論文作成における統計解析に関する留意点

富山大学大学院医学薬学研究部バイオ統計学・臨床疫学
折 笠 秀 樹

本誌編集委員の立場から，投稿論文の統計解析の記載の仕方について解説しておきたい。表の作り方など，統計解析に直接関係のない事項も含まれるが，論文作成時の留意点を周知することが目的である。

一般的事項

1. 患者背景の表では例数は縦列に記載する

群別に患者背景の数値を示す際，1行目に例数（*n*）を書いている例を見かけることがあるが，例数は縦軸の群の欄に書くのが一般的である（**図1**）。例数は*n*と小文字を使うほうが多く，本誌でも*n*を使用している。ただし，**図2**のように全体の人数を*N*と大文字で表記し，男性，女性の人数を*n*と小文字にして区別することもある。

また，測定値の推移等を図で示す場合も，**図3**に示すようにA群，B群の例数をそれぞれ図中に書くのが一般的である。

2. 図表脚注に統計手法名を書かない

図表の脚注に印をつけ，使われた統計手法名を書く例を見かけるが，どの統計手法をどういうときに使うのかについては本文の「統計解析」の節で書き，図表やその脚注には書かないことが一般的である。二つ以上の検定を用いる場合でも，「この変数では○○検定を用い，この変数では○○検定を用いた」と本文に記載しておけばよい。どうしても図表中に統計手法を記載しなければならない場合を除いて，「統計解析」を読めば，図表で使われた統計手法は何であるかがわかるような記載を望む。なお，「統計解析」の節は「方法」の章の末尾に書くのが通例である。

【悪い例】

	A群	B群
n	150	145
⋮	⋮	⋮
年齢	⋮	⋮
⋮	⋮	⋮

【良い例】

項目	A群（*n* = 150）	B群（*n* = 150）
年齢	⋮	⋮
⋮	⋮	⋮

図1　患者背景の *n* 数記載例 1

項目	A群（*N* = 150）	B群（*N* = 145）
年齢	…	…
性別男性，*n*（%）	75（50%）	58（40%）
	⋮	⋮

図2　患者背景の *n* 数記載例 2

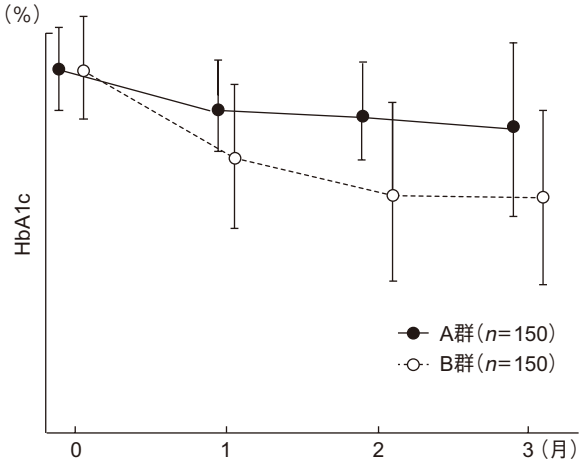

図3　図中の *n* 数記載例

3. 評価項目を記載する（RCT 論文は必須）

観察項目や測定項目は挙がっているが，評価項目を挙げていない例を見かける。臨床試験，とくにRCTでは評価項目はエンドポイントとも呼ばれる研究の骨格であり，その試験で立証したい根幹の項目のことである。それは評価項目として挙げておくようにしてもらいたい。加えて，主要評価項目と副次評価項目に分けて記載することを要望する。

4. 評価項目と統計解析は分けて記載する

評価項目ごとに統計手法を記載している例を見かけるが，両者は別々に記載してほしい。なお，探索的研究などでとくに評価項目を設けていない場合は不要である。しかしながら，臨床試験では必ず評価項目(エンドポイント)を置くことになっている。「統計解析」の節には，論文中で使用したすべての統計手法を記載する。また，有意水準(どの水準で統計学的有意と判定するか）もそこに記載する。

5. データの変動は SD，推定値の精度は SE を使う

患者背景の表では，どういった患者が組み入れられたのかを示したい。このようなデータの変動(ばらつき)を示すときにはSD(標準偏差)を使う。一方，平均値などの推定値の精度を示すときにはSE(標準誤差）を使う。一般的には患者背景ではSD，結果指標ではSEである。

平均値は取られる標本によって変動する。そのことを標本変動と呼ぶ。つまり，平均値は変動するのである。この平均値の変動を表す指標がSEであるから，SEは平均値の推定精度を表している。平均値の推移を2群で示す場合は，縦ひげとして上下にSEを付すことが多い。縦ひげとして，95％信頼区間(真値が含まれるだろう区間）を表示することもある。95％信頼区間の誤差範囲(margin of error）のことを，俗に精度(precision）と呼ぶ。

6. 有効数字をよく考えて少数桁を書く

ときに，平均値や標準偏差の数字がきわめて細かく，少数3桁も示していたりすることがある。このとき数字が細かすぎないかを考えていただきたい。有効数字はどこまでかということをもっと意識してほしい。体重を元々 kgでしか測っていないのに，平均50.15 kgとするのは少し細かすぎである。平均は元の有効数字から1桁細かくして，50.2 kgで十分だろう。

7. 考察に *P* 値は不要

「考察」の章に*P*値が記載されていることがある。*P*値とは仮説検定の結果であり，それは「結果」の章に示すべき種類のものである。特別な事情がない限り，考察には*P*値を記載しないよう気をつけたい。

8. Results のなかに手法名は書かない

英文抄録(abstract）のResultsや本文中のResultsには統計手法名を書くべきではない。統計手法名は「統計解析」の節で書くべきである。また，abstractのMethodsにも，特別なことがない限り統計手法名を挙げることはないと思われる。

9. 使用した統計ソフト名を明記する

「統計解析」の節には，使用した統計ソフト名を明記していただきたい。インターネットのサイトなどを示す場合もアドレスの明記が必要である。また，できるだけソフトのバージョンも示していただきたい。バージョンが異なるだけで解析結果が変わることもあり，また信頼できるソフトかどうかを確認することができる。

検定に関する事項

1. 検定回数は最小限にする

検定を多数実施し，*P*値がたくさん示された論文を見かける。検定には有意水準があり，それを5％に設定すると，本来は差がなくても誤って有意とする可能性が5％あることを意味する。すなわち，検定を繰り返すことによって，偶然有意差が現れるのである。有意水準を5％にすれば，20回に1回は誤って有意という結果を出している。仮に，一つの論文で100回検定を行って*P*値を示せば，何もなくとも有意な結果が五つは出ることになる。

2. 同様の項目・多時点で解析を繰り返すときは多重性を考慮する

類似した項目で検定を繰り返したり，多時点で検

定を繰り返すと，誤って有意差を出してしまうリスクが高まる。そこで，そのような場合には多重性調整(adjusting for multiplicity）をすべきである。多重性調整にはBonferroni調整，Tukey多重比較などいろいろな手法がある。適切な手法を選んで適用していただく必要がある。もし，多重性を考慮していない場合には，必ず,「なお，多重性については考慮しなかった。」という一文を,「統計解析」の節に付け加えてほしい。

3. 多重比較は分散分析で有意差を認めた後に実施する

平均値の多群比較では分散分析を使用する。たとえば3群あると，2群比較は全部で3通りできる。分散分析で3群全体に有意差が認められていないのに，中身の2群比較をしている例を見かける。全体で差がないという結論なのに，細部の比較をすべきではない。全体で差が認められれば細部の比較まで行ってもよいが，そのときは適切な多重比較の手法を用いてほしい。

4. 例数が少なく対称でないデータではデータ変換かノンパラ手法を

*t*検定を用いる前提として，①独立性，②正規性，③等分散性が知られている。例数が多くなれば，中心極限定理により平均値の標本分布は漸近正規になるためあまり気にする必要はないと思われるが，少数例(10例未満など）では下記の点を確認し，適切な手法を選択していただきたい。

独立性では，同じ人の複数データ(これは従属データと言う）が混じっていないことを確認したい。正規性は検定で点検する必要はないが，目測で対称性を確認しておきたい。歪度(skewness）という指標を利用するのもよいだろう。歪度の絶対値が1を超えるようだと非対称が疑われる。等分散性はLevene's testやBartlett's testで検定できるが，目安としてSDが2倍以上異なっていないかを見るとよいだろう。正規性や等分散性が疑われるような場合にはデータ変換(たとえば対数変換）をするか，ノンパラメトリックと呼ばれる手法に切り替えることが望まれる。

5. *P*値としてNSを避け，できるだけ直接値を示すこと

有意水準とは，*P*値がいくつ未満のときに統計学的有意と判定するかを示すものである。たとえば，$P<0.05$で統計学的有意と判定したりする。一般的には有意水準は一つにする。ときには，有意水準$P<0.05$だけでなく，別の有意水準$P<0.01$も記載していることがあるが，ダブルスタンダードのように見えて好ましくない。どうしても二つ設けたい場合は,「有意水準は$P<0.05$および$P<0.01$とした」と記載してほしい。なお，$P<0.005$など，変な有意水準は使わないほうが好ましい。

最近では，*P*値を直接値として記述する傾向にある。すなわち，$P=0.037$など有効数字2桁で記載することが多い。非有意の場合もNS(not significant)ではなく，$P=0.44$などと記載する。不等号表示は高度有意の場合(たとえば，$P<0.0001$）に限る。*P*値とは「差なし」の状況で現データの出る確率なので，$P=1$に近づくほど「差なし」に近いことがわかる。

6. *P*値の有効数字は2桁が原則

先に述べたように，最近では$P<0.05$など不等号を使うよりも，$P=0.037$のように直接値を示す方向にある。また，その際の有効数字は2桁が一般的である。あまり小さくなると，$P=0.0006$など1桁の場合もある。さらに小さく高度有意の場合は，$P<0.0001$などと不等号で示す。

7. *P*値の記載は主要なものだけに

本文中に*P*値を数多く記載している論文を見かけるが，これはあまりよろしくない。多重性の問題があるからである。主要な結果のみ*P*値を示すようにしたい。*P*値を示すことを取りやめる雑誌があるくらいであり，*P*値は「害あって利なし」という見方をする人も増えてきた。どうしても*P*値を数多く示したい場合は表の中で示せばよい。ただし，表だからといって無制限に*P*値を示すことは多重性の観点からも望ましくない。

8. よく見かける専門用語の誤り

二つの平均値の比較に*t*検定ではなく，Wilcoxon

検定というノンパラメトリックな手法を用いた論文を見かける。独立な2群比較の場合はWilcoxon順位和検定（Wilcoxon rank-sum test）であり，対応のある2群比較の場合はWilcoxon符号付き順位検定（Wilcoxon signed-rank test）が正しい。Wilcoxon符号付き順位和検定と書かれた例があるので注意したい。英語名もWilcoxon signed-ranks testなど誤った例を見かけるので，あらかじめ綴りの誤りがないことを確かめてもらいたい。

他にも，Fisherの直接確率法（Fisher's exact test）をFisherの直接確立計算法と誤った例もあった。また，two-way analysis of varianceをtwo-way of variance，Repeated measures ANOVA をRepeated measures of ANOVAと誤った例も見かけた。専門用語は誤りのないよう注意していただきたい。手元に適当な参考書籍がなければ，インターネットのWikipedia等で確認するとよいだろう。

追加事項

1. RCT論文作成時の留意事項

臨床試験のなかでももっとも重要とされるRCT（ランダム化比較試験）では，CONSORT声明に基づくことが推奨される。これは，パラレル比較試験およびクロスオーバー試験ともに当てはまる。本誌では，毎号巻末に「CONSORT 2010声明―ランダム化並行群間比較試験報告のための最新版ガイドライン―」が再掲されているので熟読されたい（初出は，本誌2010；38：939-49）。また，こうしたRCT論文の作成に際しては，下記の3点について留意していただきたい。

第一に，まったく同数に割り付けられている場合は，ランダム化の手法について記載することが望ましい。これについては，ブロック割付け，層別割付け，動的割付けなどが知られている。ブロック割付けの場合にはブロックサイズを記載することが原則である。第二に，例数設計の根拠についても記載することが望ましい。つまり，事前にどの程度の群間差を想定していたかを記載する。第三に，解析対象集団はITT（Intention-to-treat）なのか，PPS（Per protocol set）なのかを記載することが望ましい。ITTでは割り付けた全例を割り付けたとおりに解析するが，一回も投薬しなかった症例や同意撤回した症例などを除外するFAS（Full analysis set）や，modified ITTを解析対象集団と定義することもあろう。

Group	Period I	Period II
1（AB）	$\mu+\pi1+\tau1$	$\mu+\pi2+\tau2+\lambda1$
2（BA）	$\mu+\pi1+\tau2$	$\mu+\pi2+\tau1+\lambda2$

μ ＝全体平均，πi＝時期効果，τi＝治療効果（i=1 for A, =2 for B），λi＝持ち越し効果（i=1 for A followed by B, =2 for B followed by ）

図4　持ち越し効果を含めた分散分析モデル

2. クロスオーバー試験について

クロスオーバー試験では個人内での比較が可能になるため，個人差が大きい場合に有用とされるが，一方で留意しなければならないことも多い。

クロスオーバー試験ではPeriod II（第II期）へ入る前に，評価指標が元へ戻らなければならないが，元へ戻っていないことがある。このような場合は持ち越し効果（Carry-over effect）の可能性がある。薬物動態試験ではクロスオーバー試験がよく使われてきたが，そこでは評価指標である血中濃度は，薬物が消失すればすぐに元へ戻ることが自明なので問題はなかった。しかし，臨床的な評価指標では投与をやめてもすぐに元に戻るとは限らないので，十分なウォッシュアウト期間が求められる。また，持ち越し効果を含めた分散分析モデルを適用し（**図4**参照），持ち越し効果の項（λi）が非有意であることを確認する必要がある。持ち越し効果が無視できるなら，次に治療・時期の交互作用を検討すべきである。これは治療効果が時期により異なるか否かを示す。**図4**に交互作用項を入れた分散分析で確認してもよいし，簡単にPeriod I（第I期）での治療Aの効果と，Period II（第II期）での治療Aの効果に違いが見られないかを，*t*検定あるいはWilcoxon検定で確認してもよいだろう（治療Bについても同様）。こうした交互作用も無視できるなら，治療効果と時期効果の二つの項を含む分散分析モデル（**図4**でλ項を除いたモデル）で最終的に評価することが望まれる。