

共分散分析の無作為化比較臨床 試験における留意事項

ノボノルディスクファーマ株式会社

清見文明、西田朋由、西島啓二

2005年10月31日（於）吉祥寺東急イン

第4回東京吉祥寺地区月例研究会

本日の報告

- はじめに
- **Points to consider...**の要約
- 復習: 最小二乗平均、交互作用を主要な解析に含めないこと

以降、計量生物セミナーでの報告改定版 + α

- 共変量による調整を行うことについて
- 交互作用について

本日の報告(続き)

- 共分散分析におけるいくつかの留意点
 - 結果の提示で、誤解を与えてしまった例
(調整解析が偏りの補正に役立った例も含む)
(共分散分析の結果の提示(案)も含む)
 - 希薄なセルが推定値に大きく影響する例
- 共分散分析に対する指摘
- 感度(安定性)の解析
- PtCへの若干のコメント
- まとめ

付録: ベースライン共変量と平均への回帰

はじめに

興味をもった発端

検出力が高いとの理由で、2000年頃(?)から
主要な解析として、共分散分析を使いはじめた。

**Points to consider on adjustment for
baseline covariate**の案が作成された(2001年)。
(とても強く共分散分析を推奨)

- 本当に検出力が実質的に高いのか
- 交互作用があるのでは
- ベースライン共変量をどのように解釈するのか...

はじめに

前 提

- ・エンドポイントは連続変数
- ・無作為化比較臨床試験
- ・質的交互作用はない
- ・カテゴリ変数：層別無作為化に用いた因子のみ統計解析で考慮する
- ・連続変数の共変量はエンドポイントのベースラインをその代表とする

はじめに

共分散分析のモデル

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma x_{ijk} + \varepsilon_{ijk}$$

$$\sum \alpha_i = 0, \sum \beta_j = 0, \varepsilon_{ijk} \sim \text{i.i.d.N}(0, \sigma^2)$$

k は以下の通り

	群1	群2	...	群a
層1	n_1	n_1	...	n_1
層2	n_2	n_2	...	n_2
...			...	
層b	n_b	n_b	...	n_b

Points to consider on adjustment for baseline covariate の要約

層別はバランスを確実にするために利用され、また管理上の理由で用いられる。通常、層別因子は主要な解析に共変量として含められるべき。

主要変数に影響を及ぼすことが既知の変数、and/or 強い臨床的合理性を有する変数は、主要解析で共変量として考慮されるべき。このことは事前にプロトコル又はSAPに記載すべき。

事後的に偏りが見られた項目は、それを理由に解析に含めてプライマリーとするべきではない。

無作為化後に測定される変数は、潜在的に治療に影響されるので、主要な解析に共変量として含めるべきではない。

主要変数のベースラインは、通常共変量に含められるべき(主要変数が raw outcomeであろうが変化量であろうが)。

少数の共変量が主要な解析には含められるべきである。大容量データが多いの共変量を含めることをサポートするかもしれないが、それら各々の正当性を示さねばならない。

事前知識がなければ、共変量と主要変数の間には単純な関数形を仮定すべきである（線形性又は連続変数の2値化）。

モデルの妥当性をチェックしなければならない。これは特に線形／非線形モデルに対して重要である。規格・設定の誤りが(mis-specification)治療効果の不正確な推定を導くことがあるからである。通常の線形モデルの下でさえ、外れ値の影響には気を配る必要がある。

調整解析が提示されるときはいつでも、共変量のカテゴリ毎（連続変数は意味があればカテゴリ化して）の治療効果も提示すべきである（モデルの仮定の妥当性評価を可能にするために）

主要な結果の頑健性を調査するために、感度分析を事前に計画し実施すべきである。

主要な解析には治療と共変量の交互作用を含めるべきではない。実質的な交互作用が事前に予期されるならば、当該試験は部分集団解析を可能にするようにデザインされるべきである。

共変量を含めない探索的解析が、理解促進のため行われるかもしれない。

主要な解析とは、プロトコル又はSAPに曖昧さなく事前に特定され、正確に実行して解釈されるものである。この主要な解析が試験から導かれる結論をsupportすべきである。他の多くの妥当な解析が存在するので、事前に計画された解析に基づく結果が最も信頼されるであろう。

復習 1 (最小二乗平均)

$$\text{モデル: } Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma x_{ijk} + \varepsilon_{ijk}$$

群i、kの最小二乗平均は以下の通り

$$\begin{aligned} \hat{\mu} + \hat{\alpha}_i + \frac{1}{b} \sum_{j=1}^b \hat{\beta}_j + \hat{\gamma} \bar{X} \dots \\ \hat{\mu} + \hat{\alpha}_k + \frac{1}{b} \sum_{j=1}^b \hat{\beta}_j + \hat{\gamma} \bar{X} \dots \end{aligned}$$

群間比較はこれらの差: $\hat{\alpha}_i - \hat{\alpha}_k$

例

	治療群1	治療群2
層1	5, 7	6
層2	3	2, 4

各セルの平均

6 6

3 3

	治療群1	治療群2
標本平均	5	4
最小二乗	4.5	4.5

$$\mu = 4.5, \alpha_i = \pm 0, \beta_j = \pm 1.5$$

標本平均 (モデルは $E(Y_{ijk}) = \mu + \alpha_i + \beta_j$)

$$\text{群1} = \frac{[2(\mu + \alpha_1 + \beta_1) + (\mu + \alpha_1 + \beta_2)]}{3} = \mu + \alpha_1 + \frac{2}{3}\beta_1 + \frac{1}{3}\beta_2$$

$$\text{群2} = \frac{[(\mu + \alpha_2 + \beta_1) + 2(\mu + \alpha_2 + \beta_2)]}{3} = \mu + \alpha_2 + \frac{1}{3}\beta_1 + \frac{2}{3}\beta_2$$

$$\text{群1} - \text{群2} = \alpha_1 - \alpha_2 + (\beta_1 - \beta_2)/3$$

調整済みの平均の解釈 (Reviewの続き)

	群1	群2	計
層1	15	11	13
層2	9	5	7
計	12	8	10

$$E(Y_{ijk}) = \mu + \alpha_i + \beta_j$$

$$\mu = 10$$

$$\alpha_i = \pm 2$$

$$\beta_j = \pm 3$$

群	層	値	群	層	値
1	1	13	1	1	13
1	1	17	1	1	17
1	2	6	1	2	6
1	2	12	1	2	12
2	1	10	1	2	8
2	1	12	1	2	10
2	2	4	2	1	10
2	2	6	2	1	12
			2	2	4
			2	2	6
			2	2	5
			2	2	5

	群1	群2	計
層1	15	11	13
層2	9	5	7
計	11	7	9

means
←

LSmeans
←

LSmeans
←

means
↙

復習2（交互作用を含めると・・・）

- ・交互作用を含めないとき

主効果の検定は、 α 、 β の制約式の影響を受けない。
(e.g. $\sum \alpha_i = 0$ でも、 α_1 and $\alpha_2 = 0$ でも治療群や層の検定のp値はかわらない)

- ・交互作用を含めると

主効果の検定は、 α 、 β の制約式の影響を受ける。

通常

$$\Sigma \alpha_i = 0, \quad \Sigma \beta_j = 0,$$

$$\Sigma_i (\alpha \beta)_{ij} = 0 \quad (j=1,2,\cdots,b)$$

$$\Sigma_j (\alpha \beta)_{ij} = 0 \quad (i=1,2,\cdots,a)$$

一般

$$\Sigma V_i \alpha_i = 0, \quad \Sigma W_j \beta_j = 0,$$

$$\Sigma_i V_i (\alpha \beta)_{ij} = 0 \quad (j=1,2,\cdots,b)$$

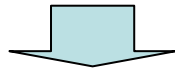
$$\Sigma_j W_j (\alpha \beta)_{ij} = 0 \quad (i=1,2,\cdots,a)$$

$$\Sigma V_i = 1, \quad \Sigma W_j = 1$$

	被験薬	プラセボ
男性	(32) -16	(32) 0
女性	(24) -8	(24) 0
合計	(56) -12.6	(56) 0

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

$$V_i = 1/2, \quad W_j = 1/2$$

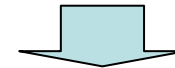


$$\mu = -6 \quad \alpha_i = \pm 6 \quad \beta_j = \pm 2$$

$$(\alpha\beta)_{11} = (\alpha\beta)_{22} = -2$$

$$(\alpha\beta)_{12} = (\alpha\beta)_{21} = 2$$

$$V_i = 1/2, \quad W_1 = 1, \quad W_2 = 0$$



$$\mu = -4 \quad \alpha_i = \pm 8$$

$$\beta_1 = -4 \quad \beta_2 = 0$$

$$(\alpha\beta)_{11} = (\alpha\beta)_{21} = 0$$

$$(\alpha\beta)_{12} = 4 \quad (\alpha\beta)_{22} = -4$$

復習2の要約

- ・交互作用を主解析に含めるべきではない.
- ・交互作用を含めた探索的解析では, 交互作用の検定のみ行うべき。LSMEANや主効果の検定結果は出力しないほうが良いかもしれない.
- ・制約式に物理的に積極的な意味がない限り、交互作用がある場合の主効果の検定に普遍的な意味をもたせるのは難しい(広津、1976).

計量生物セミナーの改定版 + α 、

— 共変量による調整について —
(PtC 周辺のコメント・考察、留意事項)

推 奨

PtC同様：無作為化に用いた層別因子とエンドポイントのベースラインは、主要な解析で考慮する。精度改善と偏りの調整（特に少数例の場合）が期待できる

合 意

- ・ ある程度の精度が確保されていれば、共変量を含めても含めなくても良い。
- ・ 交互作用があっても主効果だけを共変量に含めて解析できる。

留意点

- ・ 誤解がないように結果を提示する。
- ・ LSMeansの解釈を臨床の人たちと共有する。
- ・ ~~調整平均の群間の差が希薄なセルに影響されることがある。~~
各群の調整平均が希薄なセルに影響される。（交互作用を含めないなので、群間比較への影響は大きくない。）

- 1 共変量による調整を行うことについて
- 2 交互作用について
- 3 共分散分析におけるいくつかの留意点
 - 結果の提示で、誤解を与えてしまった例
 - 希薄なセルが推定値に大きく影響する例
- 4 共分散分析に対する指摘
- 5 感度(安定性)の解析
- 6 PtCへの若干のコメント
- 7 まとめ

1. 共変量による調整を行うことについて

PtC

通常、層別因子は主要な解析に共変量として含められるべき。

主要変数への影響が既知の変数は、主要解析で共変量として考慮すべき

主要変数のベースラインは、通常共変量に含められるべきである

(共変量調整のメリット: 精度、偶然の偏りの補正)



同意

当社は共変量で調整する解析を主要な解析にしている

(およそ2000年以降殆ど全ての新規プロトコルでANCOVA)

1. 共変量による調整を行うことについて

t-検定は？

- ーメリット:理解しやすい
- ー正確である

個人的には・・・

わずかな精度を犠牲にして理解のし易さを優先しても、複雑にならない範囲で精度を追求しても、正確な推測という本質を損なわない限り、いずれの方法でも認められるべき。

実際、前後の差や比で(等簡単な指標の工夫で)、層別因子やベースラインの影響が殆ど消える例がある

1. 共変量による調整を行うことについて

		前	後	前後差
1型	被験薬(195)	7.4	7.3	-0.1
	対照薬(98)	7.4	7.3	-0.1
2型	被験薬(67)	7.8	7.7	-0.1
	対照薬(35)	7.8	7.7	-0.1

1. 共変量による調整を行うことについて

		前	後	前後差
層1	被験薬	7.4	7.1	-0.3
	対照薬	7.4	7.3	-0.1
層2	被験薬	7.8	7.5	-0.3
	対照薬	7.8	7.7	-0.1

$$E(Y_{ij1k}) = \mu + \beta_j \quad \quad \quad \mathbf{7.6 \quad \pm 0.2}$$

$$E(Y_{ij2k}) = \mu' + \alpha_i + \beta_j \quad \quad \quad \mathbf{7.4 \pm 0.1 \pm 0.2}$$

$$E(Y_{ij2k}) - E(Y_{ij1k}) = (\mu' - \mu) + \alpha_i \quad \quad \quad \beta'_j \quad \begin{array}{l} \text{層により治} \\ \text{療の反応} \\ \text{が異なる} \end{array}$$

1. 共変量による調整を行うことについて

合 意

- ・ ある程度の精度が確保されていれば、共変量は含めても含めなくても良い.

2. 交互作用について

	被験薬	プラセボ
男性	(32) -16	(32) 0
女性	(24) -8	(24) 0
合計	(56) -12.6	(56) 0

	被験薬	プラセボ
男性	-14.3	-1.7
女性	-10.3	2.3
合計	-12.3	0.3

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

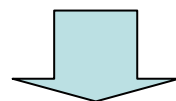
$$\mu = -6 \quad \alpha_i = \pm 6 \quad \beta_j = \pm 2$$

$$(\alpha\beta)_{11} = (\alpha\beta)_{22} = -2$$

$$(\alpha\beta)_{12} = (\alpha\beta)_{21} = 2$$

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j$$

$$\mu = -6 \quad \alpha_i = \pm 6.3 \quad \beta_j = \pm 2$$



交互作用を無視しても、少なくとも群差の推定値は標本平均の差に等しい。

2. 交互作用について

治療群*i*と*k*の標本平均の差: $\bar{Y}_{i..} - \bar{Y}_{k..}$

交互作用を含めないANOVAモデルのもとでの最小二乗平均

$$\widehat{\mu + \alpha_i} = \bar{Y}_{i..} + (1/b) \sum_{j=1}^b \bar{Y}_{.j.} - \bar{Y}_{...}$$

$$\widehat{\mu + \alpha_k} = \bar{Y}_{k..} + (1/b) \sum_{j=1}^b \bar{Y}_{.j.} - \bar{Y}_{...}$$

2. 交互作用について

連続変数の共変量についても、

$$\bar{X}_{i..} = \bar{X}_{k..}$$

が成り立てば(期待的に成り立つ)、最小二乗平均の差は、標本平均の差と等しくなる。

この場合の式は簡単でないので省略

2. 交互作用について

- 交互作用が存在するが、それを無視して主効果だけをモデルに含めて解析することは可能であり、精度の向上に貢献する(少なくとも群差の推定量が標本平均の差に等しいので)。
- ただし、各群および各群*層の最小二乗平均解釈は、明確ではない。

3. いくつかの留意点

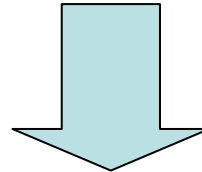
- (1) 結果の提示について(誤解を与えた例と、表の案)
- (2) 調整平均の解釈(Review)
- (3) 希薄なセルが与える影響

(3) 希薄なセルが結果に影響する例

- 希薄な層がある例は、極端な例です。一般に述べたかったこと：症例数の少ない層が、最小二乗平均に大きな影響を及ぼす。
- したがって、標本平均と大きく異なることがある。
- ベースラインからの変化も、最小二乗平均を使って解釈すべきでないかもしれない。
- しかし、交互作用を含めない限り、治療群の差に与える影響は小さい(注：希薄なセルを含む層別因子を解析で考慮することを薦めるものではない)。

4. 共分散分析に対する指摘

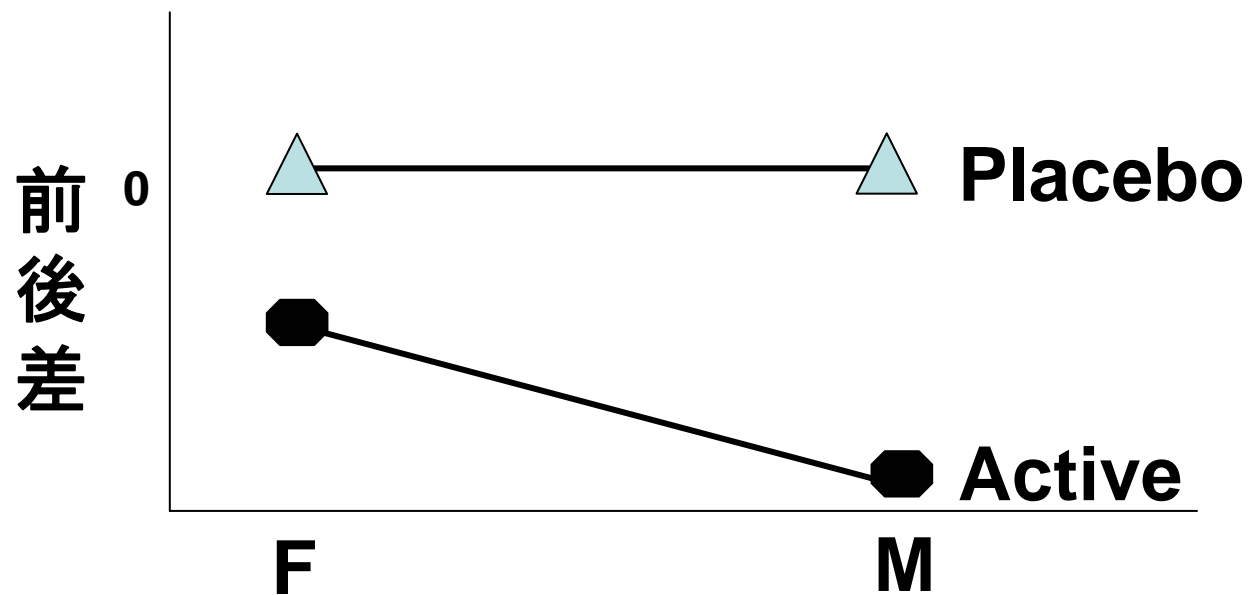
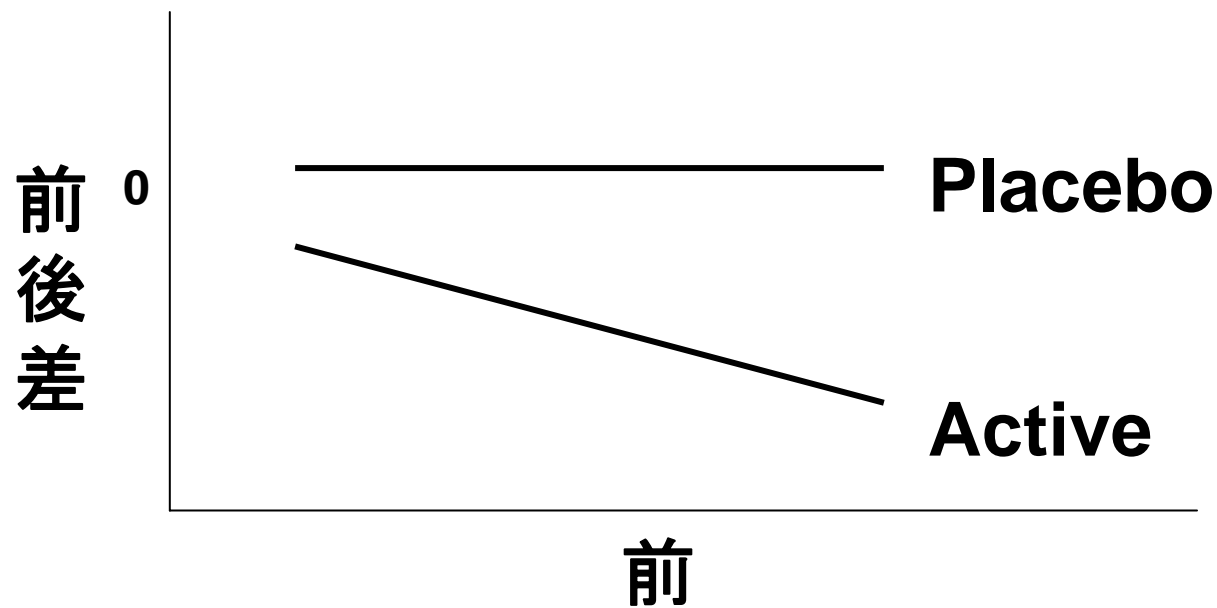
- あるモデルの成立下で比較の上で正確性である
- 精度が改善する. 偏りを調整する.

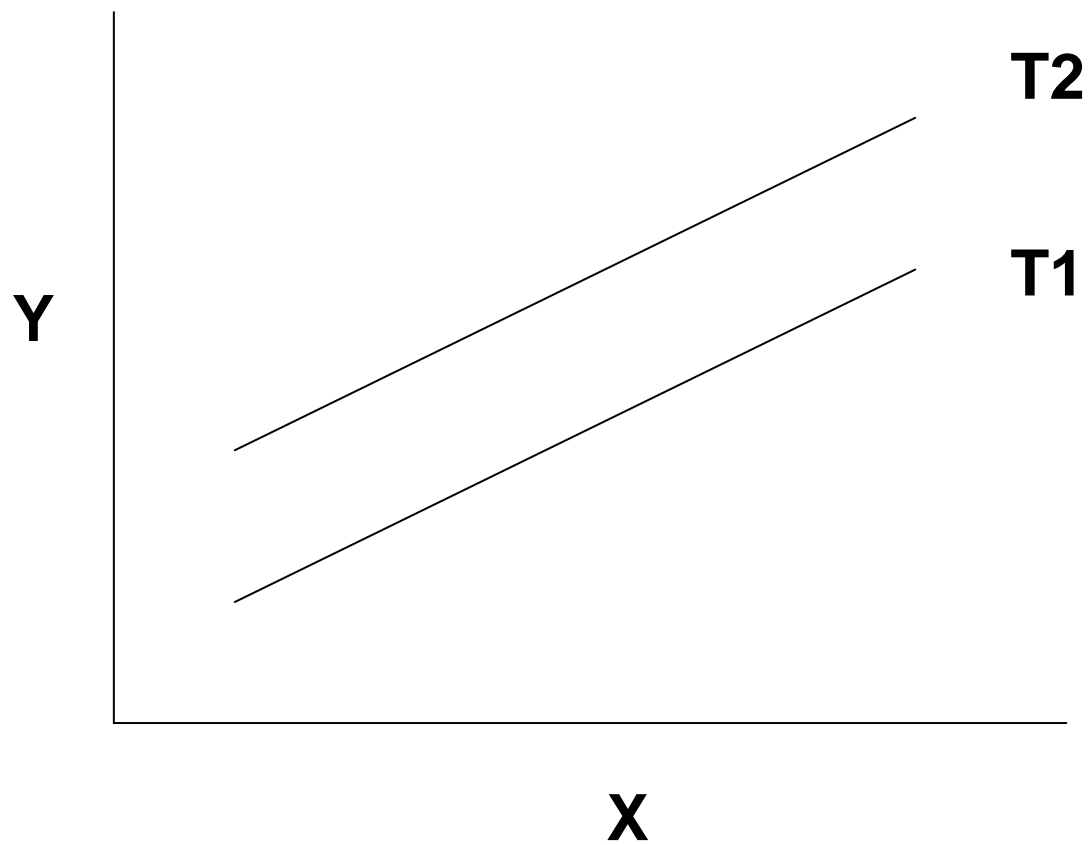


しかし、...

- 精度改善を理由に、解釈が容易とはいえない最小二乗平均が主役になる。(差の推定量は解釈できるが)
- 連続変数で、平行でもないのにX...で比較する。
- プラセボ対照の場合は、無視できない交互作用が起きている可能性が高い。これを完全に無視する解析が妥当か？(連続変数なら、治療群で傾きが異なるのに、同じとするモデルで解析する)
- ベースライン共変量においては、直線を仮定した推定量は何を表すのか？

標本平均/t-検定は上記の問題を解決しているのか？





4. 共分散分析に対する指摘

- 標本平均／t検定を用いたからといって、共分散分析での問題を解決しているわけではない。
- しかし、最小二乗平均の解釈と、直線の仮定に関する懸念はもつともで、適切な回答は見つかっていない。

感度(安定性)の解析

- ・外れ値を除いた解析(共変量の調整に限らない)
- ・PPSでの解析(共変量の調整に限らない)
- ・治療群と共変量との交互作用を含めた解析.
- ・共変量を除いた解析
- ・ノンパラメトリックな方法を用いた解析
- ・場合によっては, 経時データの解析(いわゆるmissingデータの解析を含む)

PtCへのコメント

通常、層別因子は主要な解析に共変量として含められるべき。

→Yes. しかし、層間で薬剤への反応が異ならなければ変化量(率)で影響を除去できる。

主要変数への影響が既知の変数は、主要解析で共変量として考慮すべき

→Yes. しかし、指標の工夫で共分散分析を避けられることがある。ただし、少例数のとき偶然の偏り補正を目的に共分散分析の利用するのも一考。

事後的に偏りが見られた項目は、それを理由に解析に含めてプライマリーとするべきではない。 →Yes

無作為化後に測定される変数は、潜在的に治療に影響されるので、主要な解析に共変量として含めるべきではない。

→Yes. しかし、無作為化後のベースライン測定は頻繁に見られる(e.g. ベースラインHbA_{1c}等、選択基準に関連するが採血と値入手に時間差がある場合など)。DB、無作為化と測定の時間短縮、客観的指標で影響は小さいはず(被験者に知らせないのは大前提)。HbA1Cならば、オープンラベルでも問題なし。血圧の場合は確かにバイアス混入の恐れはある。なお、本件は調整の問題ではなく、一般の留意事項。

← まだ、独り言の段階です

主要変数のベースラインは、通常共変量に含められるべきである

→Yes. しかし、指標の工夫で共分散分析を避けられることがある。

少数の共変量が主要な解析には含められるべきである。 →Yes

事前知識がなければ、共変量と主要変数の間には単純な関数形を仮定すべき → Yesと考えますが、その他の関係を仮定した経験がないので、コメントは控えたい。

モデルの妥当性をチェックしなければならない。

→Yes。しかし、安定性との区別が私には明確ではない。共分散分析では正規性の仮定。直線性も含められるかもしれない。交互作用は安定性の意味で、補足的解析と捉えています。

調整解析を実施するときは、共変量のカテゴリ毎の治療効果も提示すべき

→大賛成。交互作用解析とともにいつも出力している。

主要な結果の頑健性を調査するために、感度分析を事前に計画し実施すべき。 →Yes

主要な解析には治療群との交互作用を含めるべきではない。→Yes
実質的な交互作用が事前に予期されるならば、部分集団解析を可能にするようにデザインされるべき。

→正論でYes. しかし、常に全ての水準で検討に十分な症例数を確保できるわけではない。一般には層別割付をして、全水準で方向性が変わらないことがみられれば良いのでは？

主要な水準が存在するならば、可能なら、当該水準の部分集団だけで証明できる例数を確保する？

層別因子のあるプラセボ対照試験では、交互作用が存在するのが一般的で、常にこの条件を満足させるのは難しい？（1群30例の用量反応試験など）

共変量を含めない探索的解析が、理解促進のために実施。→Yes

事前に計画された解析に基づく結果が最も信頼されるであろう。→Yes

まとめ

指標を工夫し、t-検定を主要な解析にすることも可能だが、検出力と偏りの補正の観点から、共分散分析を推薦できる。適用にあたっては下記の留意事項があげられる。

- (共)分散分析は、最小二乗平均をよく理解した上で適用することが重要である(誤った解釈をしないために)。
- 交互作用を含めると、重みにより主効果の検定結果は異なる。補足的・探索的解析で交互作用を調べるにあたって、主効果の検定結果や治療群の最小二乗平均などの出力には、十分な配慮が必要である(多くの場合、個人的には出力しないほうが良いと思っています)。
- 交互作用がある場合、各治療群の最小二乗平均に解釈を与えるのは困難かもしれない(治療群間の差は解釈可能)

まとめ(続き)

- 希薄な層がある場合、又は、層間の症例数に大きな差がある場合、各治療群の最小二乗平均の解釈には特に注意が必要である。希薄な層でのわずかな外れ値が(データ全体からみるとはずれ値でなくても)、各群の最小二乗平均に大きく影響する。

なお、例えば希薄な層の一方の群に外れ値があっても、交互作用を含めないモデルで解析すれば、治療群の比較においては影響は小さそうである(まだ、完全には調べていません。ただし、希薄な層を持つ層別因子を主要な解析に含めるのは正当化できないと思います。)。

- 結果の提示にあたっては、標本平均、調整平均および連続変数の共変量の全例平均を誤解を与えないように出力する。

主な参考文献

広津千尋(1976). 分散分析. 教育出版.

SAS/STATソフトウェア ユーザーズガイド

Committee for Proprietary Medical Products (2003). Points to consider on adjustment for baseline covariates (CPMP/EWP/2863/99).

<http://www.emea.eu.int/pdfs/human/ewp/286399en.pdf>

厚生省医薬安全局審査管理課長 (1998). 「臨床試験のための統計的原則」について(平成10年11月30日医薬審第1047号).

岩崎学 (2002). 「処置前－処置後」データの解析と平均への回帰. 行動計量学, 29, 247-273.

清見文明(2003). ベースラインを共変量とした共分散分析に関する考察: 無作為化比較臨床試験での適用について. 計量生物学, 24, 95-115.

宮原英夫, 丹後俊郎 (1995). 医学統計学ハンドブック. 144-150 朝倉書店.

Senn, S. (1997). *Statistical issues in drug development*. John Wiley & Sons.

Bonate, P.L. (2000). *Analysis of Pretest-Posttest Designs*. Chapman & Hall.

Milliken, G.A. and Johnson, D.E. (2002). *Analysis of Messy Data Volume III: Analysis of Covariance*. Chapman & Hall/CRC.