療効果で 1, 時期効果で 1, 持ち越し効果で 1, 治療・持ち越し間



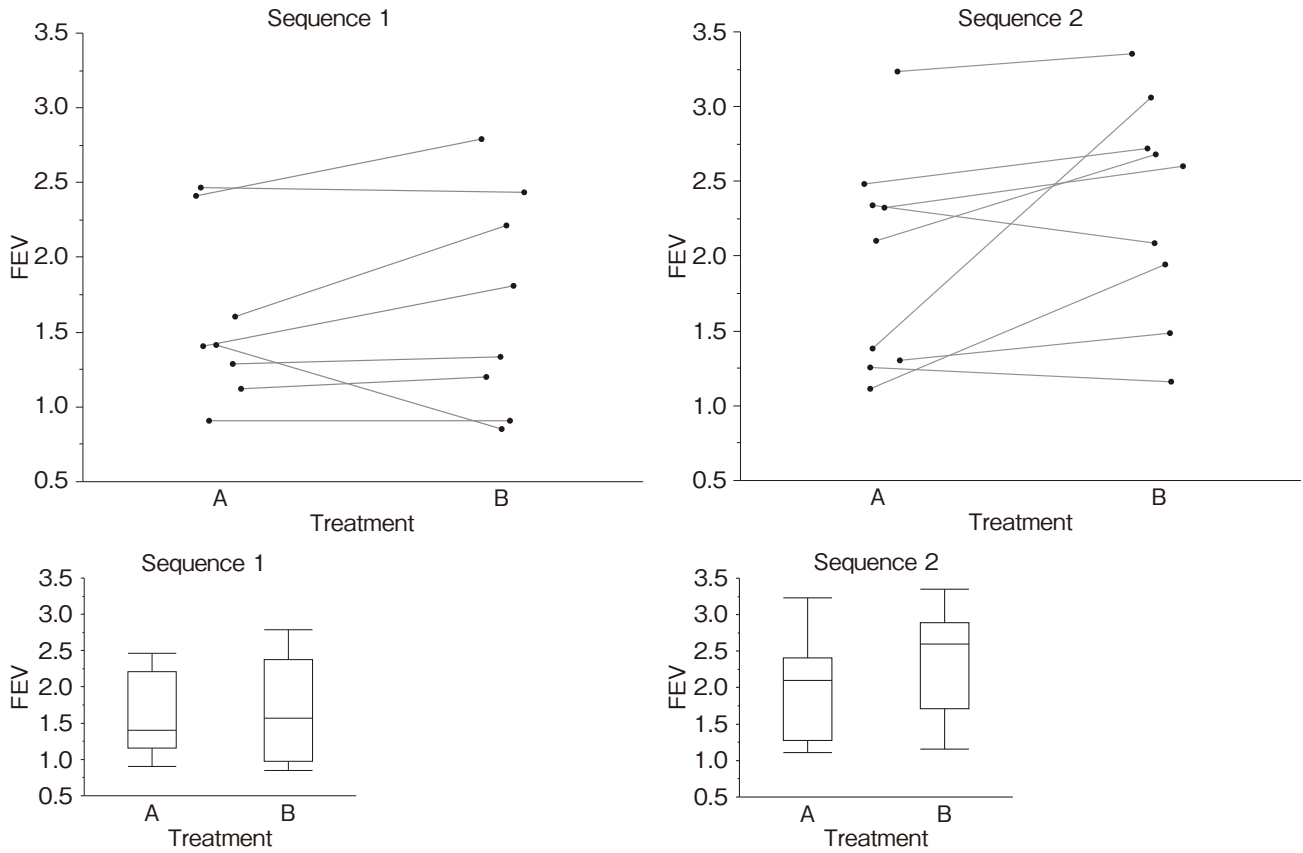


図 7 Subject profiles plot による pairwise 比較

Sequence (群) ごとに、pairwise (within subject) に二つの治療の結果を対比した。どちらの群でも治療 B のほうが FEV<sub>1</sub> 値は高そうに見えるが、その差が統計的に有意であるかは図 5 に示したように時期差を取って unpaired *t*-test (時期差に基づいているので paired analysis といえる) を実施することも可能であるし、図 8 に示すような paired *t*-test (これも paired analysis) も可能である。

交互作用で 1 に相当する。このデザインでは、持ち越し効果を推定することができる。図 13 は 2×3 デザインであるが、こちらでも持ち越し効果を推定できる。このほかにもいろいろな高次 (higher-order) クロスオーバーデザインがある<sup>16)</sup>。三つ以上の治療法を含む 2 期からなるクロスオーバーデザインも提案されており<sup>17)</sup>、ラテン方格法とよばれる実験計画法が使われることもある。

### 推奨する報告様式

クロスオーバー試験の結果をどのように報告すればよいか。それは、今後メタアナリシスを行ううえでも大変重要なポイントである<sup>18,19)</sup>。図 14 にその例を示した。群・時期ごとに平均 (標準偏差) を示し、

被験者内治療効果の平均 (標準偏差) を示す。全体の治療効果については、最小 2 乗平均値・95%信頼区間・*P* 値を提示する。最小 2 乗平均値 (LS mean) と単純平均値 (mean) は異なり、前者はモデルを用いて他の因子を調整した平均値のことである。アウトカムが複数ありスペースを割けない場合は、図 14 の下に示したような要約表でもよいと思う。

喘息患者のデータで例示したのが図 15 である。A then B は群 1 の結果であり、B then A は群 2 の結果である。要約指標は記述統計を用いて、あらためて計算する必要があるかもしれない。Treatment effect の結果については、図 10, 11 から引用すればよい。

Subject	Sequence	Pairwise difference* (A-B)
1	1	-0.05
2	1	-0.61
3	1	0.03
4	1	-0.4
5	1	0.55
6	1	-0.08
7	1	0
8	1	-0.38
9	2	-1.68
10	2	-0.58
11	2	-0.28
12	2	-0.18
13	2	0.26
14	2	-0.24
15	2	-0.83
16	2	-0.12
17	2	0.09

\*Sequence 1 では Period I - II を, Sequence 2 では Period II - I を求めた。

---

**要約統計量 (A-B)**

平均                    -0.264706  
 標準偏差                0.4939524  
 平均の標準誤差        0.1198011  
 平均の上側 95%        -0.010739  
 平均の下側 95%        -0.518673  
 n                         17

→治療 B のほうが, 平均で 0.265 L (95% CI: 0.011 L to 0.519 L) 高い。

**平均の検定**

	t 検定	Wilcoxon の符号付順位検定
検定統計量	-2.2095	-43.0000
P 値 (Prob >  t )	0.0421*	0.0250*

図 8 Paired t-test による治療効果の検証

表 1 のデータより時期ごとの差を計算したのが図上である。Pairwise difference データに関して paired t-test を実施した。ここでは, 持ち越し効果も時期効果もないことを前提としている。図 5 に示した unpaired t-test では時期効果はあってもよいので, そちらのほうが正式の方法である。どちら (図 5 も図 8) も被験者内比較に基づくので, paired analysis といえる。

## 二値データの 2x2 デザイン

これまで述べてきたのはアウトカムが数値の場合であったが, 二値データではどうすればよいか。糖

表 3 反復測定分散分析するために表 1 のデータを改変したデータ構造

Subject	Sequence	Period	Treatment	FEV <sub>1</sub>
1	1	I	A	1.28
2	1	I	A	1.6
3	1	I	A	2.46
4	1	I	A	1.41
5	1	I	A	1.4
6	1	I	A	1.12
7	1	I	A	0.9
8	1	I	A	2.41
9	2	I	B	3.06
10	2	I	B	2.68
11	2	I	B	2.6
12	2	I	B	1.48
13	2	I	B	2.08
14	2	I	B	2.72
15	2	I	B	1.94
16	2	I	B	3.35
17	2	I	B	1.16
1	1	II	B	1.33
2	1	II	B	2.21
3	1	II	B	2.43
4	1	II	B	1.81
5	1	II	B	0.85
6	1	II	B	1.2
7	1	II	B	0.9
8	1	II	B	2.79
9	2	II	A	1.38
10	2	II	A	2.1
11	2	II	A	2.32
12	2	II	A	1.3
13	2	II	A	2.34
14	2	II	A	2.48
15	2	II	A	1.11
16	2	II	A	3.23
17	2	II	A	1.25

尿病における低血糖をエンドポイントにしたクロスオーバー試験の例を, 図 16 に示した<sup>20)</sup>。結果は 2x2 表に要約され, 治療効果をオッズ比で示す。このデータは対応のあるデータなので, 条件つきオッズ比というものを計算する。そして, McNemar 検定で治療効果を検証する。この例では insulin lispro の regular insulin に対する低血糖リスクが 0.65 であることから, insulin lispro のほうでリスクは低いがある意ではないとわかる。McNemar 検定では時期効果



2×2Table:		Period I	Period II	
	Sequence 1	A	B	$n_1$
	Sequence 2	B	A	$n_2$

Source	d.f.
Between-subject	
Sequence	1
Within-subject	
Treatment	1
Period	1

Source	d.f.
Between-subject	
Sequence	1
Residuals	$n_1+n_2-2$
Within-subject	
Treatment	1
Period	1
Residuals	$n_1+n_2-2$
Total	$2(n_1+n_2)-1$

図 9 2×2 クロスオーバー試験の構造・誤差要因・自由度

2×2 クロスオーバーでは四つの平均値があるので (図上のセル数が4), 全平均値を前提としたときの自由度 (d.f.) は3になり, それが自由度=1+1+1=3と一致する (中間図)。図下では, 人数を単位にした自由度を示した。

JMP による解析  
Y: FEV actual value (表3のように)  
X: Sequence  
    Subject (Sequence) & 変量効果  
    Treatment  
    Period

SAS による解析 (参考)  
PROC GLM; CLASS Subject Sequence Treatment Period;  
MODEL FEV=Sequence Subject (Sequence) Treatment Period;  
RANDOM Subject (Sequence)/test;  
LSMEANS Treatment Period;

固定効果の検定

要因	パラメータ数	自由度	分母自由度	F 値	P 値 (Prob>F)
Sequence	1	1	15	<b>2.6357</b>	<b>0.1253</b>
Treatment	1	1	15	<b>4.6759</b>	<b>0.0472*</b>
Period	1	1	15	<b>1.3734</b>	<b>0.2595</b>

分散分析表

Source	df	SS (平方和)	MS (平均平方)	F	P-value
Between subjects					
Sequence	<b>1</b>	2.2212	2.2212	<b>2.6357</b>	<b>0.1253</b>
B-S Residual	<b>15</b>	12.641	0.8428		
Within-subject					
Treatment	<b>1</b>	0.5574	0.5574	<b>4.6759</b>	<b>0.0472</b>
Period	<b>1</b>	0.1637	0.1637	<b>1.3734</b>	<b>0.2595</b>
Residual	<b>15</b>	1.7882	0.1192		
Total	33	17.4102			

図 10 JMP による反復測定分散分析の解析結果

太字で示した統計量のみがこの分析で出力される。別の分析を行えば SS も出すことが可能である。MS は SS/df で得られる。たとえば, Sequence と Subject (Sequence) を入れた Between subjects モデルにより, Sequence と B-S residual と Total の SS は得られる。これは repeated measures ANOVA または split-plot ANOVA とよばれ, 混合モデル (mixed model) の一種である。

JMPによる解析					
Y	FEV actual value (表 3)				
X	Subject Treatment Period				
効果の検定					
要因	パラメータ数	自由度	平方和	F 値	P 値 (Prob>F)
Subject	16	16	14.862712	7.7921	0.0001*
Period	1	1	0.163726	1.3734	<b>0.2595</b>
Treatment	1	1	0.557420	4.6759	<b>0.0472*</b>
Period					
水準	最小 2 乗平均	標準誤差	平均		
I	1.9718668	0.08381330	1.97941		
II	1.8328391	0.08381330	1.82529		
最小 2 乗平均差の Student の t 検定 (Period effect)					
平均差 (Period I - II) = 0.139 (SE = 0.119), 95% CI: -0.114 to 0.39, P = 0.26					
Treatment					
水準	最小 2 乗平均	標準誤差	平均		
A	1.7740891	0.08381330	1.77000		
B	2.0306168	0.08381330	2.03471		
最小 2 乗平均差の Student の t 検定 (Treatment effect)					
平均差 (A - B) = -0.257 (SE = 0.119), 95% CI: -0.509 to -0.004, P = 0.047					

図 11 治療効果に関する推定結果 (95%信頼区間)

図 10 とは別のモデル解析で得られる。「最小 2 乗平均」は時期あるいは治療群で調整した平均値であり、「平均」が単純な平均値である。なお、Period I, II ごとの単純平均・標準偏差、および Treatment A, B ごとの単純平均・標準偏差は breakdown により得られる。

はないものと仮定している。もし時期効果があるなら、Mainland-Gart 検定<sup>21,22)</sup>などを用いたほうがよい。初期値や共変量を含む多変量解析をするには、対数線形モデルが利用できる。

### 例数設計

例数設計の公式はパラレル比較と同じだが、データ変動を示す分散の見積もりが異なる。図 17 に示したように、治療ごとの between 分散 (400 と想定) と within 分散 (100 と想定) などを用いて、クロスオーバーデザインでの分散は 200 と計算される。一方、パラレルデザインでの分散は 1800 である。すぐ想像できるように、パラレルのほうがクロスオーバーよりも 9 倍の例数を必要とする。治療効果 (群間差) の大きさは同じという仮定である。図 17 の

例では、between 分散と total 分散の比 (=400/500)、つまり第 I 期と第 II 期の級内相関係数は 0.8 である。

表 4 は、いろいろな状況でどの程度例数を削減できるかを示した。Between 分散、すなわち個人差が大きければ大きいほどクロスオーバーの効率は高い。Between 分散と within 分散の比は 2~4 倍程度と思われるため、クロスオーバーを使用することによる例数削減は 1/5~1/9 になる (表 4 参照)。別の報告によれば<sup>23,24)</sup>、パラレルのほうが 2.7~6.4 倍増えることが示された。また、例数削減および費用削減の効率について、Brown は理論的に考察している<sup>25)</sup>。

### 報告状況

Straube らはクロスオーバー論文 98 報を分析

4×2Table:		Period I	Period II	
	Sequence 1	A	A	$n_1$
	Sequence 2	B	B	$n_2$
	Sequence 3	A	B	$n_3$
	Sequence 4	B	A	$n_4$
Source		d.f.		
Between-subject				
	Sequence	3		
Within-subject				
	Treatment	1		
	Period	1		
	Carryover	1		
	Treatment×Carryover	1		
Source		d.f.		
Between-subject				
	Sequence	3		
	Residuals	$n_1+n_2+n_3+n_4-4$		
Within-subject				
	Treatment	1		
	Period	1		
	Carryover	1		
	Residuals	$n_1+n_2+n_3+n_4-3$		
Total		$2(n_1+n_2+n_3+n_4)-1$		

図 12 4×2 クロスオーバー試験 (Balaam デザイン<sup>15)</sup>) の構造・誤差要因・自由度

4×2 クロスオーバーでは八つの平均値があるので (図上のセル数が 8), 全平均値を前提としたときの自由度は 7 になり, それが自由度=3+1+1+1+1=7 と一致する (中間図)。図下では, 人数を単位にした自由度を示した。Treatment×Carryover 交互作用は想定しないのが通常なので, 図下ではその項を外した。

2×3Table:		Period I	Period II	Period III	
	Sequence 1	A	B	B	$n_1$
	Sequence 2	B	A	A	$n_2$
Source		d.f.			
Between-subject					
	Sequence	1			
Within-subject					
	Treatment	1			
	Period	2			
	Carryover	1			
Source		d.f.			
Between-subject					
	Sequence	1			
	Residuals	$n_1+n_2-2$			
Within-subject					
	Treatment	1			
	Period	2			
	Carryover	1			
	Residuals	$2(n_1+n_2-2)$			
Total		$3(n_1+n_2)-1$			

図 13 2×3 クロスオーバー試験の構造・誤差要因・自由度

2×3 クロスオーバーでは六つの平均値があるので (図上のセル数が 6), 全平均値を前提としたときの自由度は 5 になり, それが自由度=1+1+2+1=5 と一致する (中間図)。図下では, 人数を単位にした自由度を示した。

	Period		Within subject difference (A-B)
	I	II	
A then B			
Mean (SD)	$\bar{y}_{1A I}$ (SD <sub>1A I</sub> )	$\bar{y}_{1B II}$ (SD <sub>1B II</sub> )	d <sub>1</sub> (SD <sub>1</sub> )
n	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>
B then A			
Mean (SD)	$\bar{y}_{2B I}$ (SD <sub>2B I</sub> )	$\bar{y}_{2A II}$ (SD <sub>2A II</sub> )	d <sub>2</sub> (SD <sub>2</sub> )
n	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>
Treatment effect			
LSMean (95% CI)			d <sub>3</sub> (95% CI)
n			n <sub>1</sub> + n <sub>2</sub>
Paired analysis			P-value
または			
	A (n <sub>1</sub> + n <sub>2</sub> )	B (n <sub>1</sub> + n <sub>2</sub> )	A-B (n <sub>1</sub> + n <sub>2</sub> )
Outcome 1	Mean (SD)	Mean (SD)	d <sub>3</sub> (95% CI), P
Outcome 2			
...			

図 14 数値データの場合の 2×2 クロスオーバーデザイン解析結果の提示法

上の様式のほうが情報は多いが、アウトカムが複数ある場合には見にくくなるので、下の様式のほうが見やすいかもしれない。LSMean (least square mean) とは最小 2 乗平均のことであり、時期効果で調整済みの平均値である。単純平均値とは異なるので、両者を区別したほうが望ましい。

	Period		Within subject difference (A-B)
	I	II	
A then B			
Mean (SD)	1.57 (0.57)	1.69 (0.73)	-0.118 (0.354)
n	8	8	8
B then A			
Mean (SD)	2.34 (0.73)	1.95 (0.72)	-0.396 (0.581)
n	9	9	9
Treatment effect			
LSMean (95% CI)			-0.256 (-0.509 to -0.004)
n			17
Paired analysis			P=0.047
または			
	A (n=17)	B (n=17)	A-B (n=17)
FEV <sub>1</sub>			
Mean (SD)	1.77 (0.66)	2.03 (0.78)	
LSMean (95% CI), P			-0.256 (-0.509 to -0.004), P=0.047
...			

図 15 喘息患者を対象とした 2×2 クロスオーバー試験で得られた解析結果の要約表

図 14 に示した様式に合わせて解析結果を記載した。これがベストではないかもしれないが、ここで示した要約指標は含めてもらいたい。Period ごとの mean (SD) は表 3 データを使って breakdown により得られる。Within subject difference (A-B) ごとの mean (SD) は表 1 データの Period I と II の差を求めれば (Sequence 2 では II - I とする), breakdown by sequence で得られる。Treatment effect は図 11 の Treatment で得られている。

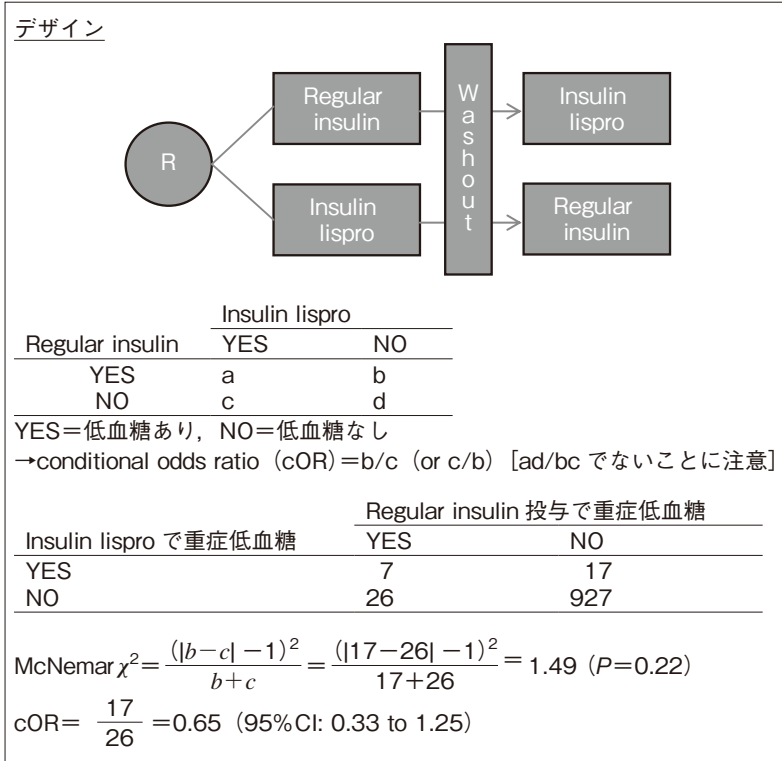


図 16 二値データを伴う 2×2 クロスオーバーデザインの解析法

対応のある二値データの際のオッズ比は通常のものとは異なり, 条件つきオッズ比 (conditional odds ratio; cOR と略す) とよばれる。McNemar 検定で効果は検証できる。ここでは sequence (or carryover) effect, period effect はないものと仮定している。

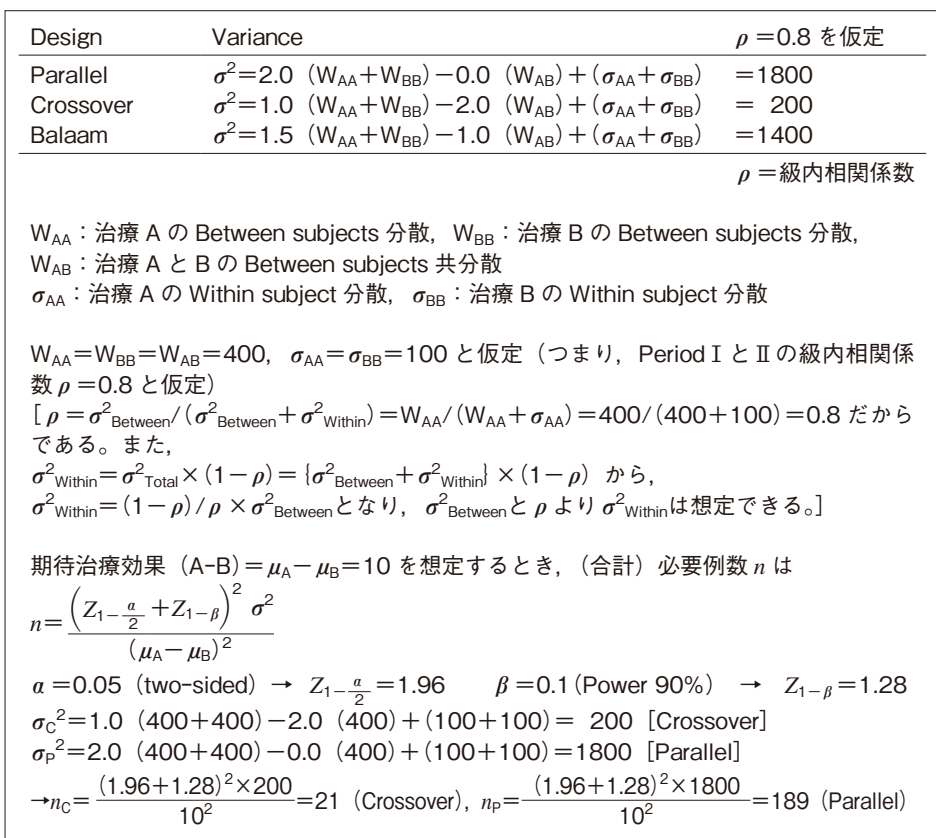


図 17 クロスオーバー試験における例数設計

Period I と Period II の相関係数が高ければ高いほど, クロスオーバー試験の検出力は高くなる。つまり必要例数は少なくてすむ。Period 間の相関係数が高いほど, クロスオーバー試験における被験者内比較をする利点が高まる。

表 4 クロスオーバー試験における例数削減

Between 分散	Within 分散	相関係数 ( $\rho$ )	分散 ( $\sigma^2$ )		例数倍率 (P/C)
			Parallel	Crossover	
1	: 1	1/2=0.5	6	2	3 倍
2	: 1	2/3=0.67	10	2	5 倍
3	: 1	3/4=0.75	14	2	7 倍
4	: 1	4/5=0.8	18	2	9 倍

時期 I - II 間の (級内) 相関係数,  $\rho = \text{Between 分散} / (\text{Between 分散} + \text{Within 分散})$  で算出した。また, 図 17 で示した分散 ( $\sigma_c^2 = 1.0 (W_{AA} + W_{BB}) - 2.0 (W_{AB}) + (\sigma_{AA} + \sigma_{BB}) \leftarrow \text{Crossover}$ ,  $\sigma_p^2 = 2.0 (W_{AA} + W_{BB}) - 0.0 (W_{AB}) + (\sigma_{AA} + \sigma_{BB}) \leftarrow \text{Parallel}$ ) を利用すると, Parallel は Crossover よりも何倍の例数が必要かを算出できる。両者ともに群間差の想定値は同じである。Margarone らは通常 2.7~6.4 倍必要であるという試算を示した<sup>23)</sup>。

表 5 クロスオーバーデザインとパラレルデザインの対比

特徴	クロスオーバー	パラレル
Size	○小さくてすむ	×大きくなりがち
Cost	○安くてすむ	×高くなりがち
Duration	○短くてすむ	×長くなりがち
Between-subject variation	○大きいほど利点	×大きいほど欠点
Attrition	×起こりやすい	○起こりにくい
Carryover	×入ると厄介	○入らない
Disease type	×安定した慢性疾患に限る (喘息, 狭心症, 糖尿病, 等)	○とくに制約はない (感染症など急性疾患も可)
Response	×早期のものに向く (早く動く血糖値など)	○早期でなくてもよい (血管イベントなども可)
Reversible	×戻らないといけない (死亡や心筋梗塞はだめ)	○戻らなくてもよい (死亡や心筋梗塞も可)

両者ともに良し悪しがある。Attrition (摩擦) とは欠測値や中途脱落のように, デザインからの逸脱事象を指す。

し<sup>26)</sup>, period ごとに図 15 で提示したような break-down を行っていたのは 23% しかなかった。

Mills らは 526 件の RCT を調査した<sup>27)</sup>。クロスオーバーは 116 件 (22%), 有効性 48%・薬物動態 28%・非薬剤 30% (重複あり), 2×2 デザインの利用が 64% であった。例数中央値は 15 例 (四分位範囲: 8~38 例) であった。EMA 生物学的同等性ガイドライン<sup>11)</sup>および FDA 生物学的同等性統計ガイドライン<sup>28)</sup>では, 評価可能例数として最低 12 例は必要と書かれている。Carryover を触れたのはわずか 29% だが, washout については 70% で触れられていた。Carryover effect 検定が 17%, period effect 検定が 13%, 治療効果の paired analysis は 96% で実施していた。例数設計は 20% でしか記載はなかった。

また, Li らは 83 件のクロスオーバー試験を調査し<sup>19)</sup>, 採用理由 (rationale) は 5%, washout したのは 41%, 事前の例数設計は 54%, 事後パワー計算は 17%, AB/BA デザインの利用は 87%, paired analysis は 76% 実施, carryover effect 検定は 10%, period effect 検定は 18%, 初期値利用は 56% であった。

### パラレルデザインとの比較

クロスオーバーデザインとパラレルデザインの対比を表 5 に示した。クロスオーバーデザインには例数・費用の削減, 期間短縮などの利点もあるが, 欠測値増加, 持ち越し効果, 安定疾患に限定, 早期反応に限定, 可逆指標に限定するなどの欠点もある。



表 6 持ち越し効果への対処法

Carryover effect がありそうだと	→ パラレルデザインを考える → 高次クロスオーバーデザイン (Balaam デザイン等) を考える → ウォッシュアウト期間を長めにとる → 第 I 期当初の値にほぼ戻るまで待ってから第 II 期へ入る → アウトカムが安定している被験者に限定する → 初期値を利用し、初期値からの差をデータにする
Carryover effect が検知された	→ 第 I 期のデータのみを用いる → やり直しをする

表 7 クロスオーバー試験を報告する際の 12 個のチェックリスト

1. クロスオーバー (crossover) という用語をタイトルに示したか?
2. クロスオーバーデザインを採用した根拠 (rationale) を示したか?
3. デザイン (2×2 など) 型を図で示したか?
4. ウォッシュアウト期間の記述をしたか、そして根拠を示したか?
5. 例数設計の根拠を示したか (事前、あるいは事後パワー)?
6. Paired (within subject) analysis に基づいて治療効果を推定したか?
7. 治療効果を評価するための統計手法を示したか? Paired <i>t</i> -test, Unpaired <i>t</i> -test, Wilcoxon test, Repeated ANOVA, Mixed model など
8. 治療効果には推定値・検定結果に加えて、95%信頼区間も示したか?
9. 持ち越し効果の検討をしたか (検定を含め)?
10. 時期効果の検討をしたか (検定を含め)?
11. 初期値 (baseline) を利用したか?
12. 欠測値データの取扱いを示したか? (除外, 補完, 多変量解析など)

これらの特徴をよく吟味したうえで、クロスオーバー試験を採用するときの根拠を考えてもらいたい。

クロスオーバーデザインの最大の難点は、何と言っても持ち越し効果である。それがありそうな場面では、表 6 に示したように、パラレルデザインを考えたり、図 12, 13 に示したような高次デザインを考えたり、ウォッシュアウト期間を長めにするなどを考えたほうがよい。より安定している被験者に絞るとか、第 II 期へ入る条件を設定することもたいせつである。実際に持ち越し効果が有意になってしまったら、第 I 期のデータだけ用いて解析するか、別のデザインでやり直しするしかないだろう。

### チェックリスト

最後に、クロスオーバー試験を実施して論文として報告するとき、どのようなことに気をつければよいかについて述べる。表 7 に 12 個のポイントを示

した。報告状況からみると<sup>18,19,26,27)</sup>、採用理由、事前の例数設計、持ち越し効果・時期効果の検討、paired analysis 結果の適切な報告 (メタアナリシスしやすいように) などは、とくに気をつけてもらいたい。欠測値の扱いについては補完するよりは、多変量解析である反復測定分散分析を行えばよい。デザイン行列が正則であれば、最小 2 乗平均値や 95% 信頼区間は求めることができる。

### 文 献

- 1) Simpson TW. Experimental methods and human nutrition (with discussion). J R Stat Soc Series B 1938; 5: 46-669.
- 2) Cochran WG, Antrey KM, Cannon CY. A double change-over design for dairy cattle feeding experiments. J Dairy Sci 1941; 24: 937-57.
- 3) Fieller EC. The biological standardization of insulin (with discussion). Supplement to the J R Stat Soc 1940; 7: 1-64.
- 4) Williams EJ. Experimental designs balanced for the estima-

- tion of residual effects of treatments. *Aust J Sci Res* 1949; 2: 149-68.
- 5) Finney DJ. Cross-over designs in bioassay. *Proc R Soc Lond B Biol Sci* 1956; 145: 42-60.
  - 6) Sibbald B, Roberts C. Understanding controlled trials crossover trials. *BMJ* 1998; 316: 1719-20.
  - 7) Sibbald B, Roberts C. Design and analysis of cross-over trials. London: Chapman and Hall; 1989.
  - 8) Chow SC, Lie JP. Design and analysis of bioavailability and bioequivalence studies. New York: Marcel Dekker; 1992.
  - 9) Senn S. Cross-over trials in clinical research. Chichester: John Wiley; 1993.
  - 10) US Food and Drug Administration. Guidance for industry-bioavailability and bioequivalence studies submitted in NDA or INDs -general considerations. Rockville; 2014; p.24.
  - 11) Patel HI. Use of baseline measurements in the two-period cross-over design. *Commun Stat Theory Methods* 1983; 12: 2693-712.
  - 12) European Medicines Agency. CHMP guideline on the investigation of bioequivalence. London; 2010. p.15 (method) and p.8 (size).
  - 13) Grizzle JE. Two two-period change-over design and its use in clinical trials. *Biometrics* 1965; 21: 467-80.
  - 14) Grieve AP. The two-period changeover design in clinical trials. Letter to the editor. *Biometrics* 1982; 38: 517.
  - 15) Balaam LN. A two-period design with  $t^2$  experimental units. *Biometrics* 1968; 24: 61-73.
  - 16) Chow SC, Lin JP. On assessment of bioequivalence under a higher-order crossover design. *J Biopharm Stat* 1992; 2: 239-56.
  - 17) Koch GG, Amara IA, Brown BW Jr, Colton T, Gillings DB. A two-period crossover design for the comparison of two active treatments and placebo. *Stat Med* 1989; 8: 487-504.
  - 18) Elbourne DR, Altman DG, Higgins JPT, Curtin F, Worthington HV. Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2002; 11: 140-9.
  - 19) Li T, Yu T, Hawkins BS, Dickersin K. Design, analysis, and reporting of crossover trials for inclusion in a meta-analysis. *PLOS ONE* 2015; DOI: 10.1371/journal.pone.0133023.
  - 20) Brunell RL, Llewelyn J, Anderson JH Jr, Gale EAM, Kolvisto VA. Meta-analysis of the effect of insulin lispro on severe hypoglycemia in patients with type 1 diabetes. *Diabetes Care* 1998; 21: 1726-31.
  - 21) Mainland D. Elementary medical statistics. 2nd ed. Philadelphia: Saunders; p.36-38.
  - 22) Gart JJ. An exact test for comparing matched proportions in crossover designs. *Biometrika* 1969; 56: 75-80.
  - 23) Margarone JE, Liebor C, Hall RE, Braun RE, Weiner M. A comparison of cross-over versus parallel-group design in the evaluation of analgesic efficacy after molar extraction. *Clin Pharmacol Ther* 1995; 58: 453-8.
  - 24) Richens A. Proof of efficacy trials: cross-over versus parallel-group. *Epilepsy Res* 2001; 45: 43-7.
  - 25) Brown BW Jr. The crossover experiment for clinical trials. *Biometrics* 1980; 36: 69-79.
  - 26) Straube S, Werny B, Friede T. A systematic review identifies shortcomings in the reporting of crossover trials in chronic painful conditions. *J Clin Epidemiol* 2015; 68: 1496-503.
  - 27) Mills EJ, Chan A-W, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials* 2009; 10: 27.
  - 28) US Food and Drug Administration. Guidance for industry-statistical approach to establishing bioequivalence. Rockville; 2001. p.8.

\* \* \*