パネルディスカッション「めまい治療の Evidence を求めて」

臨床医学でのエビデンス

深谷 卓

Medical Statistics

Takashi Fukaya

Department of Otolaryngology, Kanto Medical center

To describe methods and strategies to advance the basic science underlying all interventions in ENT, this paper reviewed some areas of medical statistics that have gained prominence over the last 5-10 years: meta-analysis, evidence-based medicine, and cluster randomized trials. This study considers several issues relating to data analysis and interpretation, many relating to the use and misuse of hypothesis testing, drawing on recent reviews of the use of statistics in medical journals.

Key words: EBM, medical statistics

はじめに

最近流行している用語の一つに「EBM に基づく治療指針」があるが、これほど自己矛盾した用語はない。EBM とは提唱者の Sackett 教授が語るように「現今の最良の証拠を、良心的、明示的、そして妥当性のある用い方をして、個々の患者の臨床決断をくだすこと」であり、眼前の患者にたいする治療、もっと広くすべてのケア介入の選択を根拠に基づき行うことである。Sackett 教授はまた「患者にとって最良の決断には、患者の既往症、経済状況、遺伝情報、家族関係や、医師の経験、技術などを考慮することは不可欠」と述べている¹⁾。つまりケア介入の選択には、医師の経験、技術などの技量と、患者の背景と特質、そしてエビデンスの3つが調和することが重要になるわけで、EBM は本質的に Bottom Up な思考であり、決して治療指針といった Top Down な思考ではない。ところが、日本ではエビデンスだけが一人歩きし、「EBM に基づく治療指針」な

どという理解困難な標語となっている。

筆者はエビデンスや治療指針を軽視したり、否定したりするものではない。逆に今までの医学では、ケア介入の決定が医療者の経験や病態生理からなされ、その有用性の評価がおろそかにされていることに同意するものである。しかし、EBM を治療指針に置き換えることには同意できない。そこをはっきりさせてから、エビデンスに関する問題点を考えてみる。

エビデンスをもとめて

エビデンスには「作る、知る、利用する」の異なった3つの段階がある²⁾。この章では、主に「作る」ことを取り上げる。まず臨床医学では統計が必須であることを理解しなければならない。臨床医学では、医療介入（手術、投薬などすべての介入を意味する）の評価はその有用性だけである。表1に示すように理学・工学などの科学実験は法則の発見を目的とし、命題は明瞭なので系の制御は容易で、未知因子の入り込むことも少ない。誤差はないことが望まれる。一方、臨床医学では事情は一変し、誤差は存在することが前提であり、

表 1 臨床医学と科学実験の差異

	科学実験	臨床医学
目的	法則発見	有用性の評価
制御	簡単	困難
未知因子	排除	ランダム化
誤差	無であるべき	存在を前提

未知因子も排除できない。この事情は例えば、スタチンのようなコレステロール降下剤の有用性を検討するとき、Endpoint を脳梗塞、心筋梗塞の死亡率低下にとってみればよくわかる。Endpoint は必ずしもコレステロールだけでは説明がつかない（系の制御が困難）。またコレステロールを低下させたとき、アテローム変化が減っても、出血性要因など別の未知因子が混入してくる。しかも対象の取り方や、数で結果もさまざまな値となる。臨床医学では有用性評価には統計が必須であることが理解いただけると思う。

臨床医学に必要なもの

ではエビデンスを作るときに、臨床医学に必要なものは何であろうか。必要な条件は表 2 に示す 3 つ、精密度、正確度、一般性である³⁾⁴⁾。この

表 2 臨床医学の統計に必要な要素

1. 精密度：ランダム誤差を少なくする
2. 正確度：バイアスを減らす(内的妥当性)
3. 一般性：より広い対象を扱う(外的妥当性)

図 1 精密度と正確度の関係

うちの精密度、正確度の関係を図 1 に示す。A はランダム誤差も少なく、真の値とも一致しており精密性、正確性とも良好である。B は正確ではあるが精密度に欠ける。C は誤差は少ないが正確度に欠ける。D は正確度、精密度ともに欠けている。

精密度を確保するにはどうしたら良いのか。ランダム誤差を少なくするには観測数を多くすることである。表 3 に示すように、95%信頼区間で検討してみると、同じ20%でも観測数が違うと意味が違うことがわかる。正確度を向上させるには無作為化 (randomization) しかない。一般化はメタアナリシスにより達成される。メタアナリシスとは治療法の効果を評価する目的で、同一目的で実施した複数の研究成績を併合する統計手段であり、利点として

1. 対象数を増やし検出力を高める
2. 結果の再現性を検証する
3. 個々の試験では検証できないサブグループ解析が可能となる

などがあげられるが、逆に対象を自分の結論に都合良いものに絞るなどの出版バイアスの危険性がある。以上見てきたように、臨床医学では無作為化され、観測数が多く（信頼区間が狭い）、メタアナリシスされたものほど信頼性が高い。さまざまな機関が論文の価値を評価してレベルを表示している。表 4 はケンブリッジ大学の評価基準である。これを見ると専門家の意見とくに生理学に

表 3 精密度

観測数	95%信頼区間
5	0.005-0.72
20	0.06-0.72
100	0.12-0.28
1000	0.18-0.22

表 4 論文の評価基準

推奨度	レベル	種類
A	1a	RCT、メタアナライシス
	1b	個々の RCT(狭い信頼区間)
	1c	All or none
B	2a	コホート研究のメタアナライシス
	2b	個々のコホート研究
	2c	アウトカム研究
	3a	症例、対象研究のメタアナライシス
	3b	個々のケースコントロール研究
C	4	ケースシリーズ
D	5	批判的吟味をもたない、生理学に基づく、基礎実験に基づく専門家の意見

基づく意見がいかに低い評価かがわかる。

ま と め

論文を評価するとき、評価基準は RCT、観測数、メタアナライシスの有無、さらには前向き試験かどうか、バイアスの避け方、などに注意を払う必要がある。またあるトライアルでの結論も、観測数を多くした施行が出てくると、結果が逆転する可能性もある。つねに、最新の評価を入手し、個々の症例の治療計画に利用していきたいものである。

文 献

1) Sackett DL, Straus S, Richardson S, et al:
Evidence-based medicine: How to Practice

and Teach EBM. Churchill Livingstone,
London, 2000

2) 大橋靖雄：臨床試験と生物統計学，そして
EBM. からだの科学 208: 80-88, 1999
3) Colton T: Statistics in Medicine. Brown and
Co, Boston Little, 1974
4) Fletcher RH, Fletcher SW, Wagner EH:
Clinical Epidemiology. Williams & Wilkins,
London, 1996

(原稿到着：平成14年 4 月23日
別刷請求先：深谷 卓
〒141-0022 東京都品川区東五反田5-9-22
NTT 東日本関東病院耳鼻咽喉科)