



呼ぶ人も多くなっています。

脳血管障害は、従来わが国では最も多い痴呆の原因であるといわれてきました。最近では欧米と同じく、アルツハイマー病のほうが多いという研究が多くなってきましたが、いまだに重要な痴呆の原因であることは変わりありません。

脳卒中から痴呆に至る過程にはさまざまな機序があり、それぞれにさまざまな経過、臨床像が異なるため、血管性痴呆は均一な疾患名というより、複数の

症候群の総称と考えるべきです。

次に代表的な3つの血管性痴呆の形を示します。

(1)大脳皮質を冒す単一もしくは複数の病巣による痴呆

前述のように、大脳皮質が傷害を受けると、その部位に応じた特定の認知機能が障害されます。その範囲が大きかったり複数の領域が傷害されると、それに応じた複数の認知機能が障害され、その組み合わせで定義的には痴呆と呼べる状態になります。

す。痴呆の症状は障害される領域により異なり、一定のパターンを示しません。

通常、脳卒中発作に引き続いて発症します。あるいは、初回発作で急激に生じた症状がある程度固定した後、再発作で再び新たな症状が加わるという、いわゆる階段状増悪の経過をとります。脳に行く大きな動脈の

動脈硬化による「血栓症」、あるいは心臓に生じた血栓が脳に行って動脈に詰まる「塞栓症」によって生じます。CT(コンピュータ断層撮影)やMRI(磁気共鳴画像)により大脳の病巣が描出され、症状を分析すれば病巣部位と対応した認知障害があることがわかります。

(2)大脳皮質機能に影響を及ぼす大脳皮質下の小病巣による痴呆

大脳皮質下の小病巣によって、大脳中心部にある「大脳基底核」や「視床」と呼ばれる神経核と大脳皮質を結ぶ経路が絶たれると、2次性に大脳皮質の

機能障害が生じ、痴呆や認知障害をもたらします。主として高血圧が原因の細い動脈の狭窄や閉塞で、「ラクナ梗塞」といわれる小さな脳梗塞によって起ります。

多くは多発性です。脳卒中発作として急性に発症することもあります。なかには発作がはっきりしないまま、いつの間にか病変が生じていることもあります。

突然に起こる痴呆が脳卒中の発作そのものであることもあり、病変の部位によっては直径1cmにも満たないような脳梗塞で痴呆が生じてしまうこともあります。SPECTやPETなどをを用いて脳血流や脳代謝を調べると、主として前頭葉に血流や代謝の低下が示されます。症状は場所により異なりますが、発動性の低下や無為、あるいは逆に脱抑制を中心とした前頭葉障害の症状を呈することが多く、記憶回路が損傷に含まれる

と記憶障害も起こってきます。しばしば運動系の神経経路も同時に障害していて、仮性球まひ、錐体外路症状、片まひなど神経症状を伴います。

大脳白質の病変による

痴呆

CTやMRIの進歩に伴って大脳白質の異常がよく発見されるようになりました。この所見は高齢者に比較的多く見られ、また高血圧と関係しています。慢性的な脳の虚血により、大脳白質がびまん性に障害されたためと考えられています。多くは痴呆とは関係がありませんが、そのようなものの中に「ピンスワンガー病」という進行性の痴呆が生じる病態があります。ゆつくりとあるいは階段状に増悪する痴呆や無為・無関心などの精神症状が見られ、仮性球まひ・錐体路症候・錐体外路症候とよばれる神経症候、歩行障害、

失禁などの神経学的障害を伴います。

痴呆と紛らわしい 認知機能の障害

一方、大脳の一部が冒されたことによる特定の認知機能の障害は、医学的には痴呆と区別されませんが、一見痴呆と紛らわしいものです。以下に、とくに紛らわしいものをあげてみます。

(1)失語症

大脳の損傷に由来する、いったん獲得された言語機能の障害を失語といいます。症状特徴の組み合わせによっていくつかの臨床型に分類することができますが、とくに痴呆と紛らわしいのがウエルニッケ失語と呼ばれる失語症です。左側頭葉後上部(ウエルニッケ領域)を中心にした領域の損傷で生じます。通常、まひは伴いません。聞いたことばの理解障害と、しゃべろうとしたことばが別の単語や音に入れ替わってしまう(錯語)

のが特徴です。重症の場合には、ことばの理解がまったくできず、発することばも意味不明(ジャーゴン)になってしまいます。そのほか、ことばの理解は比較的ましで、ことばを発することが著しく困難なプロカ失語や、文字を読む能力(失読)、書く能力(失書)、計算力(失算)が選択的に障害されるような病態も存在します。

(2)右半球症状群

大脳右半球のある程度広範な領域に損傷が生じると、外空間や自分の体の左半分に注意がとどかない、左半側空間無視、左半側身体失認という障害が生じます。そのため左側に置かれたものに気づかないことが起こります。また左片まひがあっても、その存在を否認したり無視したりします。同時に全般的に注意が障害されて、急性期に「せん妄」が見られることがしばしばあり、慢性期にも残ることがあります。そのため注意がそれや

すく集中力が続かず、病識がなく多幸的な状態を示すことがあります。

(3)失行

まひはないのに、これまでよく慣れている動作がうまくできなくなり、例えば、パイパイという動作ができない、道具をうまく使えない、一連の行為を順序だててできないなどということが起こります。これらは「失行」と呼ばれ、左大脳半球の頭頂葉の損傷で生じます。

また服をうまく着ることができなくなるものを着衣失行と呼びますが、これは右大脳半球の頭頂葉の損傷で生じます。絵の模写や積み木など、まとまりのある形を作ることができない症候を構成障害と呼びますが、部分を空間的に配置する行為能力の障害です。左右どちらの頭頂葉の病巣でも生じます。

(4)せん妄

せん妄は、痴呆と最も混同されやすい病態です。せん妄は基

本的に注意機能の障害であると理解されます。注意機能は人間の認知機能の基盤となるものであり、注意機能の低下によって

広範囲の認知機能が障害されます。選択性や持続性が低下し転導性が亢進すると、患者は五感から入力される周囲の刺激にすぐに対応してしまい、一貫した行動や思考が実現できなくなります。

せん妄の原因疾患として最も多いのは代謝障害ですが、右大脳半球の広範な損傷、後頭葉の損傷、前頭葉を冒す脳卒中の急性期にもしばしば生じてきます。急性期をすぎると注意障害

は改善し、せん妄は消失するものが一般的ですが、中にはかなり長く続くものもあります。

脳卒中の後の痴呆

脳卒中発作の急性期に最も障害が強く、次第に回復して行くのが普通です。認知障害もそのような経過をとります。

しかし脳卒中発作の後にいったん障害が回復し、その後には痴呆が生じてくることもよくあります。そのような場合は、脳卒中の再発が考えられます。脳卒中発作を繰り返せば繰り返すほど、認知機能に関連した脳部位

が損傷を受ける率が高くなり、そのため認知障害の範囲や程度が増して痴呆を生じたり、さらに痴呆が悪化します。したがって、そのようなことを防ぐには

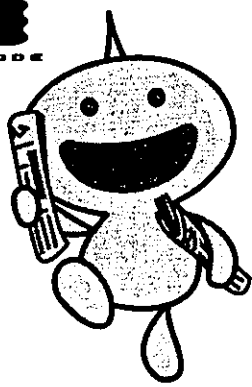
脳卒中の再発の予防が重要です。高血圧、喫煙、心疾患、糖尿病などの脳卒中の危険因子に對して、適切な医学的対応が必要で、適切な運動や食生活の改善など養生も重要です。

しかし、痴呆の原因は脳卒中の再発ばかりではありません。脳卒中後に用いられている睡眠導入剤や精神安定剤、その他の中枢神経に作用する薬剤が、認知障害や痴呆の原因となつてい

ることもしばしばあります。

運動まひなどがあって寝たきりになったような場合には、廃用症候群も認知障害悪化の理由になります。十分な刺激がなくなり、また意欲や関心もなくなっていることが問題です。またアルツハイマー病の合併もしばしば考えられます。脳卒中後に生じる痴呆の3分の1はアルツハイマー病の合併であったという研究報告も見られます。脳卒中がアルツハイマー病を引き起こすことはありませんが、その症状を顕在化させたり、悪化させたりすることは十分にありえます。

最新のニュースと 今ドキの話題は、 ボクらにおまかせ



■モードで最新ニュースを選ぶあなたの分身

マイクローン

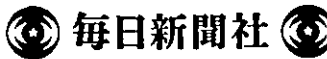
1モード公式サイト:月額200円
※パケット通信料は別途がかかります。

今すぐCheck!

Menu ▶メニューリスト ▶ ニュース / 天気 / 情報
▶ 毎日新聞 ▶ スポニチ ▶ マイクローン

ホームページ

▶ <http://www.myclone.tv>



※1モードはNTTドコモの登録商標です。
©毎日新聞社/テラネットプロダクト/オーパーワークス

“もの忘れ外来”のための予備知識：MCIの予後

松本 圭司* 森 悦朗*

抄録：高齢者の軽度認知機能障害を表す概念のうち、特に mild cognitive impairment (MCI) は、近年アルツハイマー病 (AD) の早期診断と治療についての縦断的研究が進められており注目を浴びている。MCIには標準化された診断基準はなく、研究者により定義や検査内容が違ふことに注意が必要である。Petersen の amnesic MCI の AD への移行率は年11~14%と報告されている。その一方で MCI の罹症率や経時的な安定性、および AD への進行の予測については、age-associated cognitive decline (AACD) より劣るとの縦断的研究もある。MCI の AD への進行の予測因子として Apo E epsilon 4 allele や MRI での海馬萎縮の報告があり、髄液 β -amyloid 42 や Tau の併用の有効性も期待されているが十分なエビデンスはない。今後、痴呆の早期診断に有用な客観的指標のエビデンスを踏まえ、より操作的な診断基準が確立されることが必要である。精神科治療学 17(3); 307-310, 2002

Key words : mild cognitive impairment (MCI), prognosis

はじめに

高齢者にみられる軽度の認知障害は正常加齢によるものとみなされていたが、近年痴呆性疾患、特にアルツハイマー病 (AD) の初期であることが少なくないことが指摘されている。軽度認知障害があるものの痴呆に至っていない状態は特に早期介入の点から関心が向けられている。本稿では AD の早期診断と治療の側面から縦断的研究が進められている mild cognitive impairment (MCI) を中心に高齢者の軽度認知障害の概念とその予後についてまとめる。

I. 軽度認知機能障害の概念について

正常の老化に伴う認知機能の低下と正常域を越えた状態の間の境界を表す概念として、従来より種々な用語が提唱されている¹⁾。このうち benign senescent forgetfulness, age-associated memory impairment (AAMI), age-associated cognitive decline (AACD), age-related cognitive decline 等は加齢に伴う正常範囲の認知機能の低下を意図する概念である。AAMI では、①50歳以上、②記憶障害の訴え、③標準化された近時記憶の検査で若年成人の平均より 1 SD 以上の低下、④ Wechsler Adult Intelligence Scale - Revised (WAIS-R) の単語問題の成績が 9 以上の知能が保たれる、⑤ Mini-Mental State Examination (MMSE) で 24 点以上で示される痴呆の欠如、という診断基準が初めて用いられた。その後 Levy らは AAMI の概念を注意、記憶、言語、思考、学

A prognosis of mild cognitive impairment.

*兵庫県立高齢者脳機能研究センター臨床研究科
〔〒670-0981 兵庫県姫路市西庄甲520〕

Keiji Matsumoto, M.D., Etsuro Mori, M.D., Ph.D.: Clinical Neurosciences, Hyogo Institute for Aging Brain and Cognitive Disorders, 520 Saisho-ko, Himeji-shi, Hyogo, 679-0981 Japan.

習, 視空間認知など多領域の認知障害に拡張し, かつ同年齢の平均に対して1SD以上の低下を示すものをAACDとして提唱した¹¹⁾。一方痴呆性疾患の前駆状態を意識した概念として mild cognitive impairment (MCI) や mild neurocognitive disorder 等が提唱された。MCIには標準化された診断基準はなく, 研究者により認知障害の定義や検査内容が違ふ。例えばこの用語の初出である, Flickerの論文¹²⁾では, Global Deterioration Scale (GDS)の grade 3, すなわち①知らないところで迷う, ②仕事の能力低下, ③言葉や名前の想起障害, ④読んだ内容を覚えていない, ⑤新しい人の名前を覚えにくい, ⑥貴重品をなくしたり置場所を間違える, ⑦集中力欠如, の7項目中少なくとも2つを示し, 検査上は記憶以外に言語機能, 概念形成, 視空間構成で低下がみられたものを意味する。一方 Petersen は, 認知機能, 特に記憶に関する訴えがあり, Wechsler Memory Scale-Revised (WMS-R) や Auditory Verbal Learning Test (AVLT) 等の記憶検査では同年齢者に対し1.5SD以上の低下があり, MMSE や Short Test of Mental Status 等の痴呆のスクリーニング検査や WAIS-R や Mattis Dementia Rating Scale (DRS) 等の全般的認知機能検査では正常範囲で, 日常生活能力は保たれ Clinical Dementia Rating (CDR) で0.5に相当するとしていたが⁷⁾, 最近では記憶単独の障害であることを明記し, 認知テストの具体的な基準は外している⁹⁾。しかし認知障害の範囲を記憶障害のみに限定すると mild cognitive impairment という用語は不適切との指摘¹¹⁾があり, Petersen は MCI の概念の heterogeneity について, 記憶以外の認知障害を伴わない MCI は AD の前駆状態として多いパターンであり, 一方痴呆の域には達しない軽度の多領域の認知機能障害がみられるものや, 記憶以外の認知機能の単独の障害がみられる場合は別のタイプの痴呆へ進展する可能性があるとし, 概念を明確にするため最初のタイプを amnesic MCI と記述することを提唱している¹⁰⁾。amnesic MCI の定義¹⁰⁾は, ①記憶障害の訴えがあり, それが情報提供者により証言されることが望ましい, ②年齢, 学歴に比べて記憶障害がある, ③全般的な認知機能は保たれる, ④日常生活動作は保たれる, ⑤痴

呆でない, としている。

II. 軽度認知障害の予後

軽度認知障害の予後についての縦断的研究を解釈する際には, 対象となった母集団の背景や規模, 年齢構成に加え, 用いられた診断基準や検査の違いに留意しなければならない。正常高齢者の AD への移行率は70~74歳で年0.51%, 85~89歳で3.9%なのに対し, 軽度認知障害では年6~25%と高率である¹⁰⁾。このうち amnesic MCI に一致する研究結果を列挙すると, Mayo Clinic で66人の MCI (平均81歳) を4年間経過観察したところ1年あたり12%⁸⁾, 6年目には約80%が AD に進行した⁹⁾, 記憶の問題のために家庭医を受診した, 日常生活障害が3ヵ月以上持続し (Global Deterioration Scale で stage 2 か 3), MMSE で24以上か DRS で123以上で, DMS III-R の痴呆の診断基準を満たさない107人(平均74歳)で2年間で29人(27%, 1年あたり14%)が AD に進行した¹³⁾, HMO (health maintenance organization) の登録者のうち認知機能低下を訴えた811人中, 21人(平均74歳)に記憶障害のみの認知障害が認められ, 平均4年間の経過で9人(43%, 1年あたり11%)が AD と診断された¹⁾, 等の報告がある。一方, 認知機能障害を日常生活機能スケールで定義した研究をみると, 広告募集による CDR 0.5 (questionable AD) の123人(平均72歳)では, 3年後に23名(18%, 1年あたり6%)が probable AD と診断された³⁾。また New York 大学の Aging Dementia Research Center で32人の GDS grade 3 (平均71歳) を2.2年経過観察したところ, 16人が probable AD と診断された(50%, 1年あたり25%)⁴⁾。これらの研究はいずれも病院や診療所レベル等であり大規模なものではなかったが, 最近南フランスの一般集団を対象にした正常高齢者の認知機能のコホート研究で, 60歳以上の痴呆でない833人に総合的な心理検査を施行し MCI と AACD の診断を行った¹²⁾。MCI と診断されたのは開始時, 2, 3年目で3.24%, 6.8%, 8.5%, AACD は20.9%, 48.7%, 56.8%であったが, このうち1年目に MCI とされ2年目にも同診断であったのは11.7%, 2年目に

MCIとされ3年目に同診断であったのは17.4%であり、AACDではそれぞれ56.3%、59.4%であった。また3年目に痴呆と診断されたのは1年目のMCIで0%、2年目で11.1%に対し、AACDでは17.5%と28.6%であった。この結果より、MCIはAACDに比べ有症率が低く、経時的安定性に欠け、かつ痴呆発症の予後推測において劣ると結論している。

Ⅲ. 予後に関わる因子

軽度認知機能障害のADへの変換の予測因子については縦断研究で検討されている。Apo E epsilon 4 alleleに関しては、強い予測因子である⁷⁾、記憶検査が含まれた場合には予測因子である¹⁴⁾、予測因子でない⁹⁾という報告があり一定していない。画像検査では、MRIでの海馬萎縮⁶⁾が有意な予測因子であるとされる。心理検査ではAVLTの遅延再生とWMSの精神統制¹³⁾、simple reaction time、同時注意課題における反応時間、意味カテゴリーの語想起、cueによる遅延言語記憶再生、顔と名前の組の記憶、物語の再生、複雑図形の模写¹²⁾が有意な予測因子であったという報告があるが、アメリカ神経学会のQuality Standards Subcommitteeの報告では発症前の痴呆の診断に有効な心理検査に関して十分なエビデンスは得られていないと結論している⁹⁾。また髄液中の β -amyloid 42の低値やTauの高値がADで報告されており、これらの組み合わせがMCIの予後予測に有用である可能性が示唆されている⁹⁾がやはり十分なエビデンスはない。

Ⅳ. 今後の課題

MCIの概念は早期介入の効果への期待あるいは抗痴呆薬の対象拡大という要請に促され形成されてきた面がある。しかし倫理面からは、痴呆の根本的な予防や治療法がない現在、少なくとも2、3年はその大半が痴呆を発症しないにもかかわらず、MCIと診断することの心理的、社会的影響に配慮すべきとの意見がある。少なくとも認知機能低下の自覚があり診察に訪れた人と、研究調査により

診断された人とでは違った対応が考慮されるべきであろう¹⁰⁾。今後、痴呆の早期診断に有効な客観的指標のエビデンスを踏まえ、認知機能検査や日常生活能力低下の具体的な内容を盛り込んだ操作的な診断基準が確立され、大規模な一般人口構成を背景とした縦断的研究が行われることが望まれる¹²⁾。

文 献

- 1) Bowen, J., Teri, L., Kukull, W. et al. : Progression to dementia in patients with isolated memory loss. *Lancet*, 349; 763-765, 1997.
- 2) Collie, A., Maruff, P., Shafiq-Antonacci, R. et al. : Memory decline in healthy older people Implications for identifying mild cognitive impairment. *Neurology*, 56; 1533-1538, 2001.
- 3) Daly, E., Zaitchik, D. and Copeland, M. : Predicting Conversion to Alzheimer Disease Using Standardized Clinical Information. *Arch. Neurol.*, 57; 675-680, 2000.
- 4) Flicker, C., Ferris, S.H. and Reisberg, B. : Mild cognitive impairment in the elderly : Predictors of dementia. *Neurology*, 41; 1006-1009, 1991.
- 5) Galasko, D., Chang, L., Motter, R. et al. : High Cerebrospinal Fluid Tau and Low Amyloid β 42 Levels in the Clinical Diagnosis of Alzheimer Disease and Relation to Apolipoprotein E Genotype. *Arch. Neurol.*, 55; 937-945, 1998.
- 6) Jack, C.R., Petersen, R.C., Xu, Y.C. et al. : Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, 52; 1397-1403, 1999.
- 7) Petersen, R.C., Smith, G.E. and Ivnik, R.J. : Apolipoprotein E Status as a Predictor of the Development of the Alzheimer's Disease in Memory-Impaired individuals. *JAMA*, 273; 1274-1278, 1995.
- 8) Petersen, R.C., Smith, G.E., Waring, S.C. et al. : Mild cognitive impairment : clinical characterization and outcome. *Arch. Neurol.*, 56; 303-308, 1999.
- 9) Petersen, R.C., Stevens, J.C., Ganguli, M. et al. : Practice parameter : Early detection of dementia : Mild cognitive impairment (an evidence-based review) . Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, 56; 1132-1142, 2001.

- 10) Petersen, R.C., Doody, R., Kurz, A. et al. : Current Concepts in Mild Cognitive Impairment. *Arch. Neurol.*, 58 ; 1985-1992, 2001.
- 11) Ritchie, K. and Touchon, J. : Mild cognitive impairment : conceptual basis and current nosological status. *Lancet*, 355 ; 225-228, 2000.
- 12) Ritchie, K., Artero, S. and Touchon, J. : Classification criteria for mild cognitive impairment. A population-based validation study. *Neurology*, 56 ; 37-42, 2001.
- 13) Tierney, M.C., Szalai, J.P., Snow, W.G. et al. : Prediction of probable Alzheimer's disease in memory-impaired patients : A prospective longitudinal study. *Neurology*, 46 ; 661-665, 1996.
- 14) Tierney, M.C., Szalai, J.P., Snow, W.G. et al. : A prospective study of the clinical utility of ApoE genotype in the prediction of outcome in patients with memory impairment. *Neurology*, 46 ; 149-154, 1996.



医療統計

11

臨床研究のすすめ

森 悦朗

MORI Etsuro

兵庫県立高齢者脳機能研究センター

著者プロフィール

(もり・えつろう)

1951年、福井県生まれ。

【略歴】1977年、神戸大学医学部卒業。1982年、神戸大学大学院医学研究科内科学系修了。1992年、兵庫県立姫路循環器病センター 神経内科。1990年、Scripps Clinic and Research Foundation, The Scripps Research Institute 研究員。1993年、兵庫県立高齢者脳機能研究センター。

【専門】臨床神経学。

【研究テーマ】脳血管障害、痴呆性疾患の診断・治療、行動神経学（高次脳機能）。



はじめに

本連載講座は、EBMと医療統計と名付けられ、EBMと統計学にかかわる問題をわかりやすく解説していくように企画された。EBM (Evidence Based Medicine) というのは、EBMの推奨者によれば、個々の患者の診療についての意志決定において、現在ある最良の証拠を良心的、公明にかつ思慮深く用いることであり、知りうるかぎりの疫学などの研究成果や実証的、実用的な根拠を用いて、効果的で質の高い患者中心の医療を実践するための手段であるとされている。臨床的専門技量は経験にもとづく医学のArtの部分であり、外部の根拠をもとにした批判的検証評価はScienceの部分である。EBMはその2つの統合であるとEBMの信奉者は述べている。

このように第三者的にEBMを紹介しているには訳がある。著者は臨床家であり、EBMの専門家ではなく、ましてや臨床疫学や生物統計学を専攻するものでもない。しかしこれまで脳血管障害や神経疾患の臨床研究にもっぱら従事し、またいくつかの臨床試験に関与し、いわばエビデンスを生み出す立場、すなわち臨床研究論文を書く立場にあった。またいくつかの臨床系医学雑誌の編集委員やreviewerとして研究論文を審査する立場で活動してきた。EBMの信奉者や生物統計学者ではないものが、EBMや医療統計を論じるのは差し出がましいようであるが、かえって第一線の臨床家には理解しやすい解説が可能となるようにも思われる。臨床疫学や生物統計の専門家によるEBM、臨床疫学、統計の出版物はすでになりの数にのぼっている。そのような専門家とは違って、脳血管障害や神経学に関する固有技術をもった臨床家・研究者が臨床疫学や生物統計学を論ずるところに意味がある。

EBMの専門家によってあげられるEBMの手続は、(1) 目前の患者の臨床上的問題の抽出、(2) エビデンスを示す文献の検索、(3) 文献の批判的吟味、(4) 目前の患者にそのエビデンスが適応できるかの判断、という4段階のステップである。いずれも第一線の医師にとっては困難な問題を含んでいて、とくに1と4のフロントエンドが本来的には最も重要な部分である。しかし、一般的には3は臨床疫学および生物統計学の知識の裏付けを要し、最も敬遠される部分である。EBMの専門家はこの部分の重要性を数式まであげて強調するものだから、かえっ

て第一線の医師をEBMから遠ざけてしまうという逆効果も生んできたようである。日本においては多くの医学部で臨床疫学を教える課程もなく、卒業後に医師として臨床疫学を研修する機会も少ないのが現状でもある。しかし臨床疫学の知識は臨床研究を行ううえで基本的なものであり、研究デザインを構築するうえで極めて有用である。逆にいえば、臨床疫学および生物統計学にもとづいて論文の批判的吟味ができるという能力をもつことは、臨床疫学および生物統計学な批判的吟味に耐える研究デザインをつくる潜在的な能力をもつことになるということである。EBMを十分に理解せよということは、一流の臨床研究者になれるほどのことを要求しているのに等しい。とんでもないことだといわれるかも知れないが、医師であれば一度ならず症例報告を含めて臨床研究報告をおこなった経験であろう。それならその報告は一流の臨床研究であってほしい。今回は「臨床研究のすすめ」と題してEBMと医療統計にまつわる問題をみていくことにする。



臨床研究の特徴と生物統計

臨床医学研究においては、「因果関係」、「診断の適否」、「治療の効果」が主要なテーマである。これらはつまり patient-oriented な研究である。さらに「病態生理」の分野においても貢献するところも小さくなく、疾患原因に関する重要な示唆もあり得るが、それらは直接的なものではない。つまり disease-oriented な研究手法としては限界がある。ここで基礎医学研究とは異なる臨床疫学的研究の特徴をあげてみる。まず当然だが、臨床例を対象にしていることである。臨床例は年齢、性、生活歴、職業など背景がさまざまである。個体差が大きい。また疾病の程度もさまざまである。このようなものすべてが交絡因子（主因子以外の、結果に影響を与える因子）となる。細胞や純系動物を用いた実験系では、そのような因子はすべて統制されているので通常は考慮する必要はない。そのような交絡因子の影響を最小にするために臨床研究では工夫が必要である。たとえば、多数例の無作為化によって2群が平均として同等になるようにする。交絡因子に関してマッチさせたペアをつくる方法（症例対照研究）、あるいは回帰あるいは偏相関などを用いて数学的に補正する方法などがある。前者ほど作為の介在

する機会がなく、また仮説が介在した操作も少ないので、一般に価値が高い。

一方、アウトカム（outcome：転帰あるいは結果と一般的には訳されるが、ここでは研究の目的にそって観測すべき必然的な結果を意味し、カタカナ表現することにする）の評価には疾病の病理過程および健康におよぼす影響の計測が必要だが、臨床研究ではアウトカムの計測も単純ではない。死亡とか、発作といった出来事なら計測しやすいが、神経症候、機能障害、あるいは生活の質（QOL）の程度を計測するのはむずかしい。健康に対する影響は症候、能力障害、社会的不利に分けられ、それぞれの計測方法として神経学的スケール、ADLスケール、QOLスケールあるいは社会的関係の評価が考えられている。一般的に、後者ほどその意味するところは直観的に明らかである。しかし環境、個体差、主観などによる影響（ノイズ）が大きいため信号の検出力は弱く、信頼性も低い。一方、前者ほど治療的介入の直接的効果（シグナル）の検出力は高く信頼性も高いが、その健康におよぼしている意味は分かりにくくなる。QOLは疾病をもった人間の福祉を考えるうえで理屈上重要な因子であり、アウトカムとして計測できればよいのだが、その方法は確立されていない。人それぞれの価値観が違いすぎて、確立しようということが土台無理なのかもしれない。脳血管障害で生じる神経症候や機能障害は複雑で、本来一元的に表現できるものではない。疾患の重症度を示す指標もない。たとえば、失語症と右手指の運動障害とどちらの機能障害が重症かという問いに対して落語家か時計職人かで答えは異なる。遷延性意識障害を起こした中脳の小さな病巣と、無症候のかなり大きな大脳病変ではどちらが重症かという問いに対して普遍的な答えはない。血液凝固線溶系、脳動脈の閉塞、脳循環、脳梗塞の体積などの病態生理学的指標の計測によって虚血性脳血管障害に対する治療的介入の効果の理論的裏付けを得ることができるが、それらの指標からでは患者個人の健康に対する直接的効果を知ることはできない。これらは代理的計測（surrogate measure）とよばれる。研究の目的とする臨床的に意味のある効果を評価するのが困難なとき、それらに代わって用いられる臨床検査や生理学的指標などをさす。しばしば効果のエンドポイントと代用評価項目は相反する結果となることもある。アウトカムとして何を計測するのも臨床研究では重要な問



エビデンスの水準

題である。スケールの問題はそれだけで重要な臨床疫学的課題でもあるので、アウトカムの計測とスケールについては項を改めて考えてみたい。

得られたアウトカムは統計学モデルにあてはめて検討する。むしろ、いずれかの統計学モデルにあてはまるようにアウトカムを含めて研究デザインを構築しておくというべきであろう。統計学モデルは、変数の型（出来事の頻度か、数値か、カテゴリーか）、変数の分布の型（パラメトリックかノンパラメトリックか）、変数の数などによって選ばれる。探索的研究か検証的研究かで違って来るかも知れない。パーソナルコンピュータ用の統計パッケージは安価に（無料ソフトも多い）、簡単に、相当複雑な統計分析も可能にしているが、適切に統計モデルを選び、適正に用いることが重要である。多重比較と反復比較の問題、後付けの解析やサブグループの解析など誤りやすい問題については別に考えてみたい。

以上のように臨床研究では研究デザインは研究の成否および価値を決定するほどの意味をもっている。それが臨床研究をむずかしくする理由でもあり、臨床疫学、生物統計学の知識が要求される所以である。ところが逆にこのことが研究者の自由な発想を可能にし、臨床研究の魅力ともなっている。臨床疫学、生物統計学は多くのノイズのなかからシグナルを抽出するツールを提供してくれるが、何がシグナルで何がノイズかを知り得て、研究を導いていけるのは臨床家なのである。

治療に関して判断を下す時にエビデンスを利用しやすくするため、エビデンスのランク付けの提案がなされている。いくつかの提案があるが、ここではAmerican Heart AssociationのStroke Councilで採用されているCookの分類を表①にあげておく。このようなランク付けは治療ガイドラインの作成に利用されている。すなわちランクの高いエビデンスがあれば治療として推薦されるというわけである。これに関して「EBMの目的は最良の臨床ガイドラインを作成することにある」とか「最良の臨床ガイドラインがあればどこでも誰でもEBMを実践することができる」という主張がなされて、行政サイドから、たぶん医療費の問題も絡めて、治療ガイドラインづくりが盛んにおこなわれている。そのような動きは、一面では臨床家の面倒を軽減することになるが、明らかにEBMの矮小化である。EBMはガイドラインづくりに加担し、医師の裁量にもかかわる問題を引き起こすという危惧も出てくるのも当然である。しかし、そのような押しつけの性格がEBMのなかに内在しているのではない。EBMの動きはまったくその逆で、個々人が、絶え間ない多くの情報から適切に情報を選択することを押し進めているのである。もし、利用する側の問題だと傍観しているEBMの信奉者がいるのなら許せない。EBMを理解するという事は、潜在的な一流の臨床研究者をつくり出すことだと心得るべきだろう。

いずれにせよ治療ガイドラインづくりは進行している。そのなかではエビデンスの評価がおこなわれるが、しば

表① エビデンスの水準

Level I	偽陽性 (α 過誤) および偽陰性 (β 過誤) の確率の低い無作為化割り付け試験からのデータ
Level II	偽陽性 (α 過誤) および偽陰性 (β 過誤) の確率の高い無作為化割り付け試験からのデータ
Level III	同時期のコホートとの比較による非無作為化研究からのデータ
Level IV	過去のコホートとの比較による非無作為化研究からのデータ
Level V	逸話的症例報告からのデータ

しば研究デザインにもとづいてランク付けが用いられる。しかし単純に研究デザインだけにとらわれて表面的なランク付けがなされたのでは正しい評価とはいえない。大規模な無作為化割り付け試験の結果は必ずしも信頼おけるものとはかぎらないのである。内容に立ち入って科学的批判的吟味がなされなければならない。それではどのようにして科学的批判的吟味を行うのか。多くの医学研究者は研究論文を医学雑誌に投稿し、受理されなかったという経験をもっているだろう。科学的批判的吟味の際に、なぜ論文が医学雑誌に受理されないのかという理由を考えてみると研究の質が実にわかりやすくなる。医学雑誌のreviewerはこのような点について、公正かつ慎重に批判的吟味をおこなっているはずであるからである。Greenhalghは論文が医学雑誌に受理されない一般的な理由として、(1)重要な科学的問題を扱っていない、(2)オリジナリティがない、(3)著者のあげた仮説を実際には証明していない、(4)違う種類の研究をおこなうべきだった、(5)対象を得ることがむずかしかったなどの実施の困難さが研究プロトコルを歪めた、(6)サンプルサイズが小さすぎる、(7)研究が非対照試験であったり適切な対照をおいていない、(8)統計解析が不正確あるいは不適當、(9)データから導き出されない非論理的な結論、(10) onflict of interest、(11) 書き方が悪くて理解できない、をあげている。



臨床研究デザイン

まず臨床研究のデザインの形態とその効果を考えてみよう。以下に代表的なデザインをあげる。

●無作為化対照試験 (randomised controlled trial : RCT)

無作為化対照試験では、対象はある介入と別の介入に無作為に割り付けられる。2つの群はある一定の期間追跡され最初に定義された転帰、(たとえば死亡、発作などの事象、計測値など)に関して分析される。無作為化対照試験では、特定の患者群における一つの変数に関する厳格な評価が可能で、患者群は平均において介入(独立変数)以外の点(交絡因子)で同一であり、結果の違いは理論的には介入に帰することができる。無作為化割り付けがなされていない場合、割り付けが何らかの必然的

な条件(systematic biasとよぶ)で決定されてしまうことがあり、統計的な比較は意味をなさなくなる。たとえば無作為化されずICUで治療された群と一般病床で治療された群を比較すると、おそらくICUで治療された群の方が転帰は悪いだろう。これからICU入院は一般病床入院にくらべ悪いという結論を導き出すのは誤りであるというのは簡単に理解できるだろう。未経験な研究者がしばしば冒すのはこの種の誤りである。

しかし無作為化対照試験は、費用や時間がかかり、無作為化が困難あるいは不可能、非現実的、非倫理的であることもある。また対象の選択基準および除外基準、エンドポイントなど、その試験の結果が一般化できる範囲に留意しておかねばならない。不完全な無作為化やすべての候補患者が無作為化されない(うまく反応しそうな患者だけを参加させる)といったあらゆるバイアスの存在に注意しておかなければならない。無作為化割り付けであっても結果的に2群は相同であったという保証はない。また転帰の評価に代用エンドポイントが用いられ、それが隠れたバイアスとなる。研究デザインを構築するうえでそのようなバイアスが生じる可能性をできるだけ排除することがキーとなり、研究論文の批判的吟味ではどのように配慮がなされたのかをみていくことがキーとなる。これらのバイアスについては項を改めて解説したい。

●コホート研究 (cohort study)

コホート研究では、2つあるいはそれ以上の群が特定の因子にもとづいて地域住民などから選ばれ、一定期間特定の疾患あるいはその他アウトカムを生じるかが調べられる。たとえば喫煙の有無に関して脳卒中の発生を調査するような研究がそれである。因果関係や予後の研究にはコホート研究が最も適している。

●症例対照研究 (case-control study)

症例対照研究では、特定の疾患の患者群を同定し、対照群(他の疾患の患者群あるいは健常者群、血縁者群など)と背景を一致させる。その2つの群の過去に得られたデータを引き出して比較する。症例対照研究は因果関係を導き出すのに適している。エビデンスの水準はそれほど高くないが、比較的稀な疾患で他の方法が採りにくい場合には有効である。症例対照研究の限界(あるい

はバイアス)は症例と対照を選び出す過程にあり、そこにバイアスが介在する危険がある。

●横断的調査 (cross sectional survey)

横断的調査では、疾患の代表的サンプルあるいは健常者の代表的サンプルを対象として一定の臨床的問題に関して一定の検査や調査がなされる。患者群(あるいは健常者)における特定の検査所見の頻度などが知り得るので、診断法やスクリーニングテストとの妥当性の検討などに適してる。

●症例報告 (case report)

症例報告については改めて述べる必要はないだろう。生物統計学が関与する余地も少ない。現在でも症例報告は研究形態のなかで最も多いものであろう。症例報告は安易で価値の低いエビデンスと一般にはいわれるが、非常に多くの情報が含まれていて、時には無視されていた情報が公になる有効な手段でもある。

●文 献●

- 1) Cook DJ *et al* : Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 102 (suppl. 4) : 305S-311S, 1992
- 2) Greenhalgh T : How to read a paper. Getting your bearings (deciding what the paper is about). *BMJ* 315 : 243-246, 1997
- 3) 森悦朗 : 急性期臨床試験の問題点. 脳と循環 2 : 333-339, 1997
- 4) 森悦朗 : 無作為化対照試験 : 外科治療が内科的治療に勝るとい証拠を得るため方法論. 脳神経外科ジャーナル 9 : 409-415, 2000

研究方法の質

森 悦朗, 博野信次

MORI Etsuro & HIRONO Nobutsugu
兵庫県立高齢者脳機能研究センター

著者プロフィール

(もり・えつろう)

1951年、福井県生まれ。

【略歴】1977年、神戸大学医学部卒業。1982年、神戸大学大学院医学研究科内科学系修了。1992年、兵庫県立姫路循環器病センター 神経内科。1990年、Scripps Clinic and Research Foundation, The Scripps Research Institute 研究員。1993年、兵庫県立高齢者脳機能研究センター。

【専門】臨床神経学。

【研究テーマ】脳血管障害、痴呆性疾患の診断・治療。行動神経学（高次脳機能）。



はじめに

自然科学におけるエビデンスの価値は、発見の大きさ、すなわち、真理にどこまで近づいたか、理論構築のためにどれほどの意味があるかであり、歴史的に決定される。一方、臨床医学におけるエビデンスの価値は、研究の目的や対象の重要性とともに、診療の意志決定のために用いることを前提にしているわけだから研究手続きのもっともらしさ、得られた結果の確かさが基準になる。そこで今回は研究計画のなかでもとくに重要な対象選択、アウトカム計測、エンドポイントの問題を取り上げる。Randomized controlled trial (RCT) であれば高い水準のエビデンスと機械的に判断してはいけない。その研究方法にまで立ち入って、その質を問わなければならない。今回問題にするのは研究方法の質である。今回のテーマには、研究のなかでもわれわれの日常の経験や知識を駆使しなければならない部分である。統計学的な考え方を背景にして整理している。統計学そのものではなく、むしろ統計学を用いる前提として考えていただきたい。



I. 研究の対象は何か、どのように選ばれたか

研究の目的に応じて最も適切な対象者を設定する。理想的には選定基準に合致する全員（母集団）を対象とすればよい（悉皆調査）が、たとえば脳梗塞患者を対象とする場合に、日本中の脳梗塞患者全例を調査することは通常不可能である。そのため、母集団から「それを適切に代表している」標本を研究対象として抽出することになる。統計学的解析は、標本集団から母集団全体の特性を推定することであり、標本集団が母集団を「適切に代表している」ことを大前提としている。したがってこの仮定が成り立たない対象から得られたデータを統計学的手法を用いて検討すること自体がそもそも誤りである。

臨床試験では、どのような対象を選択し、どのような対象を除外するのかが重要な研究デザインの要素の一つである。軽症や重症の患者を除外しておこなわれる薬剤の臨床試験がよくある。そこで得られた結果は軽症や重症の患者にも当てはまるとはかぎらない。拡大解釈はできないのである。このように対象の選択基準と除外基準が明示的になっていて、基準どおりに対象が選択されていれば結果の解釈は楽である。しかし、

対象の選択基準は必ずしも明示的とはかぎらないし、実際の対象は必ずしも選択基準どおりには選択されていない。実際におこなわれた臨床研究からいくつかの例をあげて考えてみよう。

まず対象の選択方法は明示的だが現実と解離している場合がある。t-PAの臨床試験 (ECASS)¹⁾ ではプロトコル違反がかなりの数にのぼり、しかも違反例に頭蓋内出血が多かったことが示されている。選択基準にはCTスキャンで早期虚血徴候が軽いという条件があった。プロトコル違反の多くはこの選択基準が守られなかったことであった。したがって、選択基準は現実の状況を表現しておらず、実際に対象とされたのは選択基準に反して多くの早期虚血徴候を有するものが多く含まれた患者群であった(そのことが試験の失敗の原因であろうと、臨床現場では早期虚血徴候の例をうまく除外できないというのが現実で、つぎのECASS IIでは教育によって実際違反例は著明に減少した)。つぎに対象の選択方法は明示的だが単純に解釈してはならない場合があり、この場合はもっと複雑である。たとえば健常対照群を新聞広告でボランティアを募って集めたとしよう。この対照群は一般的といえるであろうか。その応募者は、毎朝、新聞に目を通し、社会に貢献したいという意識の高い人々である可能性が高く、一般的な人々の代表とはいえないだろう。さらにつぎのようにまったく明示的ではなく、解釈も推測も困難な場合もある。国際EC-IC bypass trial (Barnett study)²⁾ が失敗に終わった原因として、bypass手術が効きそうだと考えられた患者は臨床試験外で手術され、効果が期待できない患者のみが試験参加したためではないかということが指摘されている。つまり選択された患者は患者全体を代表していないのではないかという疑問が呈されているのである。実際どうだったのか、効果的な患者というのをはたして選別することが可能だったのかどうか、真偽のほどは確かめようもないが、臨床試験外でのbypass手術が多くおこなわれたのは事実である。



II. アウトカムとエンドポイント

アウトカムは疾病の病理過程および健康に及ぼす影響を計測することで評価する。疾病の健康に対する直接的影響は症候、能力障害、社会的不利などであり、それぞ

れの計測方法として神経学的スケール、ADLスケール、QOLスケールあるいは社会的関係の評価が考えられている。これらのアウトカムの意味するところは直観的に明らかであるが、環境、個体差、主観などによる影響(雑音)が大きいため信号の検出力は弱く、信頼性も低くなる。一方、血液凝固線溶系、脳動脈の閉塞、脳循環、脳梗塞の体積などの病態生理学的指標の計測によって虚血性脳血管障害に対する治療的介入の効果の理論的裏付けを得ることができる。これらのアウトカムは一般的に治療的介入の直接的効果(信号)の検出力は高く信頼性も高いが、健康に及ぼしている意味はわかりにくく、これらの指標によっては患者個人の健康に対する直接的効果を知ることはできない。このような指標は代用的計測(surrogate measure)とよばれている。代用的計測とは、比較的簡単に計測できて、まれであったり時間がかかったりで計測しにくい毒性や治療効果を予測するが、それ自身は毒性とか臨床的效果を直接示すものではない変量と定義することができる。

臨床試験では研究の目的に沿った仮説が立てられ、アウトカムが評価され、仮説を証明するに足る目標(エンドポイント)が設定される。エンドポイントの設定は仮説と深く関係しているのである。たとえば脳浮腫を抑え頭蓋内圧亢進を抑制する作用をもつ薬剤があったとすると、その作用から重症例に対して脳死を減少させる効果が期待できるのでアウトカムは死亡という事象を計測し、エンドポイントとして急性期の死亡率の減少を検証するのがよいだろう。一方、軽症虚血組織の蘇生を狙うような薬剤では、死亡率の減少は仮説としにくく、機能的非自立以上の障害を減らすという仮説をもつことがよいかも知れない。また研究の目的によって、たとえば薬理学的作用の証明を目標とする比較的基礎的な研究の場合と、治療的介入の可能性をみるための探索的研究、治療的介入が健康にとって有用であるのかどうかを検討する検証的研究の場合とでは計測されるアウトカムは異なり、エンドポイントも異なって当然である。

● 1. 複数のエンドポイント

仮説が2つ以上存在するときは、主たるエンドポイント(primary end point)と副次的エンドポイント(secondary end point)に分けられるが、探索的研究ではどのエンドポイントが適切なのかを見出すため多く設定

され、検証的研究では主たるエンドポイントだけが結論的意味をもつとする意見が強いようである。ところで一つの仮説を立証するために主たるエンドポイントを2つ以上設定したときには統計の反復が問題となる。統計学的な問題については、次回詳述する予定であるが、たとえば従来わが国でおこなわれてきた脳血管障害に対する薬剤の臨床試験ではエンドポイントの設定がなされず、また神経徴候を個々に解析する方法が採用されていた。この場合エンドポイントが無数にあると考えてもよい。実際上肢の麻痺には効果があるが、下肢の麻痺には効果がないとされた薬品が存在する。冗談のようだが、右手と左手の麻痺を別々に評価していれば、右手の麻痺には効果があるが左手の麻痺には無効であるといった解析結果が出た可能性もあった。探索的な研究であれば、仮説をつくるためとして許されるだろうが、治療の有用性を最終的に判断するような第3相試験の場合には問題がある。

●2. 後付けの研究と層別の解析

賭博で負けがこんでいたとしても、賭けをつづけていればいつか勝目がやってくる。そこで止めれば、つまり勝ち逃げをすれば勝てる。研究でも同じことが起こりうる。もう少しで有意の差が得られるからといって予定していた追跡期間を延ばしたり、例数を増加させたりすること、または予定には達していないが、すでに有意の差が得られたので途中で研究を止めることがこれにあたる。予定の追跡期間および例数は研究計画のなかで仮説にもとづいて設定されている。そのことは一方の群が明らかに悪い転帰をとって研究を継続するのが非倫理的と考えられるときとは訳が違う。

またデータのなかからよいところだけを集めて解析すること（後付け層別解析）は誤った結論を導き出す危険性がある。たとえば、アスピリンの脳卒中一次予防効果の研究³⁾では、男女あわせて有意なアスピリンの効果が示されたが、後付けの層別解析でその効果は男性だけにかぎられていると報告された。この結論のために、後の研究でその結論が間違いであることが示されるまで、女性にアスピリンを投与することが控えられるようになってしまったということがある。類似の例が最近わが国で認可されたある薬品にもあるがお気づきだろうか。仮説にもとづいて予定されておこなわれた層別解析でないかぎり、後付けの層別解析で得られた知見は、つぎの研

究における仮説の合理的根拠に用いることはできても、結論を導くために用いてはならない。

●3. 代用エンドポイント

上述のように代用的計測を用いると例数や追跡期間を小さくできてコストを下げることができる。そのためしばしば臨床試験でも代用エンドポイントが用いられる。確かに病態生理学的指標の計測によって効果の理論的裏付けを得ることができる場合もある。たとえば内頸動脈-中大脳動脈の閉塞の患者にEC-IC bypassをおこなったところ中大脳動脈領域の脳血流は改善するという研究報告は多くある。この場合脳血流測定は代用的計測である。本来の効果は将来に起こるかも知れない虚血性脳血管障害の抑制である。脳血流の改善は脳血管障害の予防を直接意味してはいない。そのような報告からEC-IC bypassは脳血管障害の予防効果があるとは結論できないのであるが、そのような誤解はしばしば見受けられる。代用エンドポイントと本来のエンドポイントとの関係が明らかに証明されていないかぎり、代用エンドポイントを用いた研究の意義付けには注意を要する。逆に代用エンドポイントの意義が証明され、信頼性が高く、感度と特異度が高いもの場合には大変有用である。



III スケールの問題

脳血管障害の治療の効果を判定するための普遍的に受け入れられている方法はないが、一つのアプローチは障害を要約して等級化するスケールの使用である⁴⁾。スケールを用いれば変数は一つにまとめられ、少なくとも先に述べた反復比較の問題は解消できる。脳血管障害の臨床試験では神経学的スケール、ADLスケールがよく用いられる。スケールの善し悪しは、妥当性と信頼性に依存している。少なくとも精神測定学あるいは統計学的手法を用いて科学的に妥当性と信頼性が確認されている必要がある。スケールの妥当性とは、測定すべきものを正當に計測しているかどうかということ、スケールの信頼性とは、同一あるいは異なった計測者が測定をくり返したとき安定した結果が得られるかどうかということである。従来わが国でおこなわれてきた脳血管障害に対する薬剤の臨床試験では、神経症候が個々に評価されてきた。各症候の評価基準は示されず、妥当性や信頼性も未確認

で、その全般的な変動の判断も担当医師の主観に任されていたという問題が指摘できる。アウトカム計測法として決して妥当だとはいえない。

神経学的スケールとしてはMathew scale, National Institutes of Health stroke scale (NIHSS), Hemispheric Stroke Scale (HSSまたはGraded Neurological Scale), Canadian Neurological Scale (CNS), Scandinavian Stroke Scale (SSS), Orgogozo Scaleなどがあり、臨床試験で用いられてきた。いずれも複数の神経学的症候を障害の程度にしたがって数値化し、その合計をもって全般的スコアとしている。神経学的スケールの場合の妥当性は神経欠損の大きさあるいは重症度を真に表現しているかどうかということであるが、ここで神経欠損の大きさあるいは重症度とは何かという根本的な命題に対する普遍的な解が存在せず、基準妥当性を検討すべきゴールドスタンダードも存在しない。しかし、これまでに開発された神経学的スケールは、少なくとも多くの経験豊かな神経専門医が賛同した評価項目を備えているといった内容妥当性は有していると考えてよい。またADL、転帰の全般的評価、あるいはQOLの評価との相関があり、スケールの示す数値が線型であるという証明はないものの、構造的妥当性も示されていると考えられる。スケールの信頼性は治療の前後でスコアの変化をみるという臨床試験での使用目的からきわめて重要な因子である。とくに多施設共同の試験であれば検者間信頼性もきわめて重要な要素である。これまでのほとんどの神経学的ス

ケールでは検者内・検者間信頼性は確かめられ、生命統計学的に合理性の裏付けがなされている。従来の神経学的スケールには項目の選定や重みづけを中心にさまざまな問題点も指摘され、万人が満足しているわけではなく、現在でもよりよいスケールをめざして開発がつづいている。Japan Stroke Scale (JSS)⁵⁾はそのような線型性をめざした後発のスケールの一つである。丁寧に科学的な開発過程を経て、量的に重症度を計測できるのがうたい文句ではあるが、それでも重症度の平均的印象度を数値化した指標にすぎない。「理想的」スケールの開発にはおのずと限界があり、筆者は悲観的である。しかし実際のところは、複数の異なった神経学的スケール間の相関はきわめて高いことが明らかにされているので、いずれのスケールを用いても結果に大差はないかも知れない。神経学的スケールのすべての測定項目は重症度という仮想的因子に寄与していると考えられるものが複数選択されているため、重み付けがほぼ妥当と思われる程度であればほぼ同じような結果を生み出すということは当然ともいえよう。類似した測定項目が複数用意されていると冗長性が高くなり、計測に時間を要する一方で信頼性も一般的には高くなる。

ADLスケールおよび転帰の全般的評価としてはBarthel Index, (Modified) Rankin Scale, Glasgow Outcome Scaleが比較的によく使われている。これらのスケールの意味するものは、疾病が健康に与えた影響という点で神経学的スケールよりは具体的で、それがもつ意

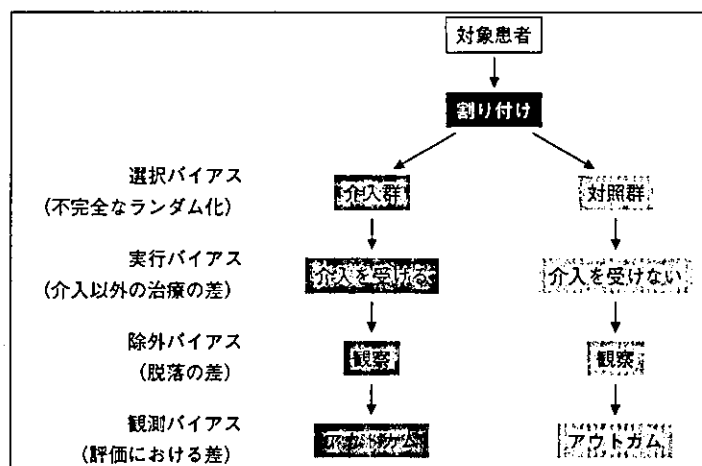


図1 ランダム化割り付け試験におけるさまざまなバイアスの源 (Greenhalgh T, 1997⁶⁾ より引用, 著者訳)

味はより明瞭である。しかし、ADLスケールおよび転帰の全般的評価の変動幅は比較的小さく、一般的には神経学的スケールよりも検出力は弱いと考えられる。Barthel Indexはそのなかでは検出力が高いことが指摘されている。

それではどのようなスケールを使うのがよいのだろうか。スケールの妥当性を追求するのは無理だという考えから、むしろ信頼性を重視したい。評価基準が具体的かつ明示的であり、信頼性が高くなるように配慮されたスケールが最良ではないだろうか。NIHSSでは検者の習熟度を上げるように、ビデオテープによる教材が用意され、習熟度の評価もできるよになっている。また、このような訓練をおこなうことによって実際に検者間信頼性が上がることが確かめられている。もう一つ重要な要件は、広く使用されているということである。このことは検者の習熟にもかかわるが、他の試験との比較が容易であり、さらに共通のスケールが用いられたデータを蓄積していけば将来の試験のための貴重なデータベースが作られることになる。



IV バイアス

研究デザインをつくるうえで最大の努力は、系統的バイアス（一方の群に必然的に偏って生じるバイアス）がないようにすることに払われなければならない。バイアスとは「測定された値を真の値から系統的にゆがめるもの」のことで、バイアスの源は研究の各時点に存在する⁶⁾ (図①)。研究参加の時点では割り付け前バイアス pre-allocation bias、標本選択にかかわるものは選択バイアス、介入研究の場合の実行バイアス、対象からデータを得るときに生じる観測バイアスがある。

割り付け前バイアスは、患者の研究に参加するか否かの意志決定にかかわって生じる系統的差異、たとえば自己選択バイアスがある。対象者自身の意志が対象選択に影響するバイアスで、たとえば脳梗塞の有病率を調査するために希望者を募ると、喫煙者などのその危険因子を多く有するものが多数参加し、結果として脳梗塞の有病率を多く見積もってしまったり、疾患対照者として健康ボランティアを募集すると、社会経済的に恵まれ健康に関心のある、「超正常者」が多く集まってくるなどがあげられる。また、インフォームド・コンセントを通

じての研究者側からの誘導もありうる。

選択バイアスは、非無作為化試験や不完全な無作為化によって導入される。たとえば非無作為化試験で、stroke care unit (SCU) での治療と一般病棟での治療の転帰に与える影響を調べたとする。SCUでは死亡率が高く、一般病棟では死亡率が低い。その結果から生命の転帰に関してSCUは一般病棟に劣ると結論することの誤りには容易に気付くだろう。無作為化でないので、重症患者は必然的にSCUに入院しただろうし、軽症であれば一般病棟に入院したからである。これが選択バイアスの例である。

実行バイアスは、比較すべき問題の介入以外の治療にある系統的な差異をいう。一般的に比較すべき治療以外はすべて同等だと仮定されるが、必然的に系統的に違いがある可能性がある。たとえばアスピリンの効果を比較する試験の場合、アスピリン投与群には必然的に制酸薬の投与が多くなっている可能性がある。副作用はアスピリンではなく制酸薬に起因しているかも知れない。

観測バイアスは観測者が予見をもって評価するなどの測定の際の系統的な差異をいう。たとえば降圧薬の臨床試験で、降圧薬を服用しているという予見をもって血圧測定をおこなったとき、血圧が高ければ、少し時間をおいてもう一度測定し、下がっていればその血圧を記録するかもしれない。降圧薬非服用患者では一回の測定で記録しているかも知れない。

脱落バイアスは、研究対象からの脱落が研究目的としている事象と関連していることによるバイアスをいう。たとえば薬物の臨床試験の際に、効果がないものや副作用を経験したもののほうがコンプライアンスが悪く、研究対象から多く脱落すると、結果として効果のあるものが最終的に多く残ってしまうことなどがあげられる。

これらのバイアスを統計学的解析で解消することはできないため、研究計画策定の段階でバイアスを最小限にするように努力しなければならない。選択バイアスをなくするためには母集団から無作為に等しい確率で選択する確率的なサンプリングをおこなうことが望まれるが、臨床事象を対象とする臨床研究では、確率的サンプリングをおこなうことは現実的にはむずかしく、非確率的サンプリングがおこなわれることが多いため選択バイアスを完全になくすことは不可能である。もちろん非確率的サンプリングのなかでも、ある選択基準を満たす全例を対

象とする連続サンプリングをおこなったり、研究の対象となりやすい人を対象とする選別サンプリングをおこなう場合にもできるだけ選択バイアスを少なくするように配慮すべきであることはもちろん必要であるが、選択バイアスのある研究イコール誤った研究という短絡的なもの見方をすると、臨床研究そのものが成り立たなくなることもまた事実である。臨床研究を発表する際には対象者の設定方法などの研究方法を詳細に記載し、読者がどのような選択バイアスが考えられ、そのことが研究結果の妥当性を損なわないかどうかを判断できるようにしておかなければならない。ある研究が病院を受診した患者データのみにもとづき脳梗塞の有病率を報告している場合には、選択バイアスを過小評価しているとみなさなければならないが、病院を受診した患者データのみにもとづき脳梗塞に対する血栓溶解療法の有効性を報告していたとしても、重症度や他疾患の合併などほかのバイアスが適切にコントロールされていれば、受診データのみであるという選択バイアスが結果の妥当性を大きく損なうとは考えにくい。

実行バイアスや観測バイアスをなくすには二重盲検化が有効である。偽薬投与やシヤム手術が倫理的に許されない場合には第三者的な、できればブラインド化された担当医師や評価者に管理や評価を委ねるような次善の策も考えられる。また、たとえ二重盲検化RCTであったとしてもいろいろなところにバイアスの源は潜んでいる。いろいろな場面を想定して慎重に研究計画を立てたほうがよい。脱落バイアスは、まず脱落や中止の基準をあらかじめ決めておくことが重要で、二重盲検試験であれば開示前に脱落や中止を決定する作業をおこなうようにする。脱落例を無視してしまうことは一般に介入側に有利にはたらく。それを回避するには、intention-to-treat (ITT) 分析をおこなうことである。ITT分析というのは、もともと割り付けされた患者のデータのすべてを、あらゆるプロトコル違反例、脱落例、割り付けと反対の治療を受けたものもすべて含めて、プロトコルに沿っておこなわれたものとして分析する方法をいう。ときにITT分析がおこなわれずに、治療が実際におこなわれたものだけで分析される場合があり protocol-compatible (PC) 分析あるいは有効性分析とよばれる。この場合、たとえ対象患者がRCTの一部であったとしても、分析の目的から実際上コホート研究、すなわち低水準の研究とみなされる。



最近大流行のガイドラインづくりのなかで、研究はエビデンスの水準で評価され、その評価は勧告や推奨の強さに反映する。RCTであれば高い水準のエビデンスだと機械的に位置づけられるようではいけない。RCTであってもその研究方法にまで立ち入って、その質を判断したうえで評価しなければならない。一方、臨床研究を計画する立場では、仮説、対象選択、アウトカムの評価、エンドポイントの設定、結果の解析、解釈という研究者の工夫が最も要求される部分なのである。

●文 献●

- 1) The European Cooperative Acute Stroke Study (ECASS) : Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke. *JAMA* 274 : 1017-1025, 1995
- 2) The EC/IC Bypass Study Group : Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. *N Engl J Med* 313 : 1191-1200, 1985
- 3) Canadian Cooperative Stroke Group : A randomised trial of aspirin and sulfipyrazone in threatened stroke. *N Engl J Med* 299 : 53-59, 1978
- 4) 森悦朗 : 急性期臨床試験の問題点. 脳と循環 2 : 333-339, 1997
- 5) Gotoh F *et al* : Amano T. Development of a novel, weighted, quantifiable stroke scale: Japan stroke scale. *Stroke* 32 : 1800-1807, 2001
- 6) Greenhalgh T : How to read paper : Assessing the methodological quality of published papers. *BMJ* 315 : 305-308, 1997



データ分析の方法
統計学的手法

博野信次, 森 悦朗*

HIRONO Nobutsugu & MORI Etsuro

愛媛大学医学部看護学科

*兵庫県立高齢者脳機能研究センター

著者プロフィール

(ひろの・のぶつぐ)

1959年、大阪生まれ。

【略歴】1984年、大阪市立大学医学部卒業。同年、大阪市立大学附属病院第一内科学教室研修医、住友病院神経内科医員。1990年、馬場記念病院神経内科医員。1992年、兵庫県保健環境部立病院局経営課脳研準備室主査。1993年、兵庫県立高齢者脳機能研究センター臨床研究科研究員。1998年、米国 University of California Los Angeles, School of Medicine 客員研究員。1999年、兵庫県立高齢者脳機能研究センター臨床研究科研究員。2002年、愛媛大学医学部看護学科教授。

【専門】臨床痴呆学、神経心理学。

【研究テーマ】痴呆の神経心理学、痴呆の臨床神経学。

【趣味・愛読書】古代中国史



はじめに

前回述べたように臨床医学研究では通常、大きな母集団から「それを適切に代表している」標本を研究対象として抽出することが必要であり、標本集団から母集団全体の特性を検討するために統計学的手法が用いられる。生物統計学者ではない臨床家が統計学的手法に通暁しておく必要はもとよりなく、また現在は種々の有用な統計解析用ソフトが利用でき、数学的知識がなくても解析をおこなうことはむずかしくない。しかし、臨床研究をおこなう場合にはデータ処理や解析を考えたうえで研究計画を立てる必要がある、そうでない場合は折角苦労して大量のデータを集めても、結局は解析に使えないなどという悲惨なことになるかねない。またデータの性質を考慮せず、明らかに誤りである統計学的手法を用いて解析をおこなうと誤った研究結果を導き出してしまうことになる。このようなことがないように、臨床研究をおこなう臨床家はある程度の医療統計知識を有しておくことが望まれる。同様に臨床研究の論文を読みエビデンスの質を検討する際にも、用いられている統計法に問題がないかどうかを判断するだけの知識が要る。厳しい査読を経て医学雑誌にあらわれる論文だが、用いられている統計法が必ずしもつねに正しいとはいえない。今回は臨床家が押さえておくべき誤った結論につながりやすい統計学の問題点をみていくことにする。筆者が医学雑誌に投稿された論文を査読する際によくみかけるものである。



誤った統計解析結果の解釈

母集団をうまく「代表しない」標本集団を抽出してしまうこと、すなわち誤差が生じる原因には、前回述べた系統誤差すなわちバイアスの他に偶然誤差がある。偶然誤差とは、標本集団にまったくの偶然により母集団にはない偏りが生じてしまうことである。この偶然誤差の程度を精度あるいは定度precisionとよぶ。たとえば男女比が同じである集団から10名をランダムに選んできた場合、どちらかが8名以上と大きく偏ってしまう確率は約10%もある。しかし100名中80名以上となると、その確率ははるかに小さくなり、標本集団のサイズが小さいと偶然の影響を受けやすくなることがわかる。精度を高くする、すなわち偶然誤差を少なくするためには標本数(サンプ

ルサイズ) を大きくする必要があり、このことは後述する第1種、第2種の過誤と密接に関係している。

統計学的解析には統計学的推定と統計学的検定がある。統計学的推定は母集団の性質を推定することで、たとえば正規分布における平均値や分散、そうでない場合の中間値や4分位値などがその指標としてあげられる。統計学的推定には点推定 point estimation, すなわち1つの値で求めたい指標の推定をするものと、区間推定 interval estimation, すなわち求めたい指標の推定値を区間で与えるものがある。たとえば脳梗塞患者の空腹時血糖値を推定する場合に、ある標本集団の平均値で真の母集団の平均値を推定する場合が点推定であり、ある信頼率(多くは95%が用いられている)で真の母集団の平均値がそのなかに含まれる区間、すなわち信頼区間 confidence interval を求めるのが区間推定である。点推定値である標本平均が同じ130mg/dlであったとしても、95%信頼区間が125-135mg/dlの場合と、100-160mg/dlの場合はその推定の精度が異なる。上述のように標本数が大きくなると精度が高くなり、信頼区間の幅は小さくなる。

統計学的検定とは、標本集団がある仮定を支持する確率を求めることである。この仮定を統計的仮説という。ある統計的仮説のもとで、ある事象が生じる確率が、はじめから定めてある有意水準(多くは5%が用いられている)と比較し、それよりも小さければその仮説を棄却するのである。このように棄てられることが望まれている(?) 仮説を帰無仮説といい、帰無仮説の反対の仮説を対立仮説という。たとえば喫煙者と非喫煙者で脳梗塞の発症率を比較する場合には、喫煙者と非喫煙者で脳梗塞の発症率に差がないと仮定し(帰無仮説)、実際に観察された喫煙者と非喫煙者の脳梗塞発症率の相違が生じる確率を計算するのである。この確率が有意水準5%よりも小さければ、帰無仮説のもとでは、このようなことが生じる可能性は5%以下であり、そのぐらいの確率で生じることはあり、そのぐらいの確率で誤りである危険はあるが(危険率)、帰無仮説を棄却し、「喫煙者と非喫煙者で脳梗塞の発症率に差がないとはいえない」とするのである。実際には帰無仮説が正しい、すなわち差がないにもかかわらず、偶然の結果、誤って帰無仮説を棄却してしまう誤りを第1種の過誤(α エラー)とよぶ。危険率とは第1種の過誤を生じる危険のある確率である。

統計解析結果の解釈の誤りの多くは、この基本的な検

定手順の理解不足が原因である。これらの解釈の誤りには、

- 1) 帰無仮説が棄却できなければ差がないことを示している。
- 2) 有意水準が小さいほど差が大きい。
- 3) 反復比較・多重比較による「有意な差」の検出。
- 4) 統計学的関連があれば因果関係がある。

などがあげられる。

● 1. 帰無仮説が棄却できなければ差がないことを示している

この誤りはたとえば喫煙者と非喫煙者で脳梗塞の発症率に差がないとする帰無仮説のもとでの確率が有意水準5%以下でない場合は、喫煙者と非喫煙者で脳梗塞の発症率が同じであると解釈する誤りである。たとえば30名ずつの喫煙者と非喫煙者を経過観察し喫煙者のうち6名(20%)、非喫煙者のうち3名(10%)が脳梗塞を発症したとすると、フィッシャーの直接確率計算法により計算した、このようなことが生じる確率は23.6%であり、有意水準5%に達しない。したがって、この場合は、「喫煙者と非喫煙者で脳梗塞の発症率に差がない」とする帰無仮説を棄却できないため、差があるとはいえないことになる。しかし、「差があるとはいえない」ことは「差がないこと」を意味しているのでは決してない。この結果から喫煙が脳梗塞の発症に関連しないと解釈してはいけないのである。もし300名ずつの経過観察をおこなっていた場合に同じ発症率が観察されたとすると、カイ自乗検定で計算した確率は0.06%となり0.06%の危険率でこの帰無仮説を棄却できるのである。このように、帰無仮説を棄却できない場合は、実際に母集団で帰無仮説が正しい、すなわちこの場合は喫煙者と非喫煙者で脳梗塞の発症に実際に差がないことによる場合もあるが、実際には正しくない、すなわち喫煙者と非喫煙者で脳梗塞の発症に差があるのに、それを検出できないことによる場合もある。このような誤りを第2種の過誤(β エラー)という。また、第2種の過誤が生じる確率を1から引いたもの($1-\beta$)を検出力(パワー)という。検出力とは実際に母集団で帰無仮説が正しくない、すなわち差がある場合に、統計学的に有意な結果が得られる確率である。第2種の過誤は偶然誤差により生じるため、標本数を大きくし精度を高めることにより、第2種の過誤を生じる確率を小さくする、換言すれば検出力を大きくす