

看護研究のデザイン

佐伯 圭一郎 Kei-ichiro Saiki, PhD

大分県立看護科学大学 人間科学講座 健康情報科学 Oita University of Nursing and Health Sciences

2002年12月27日投稿, 2003年2月24日受理

キーワード

誤差、研究デザイン、検出力、サンプルサイズ、検定・推定

Key words

error, study design, power of test, sample size, test and estimation

1. はじめに

質の高いEvidenceを得るためには、看護研究のあらゆる点において十分な配慮が必要とされる。ここでは主に統計学の立場から、研究成果の質を低下させる各種誤差を減らすためにできること、より効果的な研究デザインを行うための方法を考える。

統計学とは、データを集計するための技術であると考えている人も多いかもしれない。しかし、統計学の知識は研究計画の段階でこそ必要とされるものである。以降では量的研究を想定し、まず研究における“誤差”の問題を整理し、研究デザインを行う際の注意点にふれ、検定の考え方を整理した後、実際の看護研究においてはあまり検討されることのないサンプルサイズ設計について整理する。

2. 誤差の問題

誤差といっても、データの不正確さを指しているわけではない。私たちが知りたい“真実”からのずれ、研究結果のばらつきなども誤差として整理することができる。研究の過程において生じる誤差を大きく区分すると、偶然誤差と非偶然誤差(バイアス)に分けられる。偶然誤差は、“偶然=確率的”に生じる誤差であり、確率の問題に対処する方法としての統計学が有効である。一方のバイアスは、その存在がわかっている場合には対処法が存在するが、そのような場合はまれであり、統計学で対処することが困難な問題である。したがって、可能な限りバイアスを取り除くように研究計画において配慮する必要がある。もう少し具体的に、研究の各段階において生じるこれらの誤差を考え、対策を整理してみよう。

(1) 測定において

例えば、個人のある検査項目を考えよう。この検査値がばらつく理由としては、本来個人における値が変動するものであることや測定自体のばらつきによる場合が考えられる。この場合には、測定を行う時刻などの条件を一定にする、測定装置の精度管理などで誤差を抑えることができる。しかし、観察やアンケートによる測定などの場合には、例えば測定者による評価に偏りが生じ、バイアスとなる場合があり、測定者のトレーニングや測定マニュアルの作成といった対処が必要となる。また、先入観によるバイアス、例えばこの患者は新しい治療法の対象となっているので痛みは少ないはずだ、といった測定者の思い込みの影響を除くには、理想的にはブラインドテストが必要とされる。しかし、看護の場面においては薬剤のように完全なブラインドは不可能である。

(2) 対象の選択において

ランダムサンプリング(無作為な標本抽出)とみなせる場合、研究対象になった集団からの結果には偶然誤差のみが存在し、統計学が対処すべき問題だけとなる。しかし、現実の研究場面では、協力の同意が得られた対象のみを研究する、重症の患者は対象としない、などランダムサンプリングとみなせない場合がほとんどである。つまり、対象の選択においては常にバイアスの可能性を秘めていることになる。

単群の記述においては、標準化や重み付けといった統計手法も存在しないわけではない。また交絡という問題に対応するマッチングや層別化といった方法もないわけではない。しかし、これらはバイアスの存在

が明らかであり、その要因を事前に研究に組み入れることができた場合にのみ利用可能な方法である。現実ランダムサンプリングが看護研究の場で行いづらいたら、われわれに可能なことは、研究対象となった集団が“どのような母集団”を代表しているのかを考え、Evidenceとしてどの範囲まで一般化できるかと考察することである。

(3) 分析・考察において

推測統計における誤りについてはあとでふれるが、研究結果の分析や考察にも誤差、特にバイアスが入り込みやすい。解釈できない結果を無視し、研究者の事前の仮説を支持する文献のみを集めるなどという態度は、程度の差こそあれ研究者が陥りやすい問題であろう。

また、研究者自身は客観的な態度であったとしても、一般的に“効果がない・関連がない”などの否定的な研究結果が広く公開される可能性は低く、文献からEvidenceを収集するものにとっては、この種のバイアス(出版バイアスと呼ぶ)の存在を考慮することも必要であろう。

3. 現実的な研究デザインとその注意点

Evidenceのレベルとして、ランダム化比較試験がもっとも望ましいものであることは明らかである。しかし、ランダム化比較試験の実施は、現実的には困難であり、したがって、注目する要因の頻度や研究に必要な期間、看護業務との整合性などを考えあわせて、できるだけ質の高いEvidenceを得られる研究のタイプを選ぶ最善の選択を行うこと考えなければならない。

例えば、新しいケアの方法と従来の方法を比較する研究において、2群を設定し同時に研究をすすめることが困難であれば、従来の方法を用いた期間のデータとその後新しい方法を用いた期間のデータを比較する前後比較のデザインが考えられよう。また、ケアの効果が短期的に観察できる場合には、同一対象に前後して両方のケアを行い、それぞれの効果を比較する方法、臨床試験でいうクロスオーバーデザインも可能な場合がある。

クロスオーバーデザインは、繰り返し起きるイベントへの介入の効果が短期間に観察できる場合には、看護研究でも利用可能である。しかし、複数の介入の割付がランダムでない、例えば必ず従来法を先に、といった場合はバイアスを生じるし、前の介入の効果が

次の介入時に持ち越されていないという保証がなければ、正しい比較は不可能である。

ここで、単純な問題だが見落とされやすい点についてふれておこう。統計学の言葉では、回帰効果として知られている問題である。例えば、血圧が高い対象を選び出し、健康教育等の介入を行った後、再度血圧を測定して、介入前後の血圧を比較するといった研究の場合には、注意する必要がある問題である。理論的な詳細は参考文献にあたってもらいたい。簡単に説明すると、前の測定では偶然高めの値を示していた人が、次の測定では値が下がる可能性が高いのに対して、前の測定で偶然低めの値を示した人は研究対象とならず(対象となっていれば前よりも高い値を示す可能性が高い)、実際は何も効果がなくとも血圧の平均値は下がってしまうという現象である。介入直後の測定では介入前より平均が下がるのに対して、それ以降は横ばいとなる結果が得られた、ということが回帰効果のみによっても観察されることになる。

対照群を置かない1群のみへの介入で、血圧のように個人内での変動が比較的大きい変量に注目している場合は、このような回帰効果の影響が考えられ、対照群を設定してきちんと比較を行う必要がある。

4. 統計学の立場から

“統計”のことはデータを集めてから統計の専門家に相談すればよいと思われているかもしれないが、研究の計画段階でこそ統計的な知識を要求され、データが集まってからは、単なる処理の作業に過ぎない。例えば、データを分析しながら統計手法を選んだり、検定の有意水準を変更したりする後知恵は、研究者によるバイアスの源になりかねないことである。統計手法の選択だけでなく、研究の検出力の検討、別の見方では対象数の決定についても、事前に検討しなければ、せっかくの研究がまったくの徒労となる場合がある。ここでは、推測統計の基本的な注意点を整理し、特に研究者の関心の高い標本数の設定について考えたい。

検定をどのように利用し、どのように考えるかという問題は、ここで論じるには大きなテーマであり、ここでは実際には同じ問題の別の見方である2つの点についてのみふれることとする。ひとつは検定結果の解釈についてしばしば見られる誤解であり、もうひとつはサンプルサイズ設計、つまり研究対象の数は十分か、少なくともはないのかという研究者の疑問に対する答えである。

(1) 仮説検定における誤解

まず、先の検定結果の解釈についての誤りとは、一般に“有意差あり・なし”という結果についての誤りであるが、“有意差あり”となった場合、“本当に差がある”と積極的に主張することは正しい。しかし、“有意差なし”となった場合、“本当に差がない”と判断してしまうことは誤りである。この場合は、“差がある”と主張できないというだけであり、後述の検出力が十分であれば、別の言い方をすれば対象数が十分に大きければ“差がある”という結果になる可能性を残している。実は、“本当に差がない”ということを検定で主張することは無理であり、十分な検出力を確保して“差があるとしてもこれより小さな差しかない”と主張することしかできないのである。

(2) 検定における2種類の誤りと検出力

解釈の問題の次に、検定があくまでも確率的な方法であるということから、どうしても生じる可能性のある検定における誤りを考えよう。通常の検定においては、“差がない・関連がない”といった帰無仮説と“差がある・関連がある”といった対立仮説のどちらをとるかという二者択一の判断を行っている。本当は差や関連がなければ、帰無仮説を受け入れることが正しく、帰無仮説を捨てて対立仮説をとることは誤りである。この誤りを第一種の誤り(αエラー)と呼ぶ。逆に、対立仮説が本当は正しいときに、帰無仮説を受け入れてしまえば、これは第二種の誤り(βエラー)となる。

表 1. 仮説検定における2種類の誤り

		本当は正しい仮説	
		帰無仮説	対立仮説
判断	帰無仮説を受容		第二種の誤り
	対立仮説を採択	第一種の誤り	

どちらの場合にせよ、誤りをおかす確率は小さいほうがよいことは当然である。第一種の誤りをおかす確率の場合はわかりやすい。これは、検定における有意水準αと同一である。例えば、 $p < 0.05$ と5%を有意水準とした検定の場合、帰無仮説が正しくとも5%の確率でそれを捨ててしまう誤りをおかすことになる。検定においては、5%は十分に小さい確率であると慣習的に判断されているが、この確率を下げるためには1%を有意水準とすることもありうる。

第二種の誤りをおかす確率は少々複雑である。第二種の誤りの大きさβは、本当は対立仮説が正しいのに、観察された結果が有意とならない確率、言い換えると誤って帰無仮説を受け入れてしまう確率である。例として2群の差の検定を考えてみよう。図1に帰無仮説のもとでの検定統計量の確率分布と対立仮説が正しい(真の差=δ)場合の分布を示した。対立仮説が正しくとも、検定で有意とならない範囲に観測された結果が入ってしまえば、検定において帰無仮説を受容してしまう。その確率が図のβの部分である。別の見方では(1-β)が、正しく有意差ありという判断を下せる確率であり、これを検出力と呼ぶ。

図からの直感的な説明となるが、検出力が高くなる(=βが小さくなる)には、(i) 真の差δが大きくなる、(ii) 分布のばらつきが小さくなる、という2つのことが影響する。しかし、真の差は、たいていの場合未知ではあるが定まった値であり、研究者に可能なことは(ii)の分布のばらつきを小さくし、βを小さくすることだけである。これは現実には対象数を増やすことに相当し、十分な検出力を確保するためにはどれくらい対象数が必要かを求めることがサンプルサイズ設計である。

(3) サンプルサイズ設計

サンプルサイズの設計は、用いる検定手法によって具体的な方法が異なるが、ここでは対応のない2群の平均値の差の検定で具体的に考えてみよう。

設計にあたり、まず研究者が設定する必要のあることは、検出に値する“意味のある差”は最低どれくらいかということと、検出力をどれくらいに定めるかということである。検出力は高いほうがよいことは明らかであるが、あまり高い値はサンプルサイズが大きくなりすぎて現実的ではなくなるため、0.8(つまり第二種の誤りをおかす確率は0.2)がよく用いられている。“意味のある差”は、統計とは別の問題であり研究者が十分検討して設定する必要がある。

ここで、対応のない2群の差の検定(有意水準αの両側検定)において、検出すべき意味のある差の最小値をδ、各群のサンプルサイズをともにnとしよう。また、標準偏差は両群ともにσとする(実際に計算する場合には文献等からσの値も仮定する必要がある)。すると、

$$\frac{d}{\sqrt{2\sigma^2/n}} = z_{\alpha/2} + z_{\beta} \quad (Z_{\alpha/2}, Z_{\beta} \text{はそれぞれ標準正規分布の上側確率が } \alpha/2, \beta \text{ となる点の値})$$

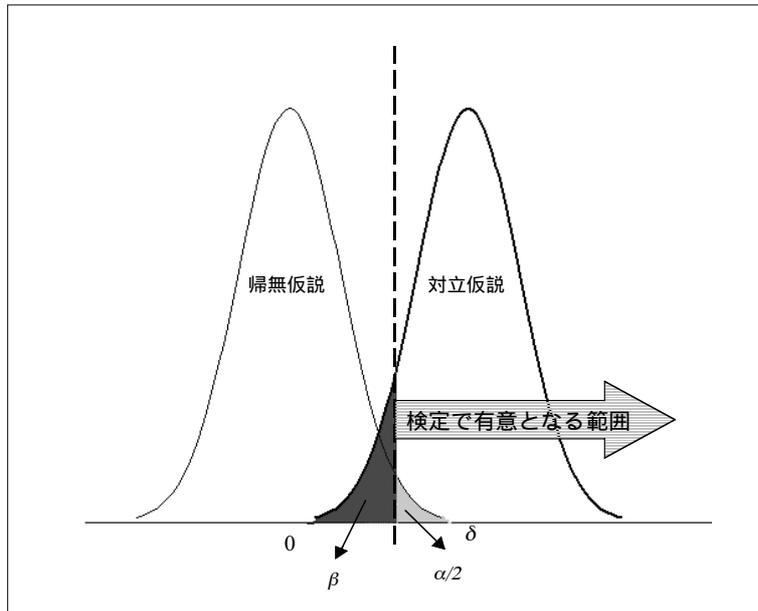


図1. 第二種の誤りをおかす確率 β

を n について解けば、検出力 $(1-\beta)$ が確保される最小の標本数が求まる。

具体的な例で考えてみよう。検出に値する最小の差を 10、有意水準 0.05、検出力 0.8 とし、標準偏差は両群ともに 20 という状況を考えて、 $Z_{\alpha/2}=1.96$ 、 $Z_{\beta}=0.81$ より

$$\frac{10}{\sqrt{\frac{2 \times 20^2}{n}}} = 1.96 + 0.81 \quad \text{を解いて、} n=62.8 \text{ となる。}$$

つまり、各群 63 人の対象が必要となるのである。

図2は、2群で割合の差の検定を有意水準5%の両側検定で行う場合を想定し、一方の群における割合を0.5に固定し、他方の群の真の割合に応じて、この割合の差を検出力0.8で検定できるための各群の必要標本数をグラフにしたものである。0.1つまり10%の差を“意味のある差”と設定した場合には 400×2 例近くの標本が必要であり、各群100例ほどの標本で検出力0.8が確保される“差”は0.2(つまり各群における割合が50%と70%のとき)以上であるということになる。

標本数がかなりたくさん必要であると感じられよ

う。実際の研究において、検出力に配慮のないまま、“有意でない”という検定結果を得ている場合が存在するものと思われる。サンプルサイズ設計の方法を使い、逆に現在の標本数では検出力はどれくらいか、という計算も可能である。十分な検出力がない標本数の場合には、本当に差がないのか、第二種の誤りをおかしているのかを考える必要がある。

(4) 区間推定のすすめ

標本数を増やすといっても実際には困難な場合が看護研究には多い。では、どのように考え、対処することが望ましいのであろうか。

ひとつには、より精度の高い測定方法を用い、研究の精度を上げるという方法が考えられる。精度の高い測定器具やよい尺度の利用が、これに該当する。また、研究デザインの工夫も可能かもしれない。例えば、単純な2群比較よりも、マッチングやクロスオーバーデザインが利用できれば、個人間のばらつきを除去できるため、少数例でも検出力を上げることができる。

これら方法が利用できない場合でも、単に有意・有意でない、といった検定結果を示すのではなく、研究結果に含まれる誤差の大きさをあわせて示す方法を

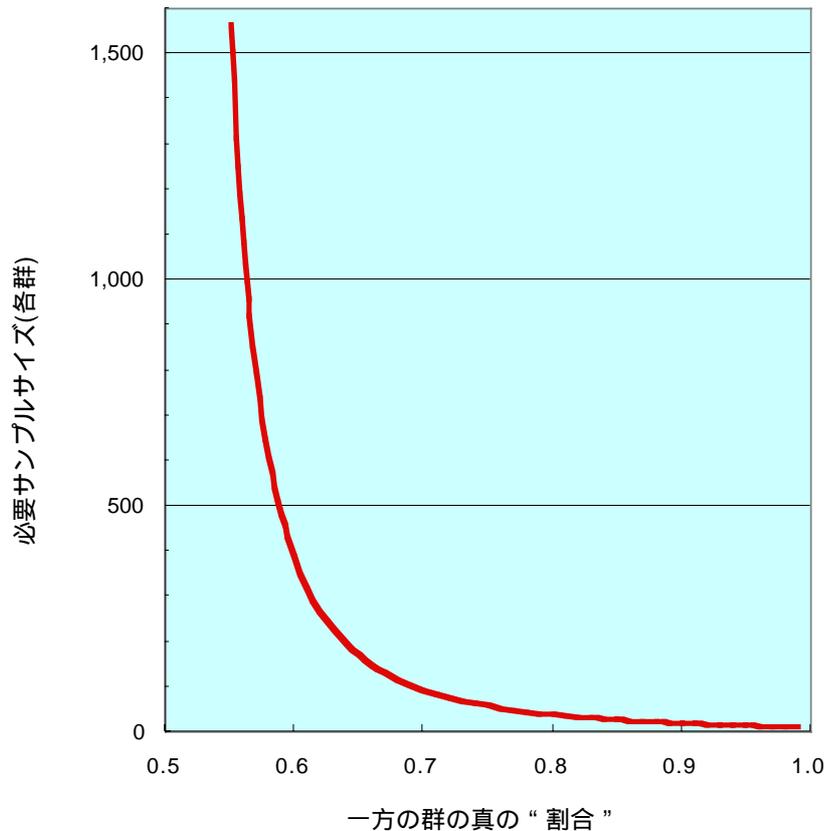


図2. 割合の差の検定で必要とされるサンプルサイズ
(一方を変化させ、他方の割合を0.5とする。検出力0.8、 α 0.05 で両側検定)

用いば、研究結果をEvidenceとして活用したい人々には有用であろう。単なる平均や割合だけではなく、それらの信頼区間を示す区間推定が、もっと活用されてしかるべきである。

5. おわりに

看護研究がEvidenceを生み出せなかったとしたら、研究者にとっての問題だけでなく、研究に協力された対象者にとっても不幸なことである。看護研究を行う者は、研究計画にも力を注ぎ、確かなEvidenceを得られる研究デザインを行う必要がある。また、単独の研究では限界があるとしても、後に続く研究者、例えばメタアナリシスを試みる者のためにも、研究デザインと結果について正確に記述し、積極的に研究成果を公開して欲しい。看護の質の向上には、絶え間ないEvidenceの積み重ねが必要とされているのだから。

著者連絡先

〒 870-1201
大分県野津原町廻栖野 2944-9
大分県立看護科学大学 健康情報科学研究室
佐伯 圭一郎
saiki@oita-nhs.ac.jp