

質疑応答

(質問)EBM に使用される統計手法になじめません。
何かよい方法はありませんか

(回答)それほどなじむ必要はありません
(和歌山県：R. M.)

(回答)名 郷 直 樹*

回答の前に

EBM を実践する上で英語の壁と並ぶ最大の敵が、質問にもある統計学の壁である。しかし、英語の壁に比べて、統計学の壁に対する処方箋は簡単だ。一言で言うと以下のようなことである。

「臨床家は統計学になじむ必要はない」

まずはっきりさせておきたいのはこのことである。臨床家は統計学になじむ必要はない。むしろ臨床家として、統計学に違和感を感じながら、個別の患者にエビデンスをどう使っていくのか、そこに集中することが重要である。

相対危険、治療必要数、有意差検定、信頼区間などの統計学的な用語がすんなり入ってくるようでは、むしろ臨床家としての考え方に問題があるのかもしれない。統計学的な考えがすんなり受け入れられ、この質問が問われないところにこそ大きな問題がある。

臨床家のリテラシーとしての統計学

なじむ必要はないとは言っても、論文の一般的な結果の解釈が統計学的にできないようではどうにもならない。なじむ必要はないが、一般的な知識として、あるいは臨床家のリテラシーとして、統計学を理解することは当然のことでもある。そこでの本稿のスタンスは明確である。質問とは若干ずれを生じるが、統計学になじむまでは行か

くても、とりあえず結果が統計学的に読める、まずはそこを目標にすることが臨床家としての第一歩である。

それでは治療に関する臨床試験論文に関し、上記のレベルをクリアするだけの最低限のポイントを解説しよう。

治療効果の評価指標

たとえば、10%の脳卒中が、治療で7%に減少したという臨床試験があったとしよう。10%と7%を比べて減っているのか減っていないのか、検討する道具立てのひとつが統計学である。

そのときにまず、割り算の指標と引き算の指標の2つを理解しておくことよ。割り算の指標の代表が、相対危険：Relative Risk (RR)，または相対危険減少：Relative Risk Reduction (RRR)，引き算の指標の代表が、治療必要数：Number Needed to Treat (NNT)である。

相対危険と相対危険減少

プラセボ群での脳卒中発症率10%、降圧薬群で7%の場合、RRはプラセボ群を基準に割り算をして、 $7/10=0.7$ と求められる。RRは1のときに差がない。1より小さければ小さいほど効果が大きい。大きければ大きいほど有害という指標である。さらにこのRRを1から引いたものをRRRという。RRが0.7の場合、RRRは $1-0.7=0.3$ 、30%となる。RRRは0のとき効果がない。0より大きければ大きいほど効果が大きく、0より小さければ有害である。

*社団法人地域医療振興協会
東京北社会保険病院臨床研修センター

治療必要数

今度は引き算の指標である。10%の脳卒中が7%に減った。今度は引いてみる。0.1-0.07=0.03, 3%の脳卒中の減少である。この3%の逆数、1/0.03=33.3, 切り上げて34をNNTという。34人治療したとき1名の脳卒中を減少させるという指標である。

表1に指標の計算例を挙げる。

統計学的有意差と臨床的有意差

臨床試験の結果は、RRRが30%で、p値が0.003と0.05未満であるため、統計学的に有意差があるなどと表現される。p値を眺め、0.05と比べることによって判定するのだが、これを紋切り型に適用すると、0.051なら差がなく、0.049なら差がある、というようなおかしなことになってしまう。あるいは、年率0.5%と0.33%の心筋梗塞を比べて、p値が0.05未満で有意差あり、なんていう論文もある。これも臨床家として受け入れがたい面がある。

しかし、これが受け入れがたいというのは当然のことで、統計学的な有意差はあくまで統計学的有意差であって、臨床的有意差とは違うのである。ここでも重要なことは有意差検定を鵜呑みにするなということである。臨床家として、統計学的有意差と臨床的有意差を分けて考えるということは、統計学の適用に当たって重要なことであるが、臨床的有意差を考える上でp値は使いにくい。そこで有用なのが推定、信頼区間である。

統計学的推定と信頼区間

有意差の有り無しというデジタルな判断をする検定に対して、推定は、差がある/ないという判定はしない。30%減少したという結果から治療効果を小さく見積もると何%、大きく見積もると何%と、ある幅を持って表示する。その幅のことを信頼区間という。相対危険減少30%(15~45)というように示される。結果は30%減少であるが、効果を小さく見積もると15%、大きく見積もると45%というわけである。通常は有意水準0.05に対応して、95%信頼区間が用いられる。真の値が95%の確率で存在する範囲が95%信頼区間である。

この95%信頼区間は、差のあり/なしという明確な判定がされない反面、臨床的にはむしろ使いやすい表現である。特に以下の2点に注目しておく。信頼区間がより有用に感じられる。

- ・検定で有意差なしの場合、最も治療効果が大きい場合にも注目
- ・検定で有意差ありの場合、最も治療効果が小さい場合にも注目

例を挙げよう。相対危険減少20%(-10~42), $p > 0.05$ という論文があったとしよう。pが0.05より大きく統計学的に有意差はない。ただ95%信頼区間の上限で見ると42%減少させるという効果があるかもしれない(もちろん10%増加させる危険があるわけだが)、そう読める。逆に、相対危険減少20%(7~35), $p = 0.02$ というような論文があったとしよう。pが0.05未満であるから統計学的有意

表1 各指標の計算結果一覧

CER	EER	RR EER/CER	RRR (CER-EER)/CER	NNT 1/(CER-EER)
0.1	0.07	0.7	0.3	34
0.01	0.007	0.7	0.3	334

CER: 対照群での発症率, EER: 治療群での発症率

RR: 相対危険, RRR: 相対危険減少, NNT: 治療必要数

表2 結果の読み方(相対危険減少の場合)

1. 相対危険減少の数字のチェック
2. 95%信頼区間の上限下限のチェック
3. 信頼区間が0を含まない場合(有意差ありの場合)
 - (ア) 治療効果を多めに見積もった際の結果: 信頼区間の上限にも注目
4. 信頼区間が0を含む場合(有意差がない場合)
 - (ア) 治療効果を少なめに見積もった際の結果: 信頼区間の下限にも注目

*相対危険の場合0を1に, 治療必要数の場合0を無限大に置き換えればよい。

差ありであるが、95%信頼区間の下限で見ると7%減少させるに過ぎないかもしれない。そんな幅をもった判断ができ、臨床的な判断につなげやすい。

治療に関する臨床試験の成績の読み方を、相対危険、相対危険減少、治療必要数、統計学的検定、信頼区間の解釈の仕方に限って解説した(表2)。最後にもう一度結論を繰り返しておこう。

「統計になじむ必要はない。臨床的な意味を考える道具のひとつとして一般的な解釈ができればよい」

文 献

- 1) 名郷直樹: ステップアップ EBM 実践ワークブック. 2009. 南江堂.