

臨床試験の統計報告

統計数理研究所, 滋賀大学 折笠秀樹

健康食品の臨床試験論文を見ていて、主に統計学的観点から気付いた点を解説します。ここで挙げた10個のポイント (Box 1) に留意して、今後論文投稿をお願いしたいと思います。また、CONSORT 声明は基本中の基本です。本誌にも掲載されています。必ず熟読してから、論文を投稿してください。

1 評価項目を主・副・探索に分けて明示すること

臨床試験では、いろいろな項目を観察あるいは測定 (まとめて観測) します。その中でも、評価項目を定義する必要があります。英語では、エンドポイ

ントやアウトカムと呼んだりします。臨床試験の肝であり、もっとも主張したい変数のことです。観測項目は記載されていますが、評価項目を定義していない例が多く見られます。どれが評価項目なのかを明示しましょう。そうでなければ、その臨床試験の目標が何なのかが明瞭になりません。何が主張したかったことなのかもわかりません。

この評価項目ですが、主要 (一次やプライマリーとも呼ぶ) と副次 (二次やセカンダリーとも呼ぶ) に分けるのが一般的です。主要は原則1つ、副次は多数でも構いません。たとえば、肌の保湿効果が主要で、肌のハリ改善・油分改善・美白・しわ改善を副次と定義します。もう一つ例を挙げると、内臓脂肪を主要、体重・体脂肪・中性脂肪・腹囲を副次とすることもできるでしょう。機能として表示したい内容に即したものを主要にすべきです。

睡眠の質改善ではどうでしょうか。睡眠時間と睡眠の質を考えたとします。両者は類似しているもので、主要はいずれか1つにすべきでしょう。OSA-MA という質問票もあります。5つの次元から成り立っていますが、やはり主要は1つの次元に絞るべきでしょう。それが難しいなら、後で述べるような多重性調整が必要になります。評価項目を何にするかは、臨床試験の肝です。何でも挙げておくのではなく、必要最小限の項目に絞ることが大切です。

また、食品は複数の異なる機能を有していることがあります。たとえば、骨密度の改善と腸内環境を整えることだったりします。どちらも主張したい機

1. 評価項目を主要・副次・探索に分けて明示すること。
2. 被験者のフローチャートを示し、解析対象集団を定義すること。
3. 有効数字に留意すること。
4. P 値は正確な値を示し、不等号表示や星印は避けること。
5. 比較試験では群内検定を避けること。
6. 二元配置分散分析の交互作用検定は誤りの元なので避けること。
7. 検定の多重性に十分配慮し、主要評価項目は多重性調整すること。
8. 多時点を考慮した解析は、適切な手法を選択すること。
9. サブ解析はメイン解析論文の後に別途作成すること。
10. 著者リストは適切か、COI は網羅したかを確認すること。

Box 1 10個のポイント

能ですから、2つとも主要に設定したいと考えるでしょう。しかし、よく考えると、骨密度と腸内環境はまったく異なる内容です。興味を持つ読者も当然違いますから、それぞれ分けて投稿するのが当然でしょう。骨密度を扱う雑誌で、急に腸内環境の話が出てきたら不思議に思うでしょう。したがって、このような場合は、骨密度と腸内環境に関する2つの論文を作成すべきです。

主要と副次だけでなく、さらに下位の探索的評価項目を設けることもあります。これについては検定の P 値は示すべきではないし、 $P < 0.05$ であっても「統計的有意」と言うてはいけません。検定をするのは、あくまでも検証的評価項目（主要と副次）のみです。検証したい仮説に対して統計的検定をするのであり、探索的評価項目はまだ仮説に至っていません。いわば、仮説形成のための評価項目が探索的と言えるでしょう。

2 被験者のフローチャートを示し、解析対象集団を定義すること

臨床試験のチェックリストとして知られるCONSORT声明では、被験者のフローチャート（流れ図）を示すことになっています。試験対象者の選別から、ランダム割付け、そして追跡期間中の中止脱落により、試験対象者数が減っていきます。その流れを表すフローチャートを必ず示してください。これによって、その研究の品質も確認することができます。

主たる解析対象集団はどれなのかも明示してください。ランダム化比較試験では、例外を除き、デフォルトはITT (Intention-to-treat) です。つまり、ランダム割付けされた全員を解析対象とします。時には、途中で脱落し、評価項目のデータがまったくなければ解析を除外することもあります。コロナウイルス感染症の人を対象としていたのに、後でPCR検査をして陰性だったら解析を除外することもあります。それらはFAS (Full analysis set) と呼ばれます。修正ITT (Modified ITT) と呼ぶこともあります。このあたりにとどめておくことが大切です。PPS (Per Protocol set) と呼んで、プロトコル完全遵

守例に限定することも考えられますが、これは開発の初期探索段階で使うべきであって、世の中へ出回る前段階には合いません。世の中へ出ると、その通りに使用する人がすべてとは限らないからです。

3 有効数字に留意すること

痛みのスケールとして、VAS (Visual analog scale) やNRS (Numerical rating scale) がよく使われます。その平均値を表すとき、小数2位も示している例を見ました。そこまで必要でしょうか。元のデータがどこまで取られたかを見てください。それより1桁下で十分だと思います。VASスケールは整数だと思われるので、小数1位が適当でしょう。年齢も小数1位で十分でしょう。余計なことですが、痛みのVASとNRSについても、主要評価項目はどちらか1つに絞るべきです。

また、データ数が100例にも満たないのに、パーセントとして小数表示している例を見かけます。小数位まで表記するのは、100例を超えて初めて意味があります。100例未満で小数は通常不要であり、24%などと表記すれば十分でしょう。

P 値の桁はどうでしょうか。通常、有効数字2桁です。3桁まで示している例が見られますが、改めるようお願いします。 $P = 0.345$ なら、四捨五入して0.35です。 $P = 0.034$ なら、そのままでもいいかもしれません。 $P = 0.0034$ はどうでしょうか。あまり小さいときは小数3位が普通なので、 $P = 0.003$ とすべきです。 P が0.001よりも小さいときは ($P = 0.0004$ など)、初めて $P < 0.001$ と不等号で表記します。

4 P 値は正確な値を示し、不等号表示や星印は避けること

P 値がいくら未満をもって、統計学的に有意と判断するか、その基準を有意水準と言います。通常は5%にしますが、厳しいと1%、甘いと10%にすることもあります。その場合、なぜそうしたかの理由も書いてください。また、検定は両側が通常ですが、もし片側検定の P 値を示すときは、その理由を方法の中の統計解析の節に明記してください。

有意かそうでないかの二分ではなく、 P 値は正確な値として出ます。それを記載するのが最近の傾向です。高度有意のときのみ、 $P < 0.001$ のように不等号で表記します。 $P < 0.05$ や $P < 0.01$ といった表記は、極力避けてください。正確な P 値が 5% (0.05) 未満なら、これは統計学的に有意な結果だと読者は判断できます。

$P < 0.05$ の時に * 印、 $P < 0.01$ の時に ** 印を見ることがあります。まず、有意水準は 1 つだけにすべきです。これは有意水準を 2 つ設けたとみなされます。どうしても基準を 1% と 5% など 2 つ設けたい場合は、方法の中の統計解析の節で理由を書いてください。 P 値は不等号ではなく、しかも星印のような記号でもなく、正確な値で示すのが原則です。

一般的には検定に頼りすぎるよりは、推定をもっと活用すべきです。すなわち、 P 値を示す代わりに、95% 信頼区間を示すのです。 P 値の大小はあまり意味がありませんし、95% 信頼区間を見ると統計的有意かは一目瞭然だからです。

5 比較試験では群内検定を避けること

単群試験では前後差、つまり群内検定のみでしょう。一方、比較試験では、主たる興味は群間比較のほうです。比較試験ではどちらも実施可能ですが、群内検定は特に主張したいことではないはず。群間では有意差が付かなかつたから、被験群のみ前後変化の検定を行う例を見かけます。そのようなときは、決して群内検定は行ってはいけません。以前解説したように¹⁾、偽陽性の結果を示すことにつながり、読者の誤解を招きます。有意差をむりやり見つけてきた可能性が高く、研究者としてはあるまじき行為です。比較試験ではもともと各群における前後比較は行う必要がありません。あくまでも群間比較が主眼です。

6 二元配置分散分析の交互作用検定は誤りの元なので避けること

群と時点の 2 要因に関して二元配置分散分析を行い、その交互作用項でもって群間比較を行う例を見かけます。時点は独立ではないので、まず二元配置

分散分析自体が誤った解析手法です。反復測定分散分析を使わなくてはなりません。その時、群と時点の交互作用検定は、(0 週からの) 変化量に関する分散分析 (2 群なら t 検定) に一致します²⁾。0 週と最終時点しか使わないのなら、あえて反復測定分散分析などせずに、単純な差 (変化量) に関する群間比較をしてください。変化量を目的変数として、0 週の初期値で調整する共分散分析がよいかもしれません。

7 検定の多重性を十分配慮し、主要評価項目は多重性調整すること

主要評価項目に類似した項目を複数設ける例を見ます。群間比較の検定を繰り返せば、その中で 1 つは統計学的に有意な結果を得ることは当然です。「数打てば当たる」と言います。有意水準 5% なら、本当は差がないのに、偶然にして有意になる可能性が 5% まで許しています。偽陽性が 20 回に 1 回ある (5%) ということです。検定を繰り返せば、偽陽性の可能性が 5% 以上に高くなります。通常 $P < 0.05$ よりも厳しい基準で判定しないと、偽陽性の結果が増えてしまいます。複数の主要評価項目を置いたのなら、必ず多重性調整を行ってください。

複数の時点でも同様です。主要評価項目に関して、4 週時、8 週時、12 週時で群間比較の検定をすれば、どこかで有意な結果が得られる確率は上がります。ここでも偽陽性の可能性が高まります。たとえば、主要を 12 週時だけに絞ることが必要になります。1 つに絞らないなら、 P 値の多重性調整が必要です。 $P < 0.05$ よりも厳しい基準を採用することです。

12 週時しか取り上げていなくても、12 週時データの群間比較だけでなく、0 週から 12 週にかけての変化量も群間比較している例があります。2 つの結果は類似してきます。ここでも偽陽性が増えるので、 P 値の多重性調整が必要です。 t 検定と Wilcoxon 検定を併用するなど、多手法の時も多重性調整をすべきです。また、事前には t 検定を明言していたが、非有意だったので、Wilcoxon 検定の結果を示すということは行ってはいけません。

1) 多重検定と多重性 (Box 2)

検定を繰り返すことにより、名目の有意水準が5%よりも大きくなります。通常は20回に1回ほど偽陽性を許しているわけですが、多重検定により偽陽性の可能性はどんどん高くなります。数打てば当たるわけです。これを補正するのが多重性調整 (Adjustment for multiplicity) とされるものです。

多重性はいろいろな場面で生じます。多群あるとき、いろいろな群間比較を繰り返すと、どこかで有意差が出やすくなります。多項目比較, 多時点比較,

多手法比較については、上ですでに述べました。多部分集団比較というのは、多くのサブグループで群間比較を繰り返すと、どこかのサブグループで有意差が得られやすくなることです。サブグループは部分集団と訳すことが多いです。

2) 多項目による多重性 (Box 3, 4, 5)

評価項目を、主要、副次、探索的に分けることはすでに述べました。主要エンドポイントは1つが望ましく、そうであれば多重性の問題は生じません。2つ以上設けたときは、多重性の調整をすべきで

多重検定 (Multiple testing)

検定を繰り返すことにより、名目有意水準 (5%) が超えてしまうこと。
平均的に20回に1回 (5%) の言い過ぎどころではなくなる。
検定回数を増やしていけば、誤って有意差 (偽陽性) が得られる (数打てば当たる)。

多重性 (Multiplicity) のいろいろ

多群比較 → いろいろな群間比較を繰り返すと、どこかで有意差が得られやすい。
多項目比較 → いろいろな項目で群間比較を繰り返すと、何かの項目で有意差が得られやすい。
多時点比較 → いろいろな時点で群間比較を繰り返すと、どこかの時点で有意差が得られやすい。
多部分集団比較 → 多くのサブグループ (部分集団) で群間比較を繰り返すと、どこかのサブグループで有意差が得られやすい。
多手法比較 → いくつか手法を変えて群間比較を繰り返すと、どの手法かで有意差が得られやすい。

Box 2 多重検定と多重性

評価項目

評価項目を主要 (Primary)・副次 (Secondary)・探索的 (Exploratory) に分けること

主要エンドポイント

1つが望ましく、その際は多重性の調整は不要である。
2つ以上あるときは、多重性の調整 (Box 4 参照) をすべきである。
多重性の調整をして有意の場合は、「統計的有意」と表現してもよい。
調整しないときは、 $P < 0.05$ でも「統計的有意」とは表現しない。
ただし、そのままの (未調整) P 値を図表に示してもよい。
その代わりに、方法の章で、多重性の調整を行わないことを必ず明記する。

副次エンドポイント

複数設けることは問題ない。
 P 値をそのまま示してもよいが、「統計的有意」という表現は避ける。
統計的有意を主張したいのなら、「Gatekeeping 法」(Box 5 参照) などで多重性の調整をする。

探索的エンドポイント

複数設けることは問題ない。
 P 値をそのまま示してもよいが、「統計的有意」という表現は避ける。

Box 3 多項目による多重性

6項目について群間比較を繰り返す例

Bonferroni, Hochberg, Holm の3手法を紹介する。

P 値の例は
Cao J, et al. *JAMA* 2014; 312 (5): 543-4. より引用

P 値を大きい順に並べて判定 (Bonferroni, Hochberg)
P 値が閾値よりも小さければ有意, 大きければ非有意と判定する。
Hochberg 法では, 一旦有意になれば, それ以降はすべて有意と判定する。

Hochberg Y. *Biometrika* 1988;
75 (4): 800-2.

検定	P 値	Bonferroni		Hochberg	
		閾値	検定結果	閾値	検定結果
1	0.40	0.05/6=0.008	非有意	0.05/1=0.05	非有意
2	0.027	0.05/6=0.008	非有意	0.05/2=0.025	非有意
3	0.020	0.05/6=0.008	非有意	0.05/3=0.017	非有意
4	0.012	0.05/6=0.008	非有意	0.05/4=0.0125	有意
5	0.011	0.05/6=0.008	非有意		有意
6	0.010	0.05/6=0.008	非有意		有意

P 値を小さい順に並べて判定 (Holm)
P 値が閾値よりも小さければ有意, 大きければ非有意と判定する。
Holm 法では, 一旦非有意になれば, それ以降はすべて非有意と判定する。

Holm S. *Scand J Statistics*
1979; 6 (2): 65-70.

検定	P 値	Holm	
		閾値	検定結果
1	0.010	0.05/6=0.008	非有意
2	0.011		非有意
3	0.012		非有意
4	0.020		非有意
5	0.027		非有意
6	0.40		非有意

Hochberg 法では3つで有意と
なったが, Bonferroni 法と Holm 法
ではすべて非有意であった。

Box 4 多重性調整のための3つの手法

評価項目が階層構造になっており, 優先順位をもって順序付けする。
主要評価項目が3つあり, ①, ②, ③に順序付けられている。
副次評価項目は2種類あり, それぞれ2つずつあり, ④と⑤, ⑥と⑦は順序付けられている。

主要評価項目①, ②, ③

① P<0.05 → No なら非有意で Stop
→ Yes なら有意 → ② P<0.05 → No なら非有意で Stop
→ Yes なら有意 → ③ P<0.05 → No なら非有意で Stop
→ Yes なら有意 → 副次評価項目へ進む。
優先順位に沿って検定を行い, 非有意になったら, それ以降はすべて非有意と判定する。非有意になったら, 副次へは進まない。

副次は2種類に分かれるので, 有意水準は1/2の0.025とする。
2種類, 同時並行で検定を進める。(3種類→1/3の0.017)

副次評価項目④, ⑤

④ P<0.025 → No なら非有意で Stop
→ Yes なら有意 → ⑤ P<0.025 → No なら非有意
→ Yes なら有意

副次評価項目⑥, ⑦

⑥ P<0.025 → No なら非有意で Stop
→ Yes なら有意 → ⑦ P<0.025 → No なら非有意
→ Yes なら有意

主要が1つのときは, それが有意なら, すぐ副次へと進んでよい。

Box 5 Gatekeeping 法による多重性調整

(Yadav K, et al. *JAMA* 2017; 318 (14): 1385-6. より改変転載)

す。偽陽性が増えるからです。Box 4 に示したように、Bonferroni 法、Holm 法、Hochberg 法がよく知られています。Bonferroni 法がもっとも保守的、つまり有意になりにくいことが知られています。調整して有意なら、「統計的有意」と主張しても構いません。調整しなかったら、 $P < 0.05$ であっても統計的有意と言っははいけません。未調整 P 値を図表に示すのは許されますが、むしろ信頼区間を示すのとどめるほうが無難でしょう。また、この P 値は多重性調整をしていないので、方法の章でその旨を明記すべきです。そうでなければ読者は誤解してしまいます。

副次エンドポイントについても P 値を示すのは可能ですが、「統計的有意」という表現をすべきではありません。多重性調整をしていないので、偽陽性の結果かもしれないからです。統計的有意を主張したいのなら、先に述べた 3 つの手法 (Box 4) のどれかで調整すべきです。主張したい評価項目の優先順位が決まっているなら、Box 5 に示した Gatekeeping

法がよいかもしれません。

8 多時点を考慮した解析は、適切な手法を選択すること (Box 6)

臨床試験では 4 週、8 週、12 週というように、多時点で同じデータを収集することがあります。その際、どのような解析をすべきでしょうか。もっとも単純なのは、12 週の最終値で群間比較することです。12 週の群間比較のみの P 値を示し、その他の時点は控えるべきです。4 週や 8 週のデータは無駄かと言うと、そうとも限りません。12 週のデータが欠損のときのみ、4 週や 8 週のデータを補完して利用できます。結果の一致性を確認するための感度解析として、4 週や 8 週のデータを利用することも可能です。

通常の臨床試験では、評価指標の値が有意に改善したかどうかに関心が集まります。改善パターンが違うかなどは興味を持たれません。したがって、0 週からの差に関する分散分析 (2 群比較だと t 検定)、

多時点

主要評価項目を複数個の時点で評価する (4wk, 8wk, 12wk など)。

原則

主要な 1 時点 (通常は最終時点) だけ P 値を示し、 $P < 0.05$ なら統計的有意が主張できる。それ以外の時点では、 P 値を示すのは控えるべきである。

統計解析計画

主要評価項目の主解析は、下記のどれにするかについて事前に明記しておくべきである。

- 1) 最終時データの群間比較
- 2) 初期値からの変化量の群間比較
- 3) 初期値を共変量とした共分散分析
- 4) MMRM (Mixed-effects Models for Repeated Measures)

12 週で NS (非有意) であり、8 週で有意の場合

8 週で統計的有意とは表現すべきではない。
なぜなら 8 週は主時点ではないし、主時点の 12 週で有意ではないから。

12 週で有意であり、8 週ですでに有意の場合

12 週で統計的有意と表現するのは当然のことである。
8 週でも統計的有意を主張するには、多重性の調整が必要である。
多重性調整しないのなら、8 週ですすでに統計的有意と主張してはいけない。
8 週の時点で差が広がり、12 週で有意差を認めた、などと書く。

Box 6 多時点による多重性

あるいは共分散分析(0週データを共変量とした解析)で十分かと思えます。ただ、もっと高尚な解析手法もあります。中間のデータも考慮するMMRM(Mixed-effects models for repeated measures)と呼ばれる多変量解析です。MMRMは複雑な解析になりますが、改善パターンが異なるということを主張したいときは効果的です。

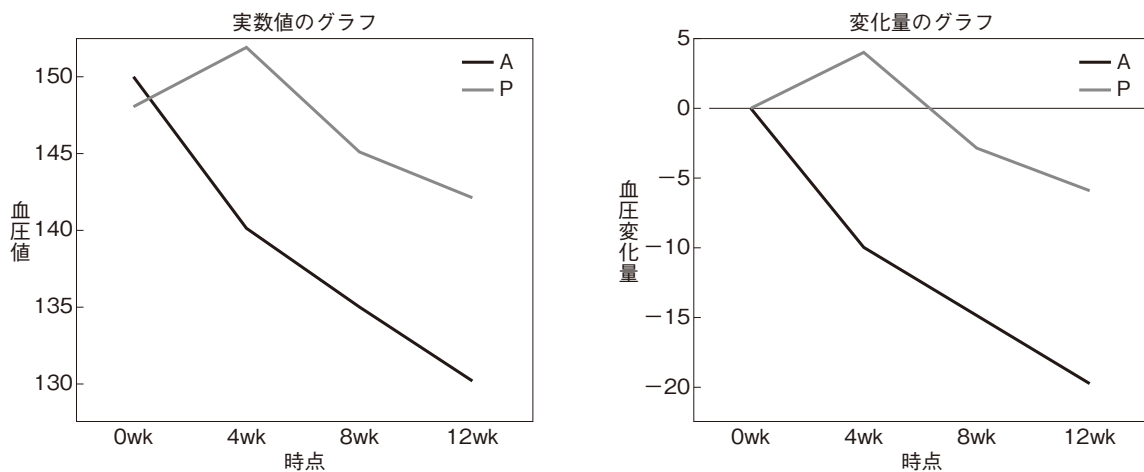
主たる時点である12週の結果が非有意であり、8週の結果が有意の場合を考えてみましょう。主の12週で非有意ですから、8週は有意だったという言うべきではありません。主が言えていないのに、副へ進むことはできません。閉手順などと呼ぶことがあります。

12週で有意であり、8週も有意であったらどうでしょうか。12週で有意と主張するのは当然のことでしょう。しかし、8週でも統計的有意かという違いがあります。複数の時点で検定を繰り返すと、いくつかの時点で有意になる可能性が増します。つまり、偽陽性が増えるのです。12週は主ですから問題あ

りませんが、その他の時点は副ですから、それは時点数が多くなるほど偽陽性が増えます。ただ、時点数に関しても多重性調整を行えば、統計的有意を主張することは可能です。このための手法も、多項目のところで紹介したのと同じものが使えます(Box 4)。仮に多重性調整をしないのなら、8週の時点で差が広がり、12週では有意差を認めたというような表現をすべきです。各時点で群間差のP値を示してもよいですが、多時点について調整は行わない旨、方法の章で明記しておきましょう。そして、統計的有意は主張しないようにしてください。

1) 評価項目推移のグラフ(Box 7, 8)

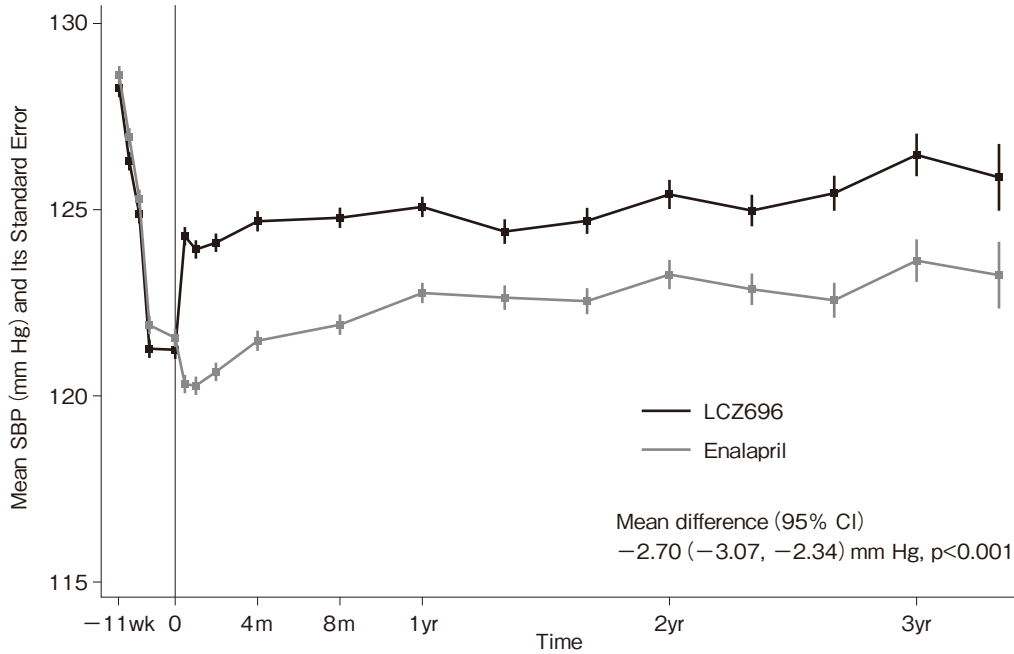
評価項目の時間推移は大変重要な情報です。それをグラフに示すことはよいことです。グラフ化には2つの方法があります。実数値で表すのと変化量で表すものです(Box 7)。初期値(0週値)に群間差があれば、変化量のほうが適切かもしれません。しかし、読者にしてみれば、初期値がいくらで、最終値はいくらという実数値に興味があるはずですから、一般的には実数値の推移グラフを作るべき



初期値に違いがあるときは変化量のグラフが良いこともあるが、原則として評価項目の数値自体のほうがわかりやすいので、実数値のグラフのほうが良い。

多時点を伴う群間比較は、最終時点の群間差、変化量の群間差、初期値を調整する共分散分析、MMRMなどの手法が考えられるので、研究実施計画書や統計解析計画書に主解析の手法を明記しておくべきである。

Box 7 評価項目の推移は実数値でグラフにしよう



Box 8 評価項目推移の図示例

(Pocock SJ, et al. *J Am Coll Cardiol* 2015; 66: 2536-49. より転載)

でしょう。日本の論文は変化量で表す例が多いようですが、海外では実数値のほうが通常のような (Box 8)。

9 サブ解析はメイン解析論文の後に別途作成すること (Box 9,10)

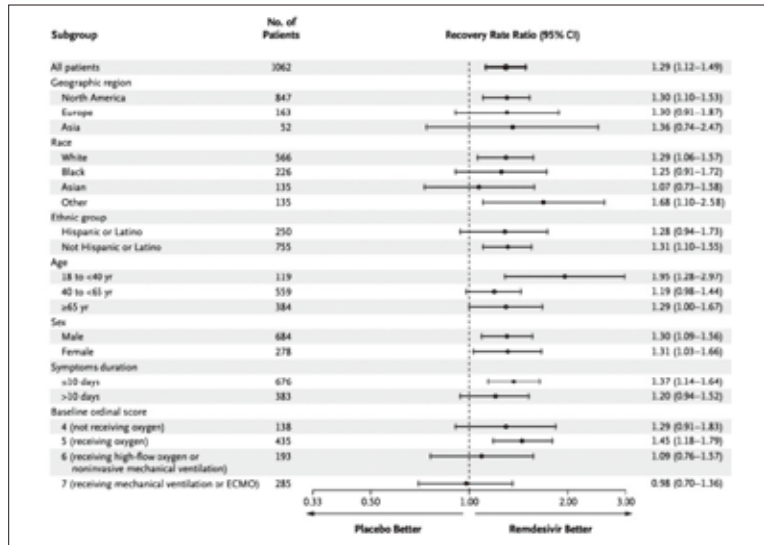
メイン解析ではもっとも主張したい仮説を立証します。その最後に、その仮説がどんなサブグループでも当てはまるかどうかを検証します。それをサブグループ解析 (Subgroup analysis) と言います。これはあくまでも、結果の一般化可能性を確認するのが目的です³⁾。有意差のみられるサブグループを探すのが目的ではありません。男女ごとのサブグループ解析では、男性でも女性でも同様の結果かを確認します。高齢者と中年者と若年者のサブグループ解析でも、年齢層により効果差がないことを確認します。サブグループ間で効果は一樣かを示すのに、交互作用検定を用いることがあります (Box 10)。この検定で非有意 (たとえば, $P > 0.2$) であれば、効果は一樣だと判定します。いろいろなサブグループ

で一樣性を確認することにより、効果はどういった人にも当てはまると理解できます。

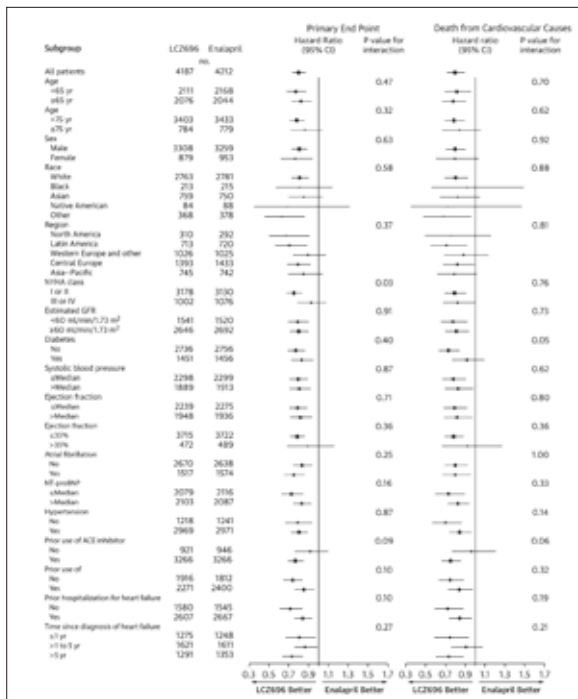
似たような用語として、層別解析 (Stratified analysis) があります。こちらは疫学研究で使われることが多い解析です。疫学では交絡がしばしば生じます。それを補正する手段として、この層別解析を用いるのです。技術的には同じことを行っていますが、その用途が異なります。臨床試験におけるサブグループ解析は結果の一般性を確認するためであり、疫学研究における層別解析は交絡補正をするためです。

時に、高齢者だけや高値の人だけのサブグループ解析を見かけます。もちろん、高齢者に限定した解析、数値が高い人だけに限定した解析は意味のあることです。しかし、メイン論文ではまずは、適格基準に合った対象者に関する全体結果を発表することです。一部の部分集団に限ったサブ解析は、目的が異なりますので、別論文でしっかり議論すべきです。

サブグループ解析に使用する変数は、事前に定義しておくべきである。
 サブグループの片面（たとえば男性）だけ示すのではなく、両面（女性も）必ず示すべきである。
 個々のサブグループの結果に P 値は示さず、この例のように 95% 信頼区間を示すほうが良い。



Box 9 主論文でサブグループ解析の結果を提示する事例
 (Beigel JH, et al. *N Engl J Med* 2020; 383: 1813-26. より転載)



サブグループの両面で効果の一致性を、交互作用検定 ($P > 0.2$ ならばほぼ一致) することも多い。

Box 10 サブグループ解析の結果を図示する例
 (Pocock SJ, et al. *J Am Coll Cardiol* 2015; 66: 2648-62. より転載)

片割れサブグループだけの結果を載せると、それは誤解をもたらします。片方で素晴らしい結果だったら、もう半分はまったくだめなはずです。いろいろなサブグループで解析してみたら、そこが強く出ているので、強調しようとしたのではないのでしょうか。これも多重性の問題であり、研究者として避けるべきことです。偽陽性かもしれないのです。サブグループ解析で有意になっても、多重性調整をしていない限り、統計的有意という言い方はしてはいけません。一般論として、片割れサブグループに限定したサブ解析は、メイン解析を出した後に別論文として書いてください。

また、骨密度に関する研究論文なのに、突如として腸内環境の結果が示される例も見かけます。その健康食品には2つの機能があるからでしょうが、それらはまったく別のテーマです。一つの論文にするのは無理があるし、読者も混乱します。内臓脂肪の論文なのに、血圧の結果が示されるのも違和感があります。ましてや、筋肉量の結果が付いていたらどうでしょうか。こうした別個のテーマについては、サブ解析と同様に、別論文として投稿すべきです。商品説明書には異なった効果を示してもよいですが、論文とは1つの仮説に対する答えを与えるものです。2つの異なる仮説であれば、それは異なる論文を書くべきです。

10 著者リストは適切か、COIは網羅したのか確認すること

その研究に貢献したと思われる人は、すべて著者に挙げるべきです。たとえばCRO (Contract research

organization, 開発業務受託機関)に委託したとしても、論文への貢献度が高いと思われる人(統計解析者など)は著者へ挙げるべきです。あえて載せなかった事例が過去にありました。COI (Conflict of interest, 利益相反)を隠したかったためです。逆に、研究費獲得以外何もしていないのに、名前の載っている例があります。これも絶対避けるべきです。その論文に対して知的貢献があった人のみを挙げるべきです。

著者のCOIについてですが、文末に必ず示してください。当該食品メーカーから謝金等を受け取ってなくても、他の同業者から受け取ってれば、それは立派なCOIです。この研究のために受け取ってなくても、その会社から何らかのお金が回ってれば、それもCOIに記載すべきです。この論文が出たら謝金を受け取ることになっているが、まだ受け取っていないから、COIは関係ないというのも変な話です。少しでも後ろ指を指されないよう、関係がありそうなものは、すべて申告しておくのが慎重な態度かと思います。なお、COI開示が必要とされるのは、論文投稿時から過去3年間とされています。

[COI] 著者(HO)は、健康食品メーカーからの謝金等は一切受け取っていない。

文献

- 1) 折笠秀樹. 群内変化の検定は誤解を招くことに注意しましょう. 薬理と治療 2020; 48 (9) : 1649-52.
- 2) 折笠秀樹. 前後比較の群間比較において、二元配置分散分析を用いた交互作用解析は誤りである. 薬理と治療 2021; 50 (8) : 1493-502.
- 3) 折笠秀樹. サブグループ解析. 薬理と治療 2020; 48 (12) : 2171-4.