

これらの指標、つまり RRR, ARR, NTT, そして NRR について, GUSTO スタディでの例で比較してみる.

表 5. GUSTO スタディ結果を用いた各 EBM 指標の比較

Arm	Event rate		
tPA	6.3%		
SK	7.3%	RR 0.86	
			RRR 14%
			ARR 1%
			NNT 100 人
		$NRR = RR / (1 - RR) = 6.3$	

GUSTO スタディでは, 血栓溶解剤である tPA と SK を比較している (表 5). イベント発生率は tPA 群で 6.3%, SK 群で 7.3% であった. 絶対値で見ると, 1% tPA のほうが優れる結果であった. 相対指標であるリスク比 (RR: Risk Ratio) は 0.86 であったので, 相対リスク減少率 (RRR: Relative Risk Reduction) は 14% になる. また, 絶対リスク減少率 (ARR: Absolute Risk Reduction) は 1% であるから, NNT は 100 人になる. すなわち tPA を使用することで, 100 人に 1 人の割合でイベントを回避できるというわけである. 臨床論文では RRR という相対指標で結果を示すことが多いが, 現場主義の EBM では絶対指標のほうを重視する. それはなぜかという点, 相対では有効性を感覚的に把握できないためであろう. 一方, 14% リスクを減少するという相対指標では, リスク発生のベースライン値の大きさを超えて比較しやすい面がある. ここで新しく 2002 年に提案された NRR を考えてみよう. NRR は RR を  $(1 - RR)$  で割った値であるので, これは相対指標になる. この例では 6.3 と算出される. これはどういった意味があるのか. tPA 治療によって救われる人と救われない人の比である. tPA による治療があっても, まだ 6.3 倍の人が救われず残っている. 当然, この値が小さいほど治療効果が高いことになる. しかし, この指標は相対であるので, やはり現場で有用な指標にはならない. すなわち, 目の前の患者へ応用する EBM にはあまり役立たない. しかし, 治療法間の比較にはこうした相対指標は有用であろう. 例えば, 他の抗血栓療法で SK に対する NRR が 5 倍だとすると, この tPA よりも優れていることがわかる. このように, NRR は医療技術評価の尺度として有用かもしれない. このような様々の指標を開発することで, 異なった解釈を可能にすることも考えてもらいたい.

ここには述べなかったが, その他にも Number Needed to Screen というような統計量も提案された (Rembold, 1998). 例えば, 乳がんを 1 例検診で回避するには何名検診しないといけないかなので, 検診の有効性を評価するための指標になる. しかし形式的に見れば, これは先に述べた NNT と同様である.

#### 4. EBM と診療ガイドライン

EBM というのは診療実践そのものであるが, 近年 Evidence に基づいた診療ガイドライン策定も進んでいる. 診療ガイドラインの目標とは, 診療の最低水準を引き上げて, どこでも同じ医療が Jpn J Biomet Vol. 24, Special Issue, 2003

受けられるようになることである。さらに言えば、医療がより透明になることも期待される。そういったガイドラインは専門家のコンセンサスで作成することも可能であるが、そこに Evidence を引用しながら作成するようになってきた（折笠, 2000b）。ガイドラインを策定するにおいては、臨床疑問をまず明らかにする。こういった内容を書くかということであり、これは主として医師が担当しなければならない。続いて、その疑問に答えるようなデータを収集するわけだが、ここではリサーチ・ライブラリアンが貢献する。もちろん、研究デザイン知識については統計家がサポートし、臨床用語については医師が協力する。得られたデータを吟味するさいには、いわゆる Evidence 水準を明らかにするわけであるが、その際には統計家や医師が関与する。最後に、ガイドライン声明を書き上げる段階では主として医師が関与することになるだろう。このように、Evidence としての臨床データに基づくガイドライン策定において、統計家が協力する余地は大きい。実際、著者も糖尿病及び脳卒中全般のガイドライン策定において、情報収集及びその吟味において協力をしてきた。

## 5. EBM の問題点

EBM というのは臨床データに基づいて、その時点でベストと考えられる標準治療を万人に提供するの目標である。しかし臨床データがさらに整ってくると、個人別治療も可能になってくるだろう。すでに高脂血症などでは、それ以外にリスクファクターの個数で治療法を変化させている。これは、Framingham スタディという Cohort 研究データなどを参考にして設定したそうである。科学を駆使してここまで進んだかという感がするわけである。このように見てくると、EBM というのは完璧で譲ることのできないものに思えるわけだが、それが万能ということはない。Evidence ではなく Experience をより重視する医療として、Narrative based medicine (NBM) もよく聞かれるようになってきた（Greenhalgh, 1999）。Evidence により画一化するのではなく、患者のエピソードや診療経験で変えるべきというものである。EBM も最終的に個別医療を目指すので、結果的には目標は同じと思われる。しかし、EBM 発祥の原典であるイギリスから、EBM とは相反する NBM が出てきたのは決して悪いことではないと思う。

さらに、EBM の総本山とも言える Oxford 大学臨床試験部の Peter Sleight 博士は、EBM に関して多くの問題点を指摘している（桑島, 1999）。例えば、EBM はあまりに単純すぎて、正しい根拠が不十分である。死亡率ばかりがエンドポイントに設定されている。RCT 万能なので、リスクの低い患者に偏りすぎている。もう一方の万能であるメタアナリシスには様々のバイアスが入りやすい。EBM の良い点を主として本稿で述べてきたが、そうではない点もあることを知っていてほしい。

## 6. ま と め

EBM において生物統計学はあまり関係ないという声もよく聞くが、本稿で生物統計学が寄与してきた、及び寄与すべき領域について述べた。Evidence 水準について整備すること、RCT やメタアナリシスであっても良質かどうかを見分ける基準を策定すること、EBM 特有の指標（統計量）について統計学的性質を検討すること、医療技術評価や診療ガイドライン策定にもっと貢献

すること、以上の可能性を促したつもりである。

## 文 献

- Altman, D. G. (1998). Confidence intervals for the number needed to treat. *Br Med J* 317: 1309-1312.
- Altman, D. G. and Anderson, P. K. (1999). Calculating the number needed to treat for trials where the outcome is time to an event. *Br Med J* 319: 1492-1495.
- Bender, R. (2001). Calculating confidence intervals for the number needed to treat. *Contr Clin Trials* 22: 102-110.
- Bjerre, L. M. and LeLorier, J. (2000). Expressing the magnitude of adverse effects in case-control studies: "the number of patients needed to be treated for one additional patient to be harmed". *Br Med J* 320: 503-506.
- Cook, R. J. and Sackett, D. L. (1995). The number needed to treat: a clinically useful measure of treatment effect. *Br Med J* 310: 452-454.
- Evidence-Based Medicine Working Group (1992). Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 268: 2420-2425.
- Greenhalgh, T. (1999). Narrative based medicine in an evidence based world. *Br Med J* 318: 323-325.
- Laupacis, A., Sackett, D. L., and Roberts, R. S. (1988). An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 318: 1728-1733.
- Massel, D. and Cruickshank, M. K. (2002). The number remaining at risk: an adjunct to the number needed to treat. *Can J Cardiol* 18: 254-258.
- Moher D., Cook D. J., Eastwood S., Olkin I., Rennie D., Stroup D., for the QUOROM Group (1999). Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. *Lancet* 354: 1896-1900.
- Moher D., Schulz K. F., Altman D. G., for the CONSORT Group (2001). The CONSORT statement. *JAMA* 285: 1987-1991.
- Rembold, C. M. (1998). Number needed to screen: development of a statistic for disease screening. *Br Med J* 317: 307-312.
- Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., and Haynes, R. B. (2000). Evidence-based medicine, 2nd edition. Edinburgh: Churchill Livingstone.
- 折笠秀樹 (1995). 臨床研究デザイン. 東京: 真興交易医書出版部.
- 折笠秀樹 (2000). EBM の実践に必要な臨床統計学のキーワード. 今月の治療 8(6): 636-643.
- 折笠秀樹 (2000a). 薬物治療に伴う有害事象に関する論文の読み方. 日本薬剤師会雑誌 52(8): 1121-1126.
- 折笠秀樹 (2000b). EBM に則った診療ガイドライン. *Mebio* 17(2): 42-46.
- 折笠秀樹 (2001). EBM 実践にさいして必要な統計学. IN: 「EBM 循環器疾患の治療 2001-2002 (三

- 田村・山科・川名・桑島, 編集)』, pp.1-20, 東京: 中外医学社.
- 桑島巖 (1999). EBM は医療の実践に不可欠か. *Medical Tribune* 8月19日号: 18.
- 名郷直樹 (1999). EBM 実践ワークブック. 東京: 南江堂.
- 久繁哲徳・監訳 (1998). 根拠にもとづく医療. 東京: じほう.
- 舟喜光一, 折笠秀樹・訳 (2001). 信頼性の統計学—信頼区間および統計ガイドライン—. 東京: サイエンティスト社.

# 特集

## 高血圧とEBM —高血圧のRCT—

### 研究の質を評価する 仮説の設定と必要症例数

折笠秀樹

富山医科薬科大学医学部 統計・情報科学

◇ 臨床研究には事前に仮説を設ける検証的な研究と、そうではない探索的な研究とがある。検証的な研究では症例数設計は必須なものとなる。症例数設計のアプローチである精度ベースと検定ベースについて述べた後、症例数設計の概念について統計用語を用いて解説する。この概念についてはぜひ理解してもらいたいし、どのようなときに症例数設計が必要になるかも、つかんでほしい。つぎに、症例数設計の実際について説明する。有効率のような二値データ、血圧値のような連続データ、そしてイベント発生のような生存時間データでの事例を示す。症例数設計が必要と判断した場合には、その時点から然るべき統計学者に相談することが望ましい。◇

#### はじめに

臨床研究を始める前にプロトコル(研究計画書)を作るのはほぼ常識化してきたが、そのなかでも困難な箇所とは症例数設定の根拠を示すことではないだろうか。それでは、なぜ事前に症例数を決めなくてはいけないのか疑問をもつ人もいるだろう。その答えは何かというと、その研究における仮説は何かを明確にさせるためである。検証的研究の場合には検証したい仮説があるはずである。しかも、統計解析では有意性検定をおこなうはずである。一方、探索的研究では仮説を探索するのが目標であるから、最初に検証したい仮説はないのが普通である。したがって、探索的研究では症例数の設計は特段必要ないことが多い。

本稿では、症例数設定に関する2つの基本的アプローチについてまず説明する。つづいて、統計概念を用いて症例数設定というのはどのようなことをおこなっている

のか、その基本的考え方を述べる。つぎに、簡単な事例をいくつか示す。二値データによる比率の場合、連続データによる平均の場合、そして生存時間データの場合である。最後に、必要症例数に影響する要因は何かについて説明する。

症例数設計については成書がいくつかあるが<sup>1)~4)</sup>、なかでも Machin ら<sup>1)</sup>の書物が最も広範に渡り数表も揃っていて有用であろう。ソフトについては、少し古いながら相互比較をしている総説を数編紹介しておく<sup>5)~7)</sup>。ちなみに筆者は、現在 FDA にいる Janet Elashoff が開発した nQuery Advisor<sup>8) 9)</sup>を使っている。臨床医向けに書かれた平易な総説<sup>9)</sup>、それに少し難解かもしれないが、症例数設計に関するランドマーク的総説論文<sup>10)</sup>も読むことをおすすめする。

#### 1. 精度ベースの方法

症例数の設定法の第一は、精度ベースのアプローチとよばれるものである。この方法は検証的研究のみならず、探索的研究の場合にも適用可能である。

まず1群だけのオープン試験で、治療法の有効率を推

#### KEY WORD

症例数設計、臨床研究、検出力

定するのが目的の試験を考えよう。有効率の推定に関する精度を定めて、症例数を設計するのがこの方法である。たとえば100例に設定した試験では、有効率の推定精度は最大で±9.8%と設計されている。なぜなら、有効率の95%信頼区間はそれが50%のとき最大になり、その近似標準誤差(SE)は $\sqrt{0.5 \times 0.5 / 100} = 0.05$ のため、95%信頼区間は $0.5 \pm 1.96 \times 0.05 = 0.5 \pm 0.098$ となるからである。もちろん、有効率は研究を始める前にはわからないが、有効率が50%のときに精度が最も悪く(つまりSEが最大)になることが知られている。仮に有効率が30%(または70%)とすると、1群100例の試験における有効率の推定精度は±9.0%( $=1.96 \times \sqrt{0.3 \times 0.7 / 100}$ )になる。2群比較の検証的試験でもこの方法は適用できるが、あまり使われないようである。

## 2. 検定ベースの方法

症例数設計というと、通常この検定ベースのほうを指す。すなわち、検証的な比較試験の場合にはおもにこの方法を用いる。そこでは、帰無仮説と対立仮説を設定することが必要となる。とくに重要で、かつ難しいのが対立仮説の設定である。例を示そう。

新薬と標準薬を比較して、その降圧効果に差があるかを検証したいとする。このときの指標は、新薬での降圧度と標準薬での降圧度の差になる。帰無仮説は何かというと、降圧度の差はない(つまり0)というものである。対立仮説は研究者が主張したいほうなので、新薬のほうが標準薬よりも降圧度が大きいとなる。これだけでは不十分である。たとえば、その差は20 mmHg というように明示する必要がある。この差20 mmHg という値を、症例数設計では $\delta$ (デルタ)であらわす。この $\delta$ は過去の情報や、研究者の予想・期待から引き出す。この $\delta$ を過大評価すればいくらかでも症例数を減らせるが、結果的に事前に設定した $\delta$ よりも群間差があまり小さくなりすぎると、統計学的に有意な結果が得られない。2群比較の場合に関していうと、検定ベースにおける検出力を50%にした場合が先に示した精度ベースの方法に等しい。

## 3. 症例数設計の考え方

知っておくべき統計学の必須概念としては5つある。第1は母集団(population)と標本(sample)の違いである。第2はばらつき(variability)の尺度である。第3は効果サイズ(effect size)であり、これは前述の $\delta$ に等しい。この3つが最も重要なものである。実はネイマン・ピアソン流の統計学的方法で検定をおこなうので、さらに有意水準(significance level)と検出力(power)が必要となる。有意水準は通常は両側5%に設定する。検出力とは、対立仮説(研究者の仮説)が正しいときに、得られたデータでの検定結果が統計学的に有意となる確率のことである。検出力は通常80%以上に設定する。検出力というのはいわば保険のことであり、検出力50%のときは保険なしの状況を意味する。すべて予想通りにいけば有意になるが、一歩まちがうと非有意になってしまう。また、保険を高くかけることで検出力を上げようとする、必要とする症例数は増える。検出力はいわばその研究が成功する確率といえる。このため、研究者はしばしばこの症例数を考えているが、これでうまくいきますかと相談にくる。このとき、統計学者は検出力を計算してあげることになる。したがって、症例数設計のことをpower calculationと欧米ではよくいう。

有効率の2群比較を用いた症例設計の概念を図1に示す。2標本間の真の差が0%と仮定した標本分布が帰無仮説であり、真の差が20%と仮定したほうが対立仮説をあらわす。これは比率データであるが、その標本分布は症例数が増えると近似的に正規分布となる。比率に関する差の値が $\delta$ (この例では20%)である。データから推定された差の値が図の点線よりも右側になれば、有意差ありと判定する。左側ならば有意差なしと判定する。点線のラインのことを限界値(critical value)とよぶ。このような決定をすると、対立仮説が正しいとき差なしとする誤り( $\beta$ 過誤)と、帰無仮説が正しいとき差ありとする誤り( $\alpha$ 過誤)が存在する。通常 $\alpha$ は5%(この図では限界値の右部分を2.5%)となるように限界値を決める。そのときの限界値の左部分(斜線)が $\beta$ 過誤となり、 $\beta$ は20%以下となるように症例数を決める。症例数が大きくなれば、推定値のばらつきが小さくなる。したがって、これらの

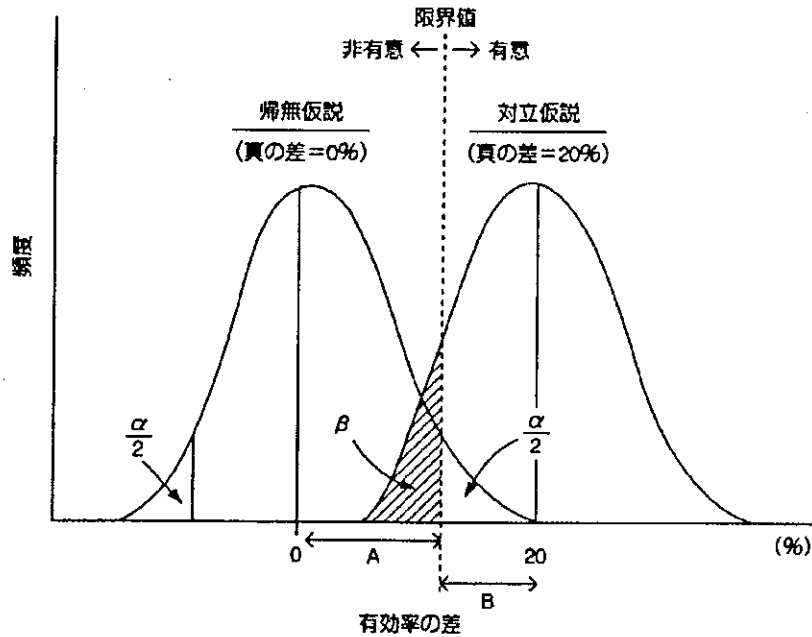


図 1. 有効率に関する 2 群比較を例にとった症例数設計の概念図

標本分布はもっと細長くなり、重なり部分がほとんどなくなる。症例数が増えると  $\beta$  が小さくなるので、検出力が高い研究ということになる。

#### 4. 症例数設計の事例

有効率に関する 2 群比較を例にとる。仮に、標準治療での有効率を 50% ( $=p_1$ ) として、新規治療では 20% 上回り 70% ( $=p_2$ ) という仮説を考える。これは統計学的には対立仮説とよばれ、真の差は 20% ( $p_2 - p_1 = 0.2$ ) となる。帰無仮説は有効率について群間差がないことであるから、真の差は 0% ( $p_1 = p_2 = 0.5$ , あるいは  $p_2 - p_1 = 0$ ) である。帰無仮説のほうの分布を用いて (図 1 を参照)、差が 0% のラインから限界値まで (A で示した) を計算すると、 $1.96 \times \sqrt{\{p(1-p) + p(1-p)\}/n}$  となる。ここで、 $p = (p_1 = p_2) = 0.5$  であり、 $n$  は各群共通の症例数である。係数が 1.96 になっているのは、両側 5% 有意水準としたためである。対立仮説の分布のほうで眺めると、図の B は  $1.28 \times \sqrt{\{p_1(1-p_1) + p_2(1-p_2)\}/n}$  になる。ここで、 $p_1 = 0.5$  であり、 $p_2 = 0.7$  である。また、 $n$  は各群共通の症例数である。係数が 1.28 となっているのは、 $\beta$  を 10% (つまり検出力 90%) としたためである。図 1 からわかる

ように、 $A = 20 - B$  (小数表示では  $0.2 - B$ ) であるから、

$$1.96 \times \sqrt{\{p(1-p) + p(1-p)\}/n} = 0.2 - 1.28 \times \sqrt{\{p_1(1-p_1) + p_2(1-p_2)\}/n}$$

となる。この等式を解くと、

$$\begin{aligned} n &= [1.96 \times \sqrt{2p(1-p)} + 1.28 \times \sqrt{\{p_1(1-p_1) + p_2(1-p_2)\}}]^2 / 0.2^2 \\ &= [1.96 \times \sqrt{2 \times 0.5 \times 0.5} + 1.28 \times \sqrt{\{0.5 \times 0.5 + 0.7 \times 0.3\}}]^2 / 0.04 \\ &= 128 \text{ 例/群 (検出力 90\% 以上ということ切り上げ)} \end{aligned}$$

と算出される。有意水準のほうの係数 1.96 は  $Z_{\alpha/2}$ 、検出力のほうの係数 1.28 は  $Z_\beta$  と表記する。検出力を 80% に変えると、 $Z_\beta$  は 1.28 から 0.84 に変わる。

つぎに、連続データに関する平均値の 2 群比較を考える。有効率に関する標準偏差 (つまり標準誤差) は、有効率の値から直接正規近似で求めた。連続データでは、さらにそのデータについての標準偏差の予想値が必要となる。過去の情報などから予想し、各群で共通にすることが多い。同様の議論から、対応のない t 検定をおこなうときの症例数  $n$  は

$$n = 2(Z_{\alpha/2} + Z_\beta)^2 \sigma^2 / \delta^2$$

と算出される。ここで、両側 5% 水準の検定をするなら  $Z_{\alpha/2} = 1.96$ 、検出力が 80% なら  $Z_\beta = 0.84$ 、 $\sigma$  は共通の標準

表 1. 生存率解析での症例数設計(1群あたりの症例数)

年率	リスク低下率	平均追跡期間		
		3年	4年	5年
3%	30%	1,638	1,226	979
	20%	3,909	2,927	2,330
4%	30%	1,226	917	732
	25%	1,820	1,362	1,087
	20%	2,927	2,191	1,750
5%	30%	979	732	583
	20%	2,339	1,750	1,396

簡単にするため、ログランク検定にもとづく設計法を用い、計算には nQuery<sup>®</sup> ソフトを使った。なお登録期間は2年とした。

偏差、 $\delta$  は平均に関する予想した群間差である。例を示そう。標準治療にくらべて新規治療は、血圧を 20 mmHg さらに下げる (たとえば、血圧値を標準治療での平均 180 mmHg なのが、新規治療では平均 160 mmHg まで下げる) という仮説をもったとする。血圧測定のばらつきを示す標準偏差を 30 mmHg だと想定する。有意水準を両側 5%、検出力 90% とすると、

$$n = 2(1.96 + 1.28)^2 \times 30^2 / 20^2 = 48 \text{ 例/群}$$

と算出される。なお、血圧値を 160 mmHg から 140 mmHg に下げると予想しても、必要症例数は変わらない。もし対応のある t 検定の場合では、

$$n = (Z_{\alpha/2} + Z_{\beta})^2 \sigma^2 / \delta^2$$

で表される。ここでは、 $\sigma$  はペアごとの差に関する標準偏差、 $\delta$  はペアごとの差に関する予想値である。

最後に、イベント発生までの期間について症例数設計の例を述べる。コントロール群でのイベント発生率(年率)が、まず必要になる。それから、新規治療によりイベント発生を相対的に何%減らすかの情報(つまり  $\delta$ )が必要となる。標準治療ではイベント発生が年 5% であったのが、新規治療でそれを 4% にまで減らすと想定すると、相対リスク低下率は 20% になる。ここまでは比率や平均の場合と同様であるが、さらにもう 1 つの情報が必要となる。それは登録期間と追跡期間の情報である。観察される総期間(年)を知るために、この 2 つの情報が必要となる。あとは通常の有意水準と検出力でよい。表 1 に例を示した。これは 2 群比較の例で、登録期間 2 年で追跡期間は平均で 3~5 年とした。コントロール群での年率を 3~5%、相対リスク低下を 20~30% として、必要症例数を nQuery Advisor<sup>®</sup> というソフト<sup>6)</sup> により算出した。いくつかの手法があるが、ここでは最も簡単なログラン

ク検定ベースの手法で計算した。年率が小さくなると症例数が増え、リスク低下が小さいと症例数は急増し、追跡期間が短いと症例数が増えることが読みとれる。

最後に一つだけ注意点がある。これまで理想の計画で進行した場合の症例数設計を示したが、実はプロトコルどおりにいかないことがある。これを不完全例(attrition)とよぶことがあるが、それを考慮すると症例数は多少増える。たとえば脱落率 10% であれば、症例数は 11% ( $1 \div 0.9$ ) 増える計算になる。

## 5. 知っておくと役立つ性質

症例数はどのような条件で上下するのだろうか。まず、検出力を上げたいければ症例数は増える。効果サイズが大きい(つまり差が大きい)試験では症例数は減る。効果サイズが 2 倍になると症例数は 4 分の 1 になり、逆に半分だと症例数は 4 倍必要になる。このように二乗に反比例するので影響が大きい。データのばらつきが大きくなると症例数は増えるが、これは単なる比例関係で影響する。比率の場合の数値例を表 2 に示す。比率の差が 0.2 に等しいとき、比率の絶対値が 0.5 から小さくなるほど症例数は少しずつ減る。ここには示していないが、0.5 から大きいほうでも同様に少しずつ減る。比率での標準誤差は 0.5 のときに最大となるためである。表 2 の右側をみると、比率に関する相対リスクは 0.5 と一定であるが、比率の絶対値が 10 分の 1 に低下すると必要症例数は約 10 倍に増えていることがわかる。イベント数がきわめて低いエンドポイントを扱う場合には、症例数は膨大になることがわかる。



表 2. 比率に関する 2 群比較の必要症例数

両側有意水準 検出力	5% 80%	5% 80%	5% 80%	5% 80%	5% 80%
$P_1$	0.4	0.2	0.1	0.1	0.01
$P_2$	0.6	0.4	0.3	0.05	0.005
必要症例数 (1 群あたり)	97	82	62	435	4,673

### おわりに

筆者のメッセージとしては、臨床研究を計画する研究者にとっては症例数設計の考え方と、それはどういう場合に必要かについてはよく知っておくべきであろう。しかし、症例数設計のための計算 (power calculation) まで研究者がする必要はないと思う。もし、ある研究で事前の症例数設計が必要だと判断したら、適切な統計学者に相談することをお勧めする。チームで臨床研究するほうが効率的であり、しかも内容についても充実する。

### 参考文献

- 1) Machin D *et al* : Sample size tables for clinical studies, 2nd ed, Blackwell, Oxford, 1997
- 2) Shuster JJ : Practical handbook of sample size guidelines for clinical trials, CRC Press, Boca Raton, 1993
- 3) Desu MM *et al* : Sample size methodology, Academic Press, New York, 1990
- 4) Kraemer HC *et al* : How many subjects? CA : SAGE Publications, Beverly Hills, 1987
- 5) Iwane M *et al* : *Control Clin Trials* 18 : 65, 1997
- 6) Dupont WD *et al* : *Control Clin Trials* 11 : 116, 1990
- 7) Goldstein R : *American Statistician* 43 : 253, 1989
- 8) Elashoff JD : nQuery Advisor<sup>®</sup> Release 4.0. Statistical Solutions, Boston, 2000
- 9) Florey CD : *BMJ* 306 : 1181, 1993
- 10) Lachin JM : *Control Clin Trials* 2 : 93, 1981

1) Machin D *et al* : Sample size tables for clinical

# 介入試験の原則

富山医科薬科大学臨床統計学 教授  
折笠 秀樹

## [Summary]

臨床研究の研究デザインは大きく分けて四つあるが、その中でも介入試験というのは最も厳密かつ困難な研究方法である。しかし、新しい治療法の確立には欠かせない研究であることも事実である。ここでは介入試験の種類として、症例集積法、個人内比較法、個人間比較法について、その特徴について解説する。介入試験で最上位に位置付けられるものが、個人間比較の中のランダム化比較試験 (RCT) というものである。このランダム化比較試験 (RCT) に関して、試験を実施する際の組織作りの大切さについても述べる。最後に、このような RCT 研究論文の読み方と探し方について簡単に紹介する。

## Key Words :

臨床試験 □ 統計学 □ 研究デザイン □ 疫学 □  
メタアナリシス

## 介入試験とは

臨床研究というのは、患者あるいは住民を対象とした研究をいう。その計画法には4種類あるといわれている(表①)。観察研究はあるがまますべてを観察するもので、アンケートなどを用いる調査がそうである。医療では疫学研究と呼ばれることもある。介入研究とは日常診療と異なる新たなことを行う研究である。これはいわば臨床実験なので実行は難しいが、その結果の価値は高いとされる。介入研究を実施する際には、施設審査委員会 (IRB) と文書同意 (IC) が不可欠である。介入を伴う実験なので介入試験とも呼ぶことがある。本稿では同一語として用いている。同じようにして、臨床試験ということもある。観察研究と介入研究が臨床研究の代表であるが、特に看護研究などでは質的研究が実施されることもある。エイズの患者数名にインタビューして、いろいろ困っていることや悩みを話してもらおう。そのテープを聞きなおして、問題点を構造的に整理する研究方法である。最後の統合研究は別名メタアナリシス、あるいは系統的レビューと呼ばれているものである。近年、EBM とともに最高位に位置づけられるようになり、一躍注目を集めるようになった研究方法である。それは、過去の同様の研究成績を統計学的手法で併合するものである。

たとえば、有酸素運動と免疫防御機能の関係を調べるとしよう。観察研究であれば、有酸素運動を行っている人を選んできて、血液を採取して免疫防御機能を調べてもらう。血液採取は日常行わないから介入研究ではないかと思うかもしれないが、介入というのは治療・予防法に関与するものを指す。この場合は有酸素運動がそうなので、それに介入をかけたかどうかポイントになる。介入研究だとすれば、これまで有酸素運動を行ったことのない人に数ヵ月行ってもらい、前後で血液採取して免疫防御機能の変化を検討する。介入研究では、このように因果関係 (effect A on B in C) をみることが多い。別の例をあげよう。慢性心不全の患者に症状などの QOL アンケートを取り、それと神経体液因子との関係を調べるとしよう。これは関連性 (association between A and B in C) をみる研究であり、研究法は観察研究になる。介入的な治療や予防は行っていないからである。神経体液因子の測定は新規的で介入のようにも思われるが、このような研究は介入研究とはいわない。いわゆる疫学研究になる。こうした研究の場合にも、できることなら IRB と IC を経た方がよい。

新しい治療法の確立において、介入研究は欠かせない。観察研究だけでは医療は変わっていかない。遺伝子治療や臓器移植というのも介入研究から出発した。それは症例集積という単純な計画法ではじまり、最終的にはランダム化比較試験という高尚な方法で確立されるというステップを取ることが多い。

介入研究の条件として、それは前向き (prospective) なことがある。前向きなら介入研究というわけではないが、介入研究なら前向きでないといけない。前向きというのは、研究仮説を立ててから将来に渡ってデータをとる研究である。後ろ向き (retrospective) 研究というのは、過去のデータを収集する研究である。過去のデータをとる際に恣意が入りやすく、前向き研究の方が優れている。

介入研究は観察研究よりもいろいろな点で優れている

表① 臨床研究の種類

1. 観察研究	ありのままを観察する、いわば疫学研究
2. 介入研究	実験的な手法で行う研究
3. 質的研究	インタビューなどを通じて行う研究
4. 統合研究	過去のデータを統計学的に併合する研究

表② 介入試験の種類とその特徴

1. 症例集積 (case series)
2. 個人内比較 (非比較研究)
(ア) 前後比較研究 (before-after study)
(イ) N-of-1 研究 (N-of-1 study)
3. 個人間比較 (比較研究)
(ア) 非ランダム化比較試験
(controlled clinical trial ; CCT)
patient-preference method
alternate allocation method
(イ) ランダム化比較試験
(randomized controlled trial ; RCT)
並行群間比較法 (parallel design)
交叉比較法 (cross-over design)

が、介入研究をすることができないテーマもある。たとえば、受動喫煙による肺癌リスク増大に関する研究がそうである。介入内容として受動喫煙があるが、今まで受動喫煙ではなかった人に対してそれをさせることは倫理的にも問題があるし、参加してくれるわけがない。したがって、このようなテーマではアンケートなどで観察研究をするしかない。

### 介入試験の種類と特徴

介入試験にはいくつかの種類がある (表②)。最も素朴な研究法は症例集積である。これは、たとえばある術式を 10 名の患者に適用し、そのうち 8 名が 1 ヶ月延命したというような研究である。ある介入的なことを行って成績をまとめる。それと類似するのが前後比較法であ

る。これは10名の患者に、コレステロール低下作用があると思われるドリンク剤を1ヵ月間飲んでもらい、1ヵ月の前後でコレステロール値の変化を検討するものである。N-of-1 デザインというのは、1名の患者の中で介入をいくつか設ける方法である。AとBという介入があれば、最初の1ヵ月間はA治療、その後1ヵ月間はB治療などとする。そして、個人内でAとBという治療の善し悪しを比較する。個人別治療の策定には、このN-of-1 的な発想が大切と思われる。

これらはN-of-1を除き介入そのみであり、別の代替法とは比較していない。その意味では、結果の信憑性は少し劣ることになる。一方、ほかの代替法と比較する比較試験が考えられる。N-of-1のように比較を逐次にするのではなく、同時に行う方が科学的に妥当といわれる。同時に分けるためには、それぞれの患者はどちらか一方の介入だけを受ける。それをどう選ぶかにおいて、いくつかのオプションがある。一つは患者が好きな方を選ぶというものである。これは、膵臓癌末期の患者でのコーヒー浣腸の臨床研究で用いられた<sup>1)</sup>。英語でpatient-preference designなどと呼ぶ。第二は、交互に与えるとか、曜日を決めて与えるという方法である。これはcontrolled clinical trial (CCT)と呼んだりする。最後はランダム化比較試験(randomized controlled trial, 略してRCT)である。ここでは、ランダムに(つまり確率を用いて)割り振りを行う。最後の方法が最も公平に分けられる方法とされる。つまり、分けられた2群はその背景がきわめて似てくる。最初のデザインは患者に選ばせているので、その背景には大きな違いが生じることが多い。背景に違いが生じたら、結果に差があっても、それは背景の差を反映しただけではないかといった批判が生まれる。そこで、そういった批判の生まれにくいRCTがベストなのである。ただRCTをやったつもりでも、たとえば封筒法を用いるとか、非盲検のため割り付け後すぐ変更してしまったりすると、背景は似てこなくなる。わが国の臨床試験ではこういう事例が多々みられ

た。RCTをやったつもりなのだが、最後に背景を比較すると違いが大きかったというのである。

いわゆるRCTにも並行群間比較(パラレル)というのと交叉試験(クロスオーバー)が知られている。効果のみられ方に個人差が大きいときには、交叉試験の方がよい。しかし、交叉試験は実施が困難なので、あまり勧められない。

介入試験、特にRCTを行うときにさらに二重盲検だとよい。しかし、わが国の市販後臨床試験ではほとんど二重盲検のものはない。当然のことながら、二重盲検の方がベターであるが、その実施は困難になる。まずプラセボを用意しないといけない。市販後では長期試験が多いため、長期にわたりプラセボを使い続ける難しさも生じる。

オープンラベルの試験では注意が必要である。二重盲検でない、途中でも比較結果が随時分かってしまうので、それが漏れないよう注意する必要がある。ちょっと形勢不利などという情報が漏れると、細工をすることが容易にできるからである。しかし、よく考えると癌臨床試験はすべてオープンである。投与スキームが大きく異なったりするので、二重盲検は不可能に近い。オープンであっても、エンドポイントがハードであれば問題は少ない。癌の場合でも、エンドポイント判定の結果は実施者には分からないようにしている。また、効果安全性委員会(Data and Safety Monitoring Board ; DSMB)<sup>2,3)</sup>を設けている。エンドポイントの判定を盲検にして行うこと、それと第三者からなるDSMBで途中結果をモニターし、それ以外の関係者は途中でむやみに比較できないよう注意する必要がある。そのためにはデータセンターというものを設けて、そこに集まる情報はいかなる関係者へも流さない体制をとればよい。

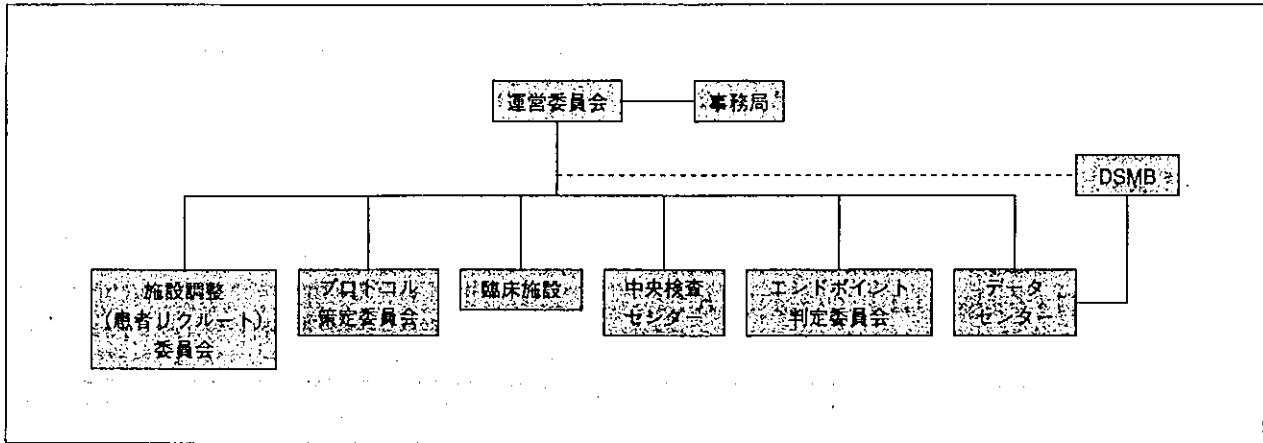


図 多施設共同試験の組織体制のプロトタイプ

DSMB : Data and Safety Monitoring Board

### RCT の組織

RCT は多施設で実施されるのが通常である。その理由は二つある。単一施設では外部妥当性（ほかの施設でも同じ結果になるか）はみれないことと、多施設でない必要症例数を集められないことである。こうした多施設共同試験を行うときには、その組織をがっちり組んで、役割分担することが成功の秘訣とされる。そのプロトタイプを図に示した。多施設というからには、病院（臨床施設）という組織があるのは当然だが、その他にもいろいろある。先ほど述べたデータセンターがそうである。ここに臨床試験のデータがすべて集まる。登録割り付けセンターもある。症例登録を受け付け、適格条件を確認して、ランダム割り付けした群を知らせるセンターである。これはデータセンターでなされることも多い。また、エンドポイント判定委員会も、特にオープン試験でイベント（脳卒中や心血管など）の判定には重要である。エンドポイントは介入試験で最も大切な事項である。運営委員会はその臨床試験の推進を担い、コアとなるメン

バーから構成されている。プロトコル策定もこの委員会の一部で構成されることが多い。事務局というのは事務的サポートを行う。研究班などであれば班長の施設で行う。経理の手続き、文書管理、施設への連絡文書などを扱う。ケースカードは事務局ではなく、データセンターで集めた方がよい。事務局にはスポンサーも関係することがあり、データセンターを兼ねると、大切なデータ情報が事務局を通してスポンサーに流れやすくなるためである。最後に、特に長期試験で大切な組織が DSMB である。これは第三者からなる組織であり、上の運営委員会とは別になる。運営委員会は結果がよくなるよう折る方だが、こちらは途中で問題なく進行しているかを公平に判断する。場合によって中間解析を実施することがあるが、その結果の検討もここでなされる。中間解析は主にデータセンターで実施される。

### RCT 論文の読み方

RCT 論文の読み方については、すでに書いているので<sup>46)</sup>、簡単に述べる。論文の書き方が CONSORT 声明

として発表されたが<sup>7-9)</sup>, これは読むときにも参考になる。読み方の第一ポイントは、患者背景の表をよく眺めることである。比較群の患者背景が類似しているかを確認し、さらにはどういった患者層が対象かを平均的に捉えることが肝要である。第二にはフロー図が示されることが多いが、これによりRCTの質を確認する。具体的には、途中での脱落や不遵守症例がどれくらいあるかを読む。患者の消息に関する追跡率はほぼ100%が望まれている。副作用、同意撤回、不遵守などによるプロトコル逸脱は年に20%以内が目安であろう。長期になればもちろんもっと高くなるだろう。第三のポイントは主要エンドポイントに関する結果である。イベントがエンドポイントであるような場合には、相対リスク (relative risk) をよく読む。これが20%以上あればかなり治療効果ありと思われる。また、効果サイズの区間推定である95%信頼区間もよくみておきたい。それは sampling variation を反映しており、結果の信憑度が分かる。NNT (number needed to treat) などという指標も読めれば、それは患者中心の医療へも生かせる。つまり、NNTが20人であれば、20人に1人余計に救える効果サイズと分かる。その他、RCTでよく示される情報として、サブ解析の結果と副作用の発生割合のデータがある。

### RCTの探し方

RCTが介入試験の王様であるが、それをMEDLINEで探すにはまずキーワードで検索し、その後limit (絞り込み) でPT (Publication Type) = Randomized Controlled Trial とすればよい。その他、RCTのほかにCCT

も含めたものとして、COCHRANE LIBRARYの中にデータベースがある。2002年現在で約35,000件のCCTが含まれている。また、RCTのメタレジスター (<http://controlled-trials.com/>) からは約14,000件のRCT (過去および現在進行中) を調べることができる。

### ■文献

- 1) Vanchieri C : Alternative therapies getting notice through best case series program. *J Natl Cancer Inst* 92 : 1558-1660, 2000
- 2) 折笠秀樹 : 長期臨床試験の評価. 医学のあゆみ 167 : 633-635, 1993
- 3) Armstrong PW, Furberg CD : Clinical trial data and safety monitoring boards. *Circulation* 91 : 901-904, 1995
- 4) 折笠秀樹 : EBM実践にさいして必要な統計学—ランダム化比較試験とメタアナリシス論文の読み方を中心に. 「EBM循環器疾患の治療」, 三田村秀雄, 山科章, 川名正敏, 桑島巖 (編), pp.1-20, 中外医学社, 東京, 2001
- 5) 折笠秀樹 : EBM実践のための臨床試験論文の読み方. *Therapeutic Research* 23 (7) : 1287-1294, 2002
- 6) 折笠秀樹 : 糖尿病診療に役立つエビデンス. 新薬と臨床 51 (7) : 557-575, 2002
- 7) 折笠秀樹 : 標準化へ向けてのRCT論文の書き方(1). 薬理と治療 28(5) : 363-366, 2000
- 8) Moher D, Schulz KF, Altman D, for the CONSORT Group : The CONSORT statement. *JAMA* 285 : 1987-1991, 2001
- 9) 野村英樹ほか (訳) : 「生物医学雑誌への統一投稿規程」(2001年10月改訂版). 医学のあゆみ 201 (10) : 790-798 ; 862-867, 2002 (医歯薬出版のHP : <http://www.ishiyaku.co.jp/magazines/ayumi.html> よりダウンロード可能. 原典 : <http://www.icmje.org/index.html>)

## Striatal infarcts mimicking frontotemporal dementia: a case report

Y. Nishio<sup>a</sup>, Y. Nakano<sup>a</sup>, K. Matsumoto<sup>a</sup>, M. Hashimoto<sup>a</sup>, H. Kazui<sup>a</sup>, N. Hirono<sup>a</sup>, K. Ishii<sup>b</sup> and E. Mori<sup>a</sup>

Divisions of <sup>a</sup>Clinical Neurosciences and <sup>b</sup>Neuroimaging Research, Hyogo Institute for Aging Brain and Cognitive Disorders, Himeji, Japan

### Keywords:

anterior thalamic peduncle, caudate nucleus, fronto-subcortical circuits, stereotyped and disinhibited behavior

We described a patient with bilateral striatal infarcts, in whom stereotyped and disinhibited behaviors were insidiously emerged over 2 years mimicking frontotemporal dementia (FTD). A positron emission tomography with 18-fluorodeoxy glucose showed a hypometabolism in the frontal lobes, basal ganglia, and thalami. The peculiar behavioral alterations remained unchanged for the following 7 years, suggesting that the disease is not degenerative but of vascular origin. A disruption of the fronto-subcortical circuits at the level of the striatum or the anterior thalamic peduncle is attributable to the FTD-like behavioral and cognitive syndrome.

Received 17 November 2002

Accepted 10 March 2003

Frontotemporal dementia (FTD) is characterized by pronounced changes in personality and social conduct, including breaches of interpersonal etiquette, tactlessness, and disinhibition, with relative preservation of instrumental functions of perception, spatial skills, praxis, and memory (Neary *et al.*, 1998). The similar behavioral syndrome may occur in patients with vascular lesions affecting the subcortical structures, especially the caudate nucleus and thalamus (Mendez *et al.*, 1989; Cummings, 1993; Bhatia and Marsden, 1994; Mori *et al.*, 1999). When the behavioral changes emerge abruptly after a stroke attack with neuroimaging evidence of lesions in the relevant regions, the differentiation from FTD may not be a big issue. The patient described here, whose magnetic resonance imagings (MRIs) showed bilateral striatal infarcts, developed behavioral changes of insidious onset and gradual progression over 2 years including stereotyped and disinhibited behaviors and decreased spontaneity. These peculiar behavioral alterations and findings on functional neuroimaging remained unchanged for the following 7 years, suggesting that the disease is not degenerative but of vascular origin.

### Case report

A 65-year-old, right-handed hypertensive man was referred to our hospital in May 1994 because of behavioral and personality changes. The patient was serious and diligent, and was a bank branch manager. There was no personal and familial history of mental or neurological illness. In April 1992, he was forcedly retired because of increasing mistakes and failures on

business. He repeatedly mailed important postal matters to wrong addresses and misunderstood customers' orders. His wife could not understand the reason of dismissal, as she noticed no behavioral changes in his domestic life. In June 1992, he suddenly developed a left hemiparesis and was admitted to a nearby hospital, where a diagnosis of cerebral infarction was made. In the autumn of 1992, his wife noticed that he often picked up cigarette butts on the street and spoke to strangers with undue familiarity. Then, it became apparent that he used up his retirement allowance (about 30 million yen) on mah-jong games in a few months. He gradually became restless, and paced in and around his house. He was once arrested for stealing cigarettes from a car on the way of pacing.

The patient was admitted to our hospital for examination in May 1994. A neurological examination revealed bilateral hyperreflexia more accentuated on the left side, extensor planter response on the left side, exaggerated jaw jerk, and positive snout reflex. Grasping, sucking or utilization behavior were not elicited. He lacked insight for his own problems. He paced around the ward in the daytime everyday, and rarely had a conversation with other patients. He often intruded his body on other patients at midnight, and occasionally requested female patients to kiss him. Sexual approaches to female hospital staffs were also often noted. Results of neuropsychological tests were summarized in Table 1. The Mini Mental State Examination score was 26. Whilst the memory, language, and visuospatial functions were relatively well preserved, the performance on the tests for frontal executive functions was considerably declined. Denkfaulheit (answering questions without thinking over) was apparent during neuropsychological testing. Magnetic resonance imagings of the brain showed a mild frontotemporal atrophy with dilatation of the anterior horn of the lateral ventriculi. Infarcts were noted in the bilateral striatocapsular

Correspondence: Etsuro Mori MD, Department of Humanities and Sciences, Kobe Gakuin University, 518 Arise Ikawadanicho, Nishi-ku, Kobe 651-2180, Japan (fax: 81-78-974-5689; e-mail: e-mori@human.kobegakuin.ac.jp).

	1994	2001
Mini Mental State Examination	26	27
Alzheimer's Disease Assessment Scale	12.3	10.3
WAIS-R VIQ/PIQ/FIQ	100/82/92	–
Western Aphasia Battery Aphasia Quotient	86	–
WMS-R Verbal/Visual/General/Delayed Memory Index	84/95/88/77	–
Rey's complex figure copy	36/14	–
Wisconsin Card Sorting Test (categories achieved)	1	0
Verbal fluency, category cue/phonetic cue (/min)	13/2	6/6
Porteus maze	10.5 years	9.5 years
Trail making A/B (sec)	47/161	82/401
Raven's color progressive matrices	24	28

**Table 1** Performance on neuropsychological tests of the baseline and 7 years later

WAIS-R VIQ/PIQ/FIQ, Wechsler Adult Intelligence Scale-Revised Verbal/Performance/Full-scale Intelligence Quotient; WMS-R, Wechsler Memory Scale-Revised.



**Figure 1** Axial and coronal T1-weighted MR (three-dimensional spoiled gradient echo) and 18-fluorodeoxy glucose PET images of the baseline (upper) and 7 years later (lower). MRI demonstrates multiple small infarcts affecting the striatum and thalamus bilaterally, and a moderate atrophy of the frontal lobe. PET demonstrates glucose hypometabolism in the frontal lobe, temporal lobe, striatum and thalamus, more accentuated in the right hemisphere. No remarkable changes are noted on MRI and PET in 7-year interval.

regions and thalamus. The striatocapsular infarcts affected the head of the caudate nucleus, anterior part of the putamen, and the anterior limb of the internal capsule, and the thalamic infarcts involved the ventrolateral nucleus, dorsomedial nucleus, and internal medullary lamina. White matter hyperintensity was also noted on T2-weighted images. Positron emission tomography (PET) with  $^{18}\text{F}$ -fluorodeoxy glucose showed a hypometabolism in the frontal lobes, basal ganglia, and thalami, more accentuated in the right side (Fig. 1).

The patient was treated with a low dose aspirin. His behavioral traits remained unchanged for following 7 years. In neuropsychological and neuroimaging follow-up studies performed in 2001, there are no

conspicuous deteriorations in neuropsychological tests except for the Trail Making test performance. Remarkable changes were noted neither on MRI except for a slight increment of brain atrophy and white matter hyperintensities, nor in cerebral glucose metabolism (Fig. 1).

## Discussion

The changes of personality and social conduct in our patient, i.e., stereotyped behaviors, disinhibition, and decreased spontaneity were those typically seen in patients with FTD. The results of neuropsychological tests, i.e., a low performance on executive function tests



with preserved memory, visuospatial, and language functions were similar to the features in patients with FTD. Together with gradual progression over the initial 2 years, all the signs and symptoms were compatible with a clinical diagnosis of FTD (Neary *et al.*, 1998). Glucose hypometabolism in the frontal lobes, basal ganglia, and thalami on PET was also consistent with the findings in FTD (Ishii *et al.*, 1998). Nevertheless, the subcortical infarcts can explain the peculiar behavioral changes and cognitive deficits. The lack of progression over the subsequent 7 years in behavioral impairment and glucose hypometabolism were not supportive of the diagnosis of degenerative disease. In FTD, behavioral disturbances evolve with time, and changes on functional brain imaging increase (Pasquier *et al.*, 1999). Moreover, cognitive decline was not evident on the global cognitive tests as well as the frontal executive tests. This stable cognition over 7 years is unlikely to be the course of FTD. Increase of brain atrophy, which was noted in slight degree in our patient on repeated MRIs, is not always the hallmark of degenerative diseases but seen in aging and vascular dementia (O'Brien *et al.*, 2001). Although each of small subcortical infarcts did not cause abrupt onset of noticeable symptoms in our patients, they occurred sporadically and consequently result in gradual development of symptoms. Furthermore, difficulty in identifying subtle behavioral and cognitive changes may have made stepwise progression of symptoms unclear.

The subcortical structures such as the thalamus, striatum, globus pallidus are central components of the fronto-subcortical circuits. Small subcortical infarcts involving them and/or the neural pathways connecting these structures disrupt the circuits and may lead to cognitive and behavioral impairment with striking frontal lobe features. Five circuits that link the frontal lobe, basal ganglia, and thalamus are currently recognized: (i) a 'motor' circuit originating in the supplementary motor area, (ii) an 'oculomotor' circuit with origins in the frontal eye fields, (iii) 'dorsolateral prefrontal' circuit, (iv) 'lateral orbitofrontal' circuit, and (v) 'anterior cingulate (limbic)' circuit. The prototypic structure of all circuits is an origin in the frontal lobes, projection to striatal structures, connections from striatum to globus pallidus and substantia nigra, projections from these two structures to specific thalamic nuclei, and a final link back to the frontal lobe (Alexander *et al.*, 1990; Parent, 1990). In frontal damage, three distinct neurobehavioral syndromes have been recognized, and each corresponds to damage of one of the three prefrontal-subcortical circuits: (i) a dorsolateral prefrontal syndrome with neuropsychological deficits, including decreased verbal

and design fluency, abnormal motor programming, impaired set shifting, reduced learning and memory retrieval, and poor problem solving; (ii) an orbitofrontal syndrome with prominent disinhibition and irritability; and (iii) an anterior cingulate syndrome with apathy and diminished initiative (Cummings, 1993). The neurobehavioral syndromes equivalent to those observed with frontal lobe lesions may occur also in patients with striatal, pallidal, or thalamic lesions. Disinhibition syndrome including talkativeness, jocularity, restlessness, irritability, agitation and sexual bluntness was reported in caudate and paramedian thalamic infarcts (Bhatia & Marsden, 1994; Mendez *et al.*, 1989; Cummings, 1993). Stereotyped activities with compulsive and obsessive behavior were previously reported in bilateral basal ganglionic lesions (Laplante *et al.*, 1989). In a review of the behavioral and motor consequences of lesions of the basal ganglia, abulia and disinhibition were noted as the commonest behavioral disturbance in patients with caudate lesions (Bhatia & Marsden, 1994). However, the symptomatology associated with vascular subcortical lesions have been hardly mentioned in the context of that in FTD. In our patient, a disruption of all the three fronto-subcortical circuits is assumed to cause his FTD-like behavioral and cognitive syndrome. In particular, lesions in the bilateral caudate nuclei are likely to play a key role. The dorsolateral part of the caudate head is involved in the dorsolateral prefrontal circuit for executive functions and the ventromedial part of the caudate nucleus is involved in the orbitofrontal inhibitory circuit. Furthermore, a damage to the anterior thalamic peduncles, which conveys diverse connection fibers to all three frontal cortices through the anterior limbs of internal capsule, might be involved. Functional brain imaging may document a functional deprivation of the remote cortices (Mori *et al.*, 1999). In patients with lesions disrupting the cortico-subcortical circuits, hypometabolism or hypoperfusion may be demonstrated in the relevant cortical area on functional brain imaging. The frontal involvement demonstrated on PET was functional brain imaging evidence of subcortical lesions leading to a functional deprivation of the remote cortices in our patient.

Finally, the clinical implication of the present case is evident. In diagnosis of vascular dementia, stress has been laid on differentiation from Alzheimer's disease. However, as described here, subcortical structures have close connections with frontal cortices, and damage to those structures often causes frontal lobe dysfunctions. FTD is an important condition that should be taken into consideration in differential diagnosis of vascular dementia.

## References

- Alexander GE, Crutcher MD, DeLong MR (1990). Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, 'prefrontal' and 'limbic' functions. *Progr Brain Res* 85:119-146.
- Bhatia KP, Marsden CD (1994). The behavioural and motor consequences of focal lesions of the basal ganglia in man. *Brain* 117:859-876.
- Cummings JL (1993). Frontal-subcortical circuits and human behavior. *Archiv Neurol* 50:873-880.
- Ishii K, Sakamoto S, Sasaki M *et al.* (1998). Cerebral glucose metabolism in patients with frontotemporal dementia. *J Nucl Med* 39:1875-1878.
- Laplane D, Levasseur M, Pillon B *et al.* (1989). Obsessive-compulsive and other behavioural changes with bilateral basal ganglia lesions. A neuropsychological, magnetic resonance imaging and positron tomography study. *Brain* 112:699-725.
- Mendez MF, Adams NL, Lewandowski KS (1989). Neurobehavioral changes associated with caudate lesions. *Neurology* 39:349-354.
- Mori E, Ishii K, Hashimoto M *et al.* (1999). The role of functional brain imaging in evaluation of vascular dementia. *Alzheimer Dis Assoc Disord* 13(Suppl. 3):S91-S101.
- Neary D, Snowden JS, Gustafson L *et al.* (1998). Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 51:1546-1554.
- O'Brien JT, Paling S, Barber R *et al.* (2001). Progressive brain atrophy on serial MRI in dementia with Lewy bodies, AD, and vascular dementia. *Neurology* 56:1386-1388.
- Parent A (1990). Extrinsic connections of the basal ganglia. *Trends Neurosci* 13:254-258.
- Pasquier F, Lebert F, Lavenu I, Guillaume B (1999). The clinical picture of frontotemporal dementia: diagnosis and follow-up. *Dementia Geriatric Cogn Disord* 10(Suppl. 1):10-14.

## Evolving atherosclerosis at carotid and intracranial arteries in Japanese patients with ischemic heart disease: a 5-year longitudinal study with MR angiography

T. Uehara<sup>a</sup>, M. Tabuchi<sup>a</sup>, E. Mori<sup>b</sup> and A. Yamadori<sup>c</sup>

<sup>a</sup>Neurology Service, <sup>b</sup>Institute for Aging Brain and Cognitive Disorders, Hyogo Brain and Heart Center at Himeji; and <sup>c</sup>Section of Neuropsychology, Division of Disability Science, Tohoku University Graduate School of Medicine, Japan

### Keywords:

atherosclerosis, carotid artery diseases, cerebral arteries, coronary artery disease, magnetic resonance angiography, progression

Received 17 September 2002  
Accepted 26 March 2003

Progression of atherosclerosis at extracranial carotid and intracranial arteries in patients with ischemic heart disease (IHD) is not well defined. We carried out a 5-year longitudinal study with magnetic resonance angiography (MRA) of patients with IHD to assess the incidence of progression of atherosclerosis at extracranial carotid and intracranial arteries and to determine predictors of the progression. We previously performed carotid and intracranial MRA on 67 patients who had received selective coronary angiography for the clinical diagnosis of IHD. Of these 67 subjects, 41 patients gave informed consent to undergo MRA reexaminations to evaluate changes of extra- and intracranial arteries over a 5-year period. The degree of stenosis was divided into five grades depending on the narrowness of the arteries, i.e. normal, mild, moderate, severe and occluded. The average of follow-up period with MRA examination was 58.8 months. The progression of atherosclerosis, as defined as an increase of one grade of the stenosis rating, including both the exacerbation of pre-existing stenosis and the appearance of new stenotic lesions, were found in five patients (12.2%) for the cervical carotid artery and in only one patient (2.4%) for the intracranial artery. A multiple logistic regression analysis revealed that baseline carotid artery stenosis ( $P = 0.008$ ), age ( $P = 0.047$ ), and coronary events during the follow-up period ( $P = 0.048$ ) were significant and independent predictors of progression of carotid atherosclerosis. In conclusion, our findings suggest that follow-up evaluation of the carotid artery is indicated for patients with IHD in whom carotid artery stenosis was detected on an initial examination. Further study is needed with larger numbers of patients to confirm these findings.

### Introduction

Patients with ischemic heart disease (IHD) are more probable to have stroke (Kannel *et al.*, 1983). In patients with IHD, attention must be paid not only to secondary prevention of cardiac attack but also to stroke. Previous studies (Crouse *et al.*, 1986; Craven *et al.*, 1990; Howard *et al.*, 1990) have shown a correlation between the extent of coronary atherosclerosis and carotid atherosclerosis. In a 2-year longitudinal study of 50 Japanese patients with IHD using carotid ultrasonography, a strong association was found between the extent of coronary atherosclerosis and the progression of carotid atherosclerosis (Tanaka *et al.*, 1992). However, long-term changes of carotid and intracranial artery in patients with IHD have not been well defined.

In our previous cross-sectional study, we evaluated occlusive lesions in the carotid and intracranial arteries by using magnetic resonance angiography (MRA) in 67 patients who received selective coronary angiography for evaluation of IHD (Uehara *et al.*, 1996). Stenoses of more than 25% narrowing of the diameter were found in 15 patients (22.4%) in the extracranial carotid arteries, and in 11 patients (16.4%) in the intracranial arteries. Most of the stenotic lesions were mild. In the present study, to determine the progression of carotid and intracranial artery disease in patients with IHD, we reevaluated the cohort examined in the previous study after a 5-year interval.

### Materials and methods

#### Subjects and study design

Amongst patients who received selective coronary angiographic examinations in our hospital between October 1993 and February 1994 for evaluation of

Correspondence: Toshiyuki Uehara MD, Neurology Service, Hyogo Brain and Heart Center at Himeji, 520 Saisho-ko, Himeji, 670-0981, Japan (fax: 81-792-95-8199; e-mail: tuehara@hbhc.jp).

myocardial infarction (MI) or angina pectoris (AP), 67 subjects who agreed to participate in the study program were given MRA studies of extracranial carotid and intracranial cerebral arteries within 3 months of the coronary angiography. They included 49 males and 18 females. At the time of the initial MRA examination, mean age of the subjects was 60.1 years ranging from 40 to 78 years. Forty-five patients had MI. None had a history of stroke.

Five years later, all subjects except for one patient who moved from our area and one who died from pancreatic cancer were invited to a reexamination between January and June 1999. Patients were contacted during their regular visit to the clinic or by telephone. Information regarding death, cerebrovascular events, and coronary events during the follow-up period was obtained from hospital records and telephone interviews. Coronary events comprised acute MI or myocardial revascularization procedures such as percutaneous transluminal coronary angioplasty or coronary artery bypass grafting. One patient who received a carotid thromboendarterectomy for asymptomatic severe carotid artery stenosis on the initial MRA examination was excluded from the study. Twenty-three patients refused the reevaluation with MRA. The remaining 41 patients gave informed consent and were reexamined with MRA. The baseline characteristics of the 41 subjects who were reevaluated with MRA and 26 subjects who were not reevaluated are listed in Table 1. Of the 41 subjects, 33 received antiplatelet agents, and 12 received statins throughout the period.

**Table 1** Baseline characteristics of patients that were reevaluated with magnetic resonance angiography (MRA) and patients that were not

	Follow-up MRA		P-value
	Evaluated (n = 41)	Not evaluated (n = 26)	
Age at initial MRA (mean $\pm$ SD)	60.5 $\pm$ 9.3	59.4 $\pm$ 8.9	0.618
Sex (male/female)	29/12	21/5	0.266
Hypertension (present/absent)	14/27	5/21	0.149
Diabetes mellitus (present/absent)	11/30	9/17	0.830
Hyperlipidemia (present/absent)	12/29	12/14	0.127
Smoking habit (present/absent)	19/22	15/11	0.257
Baseline coronary artery disease			
Two- or three-vessel disease (present/absent)	18/23	11/15	0.550
Baseline carotid artery stenosis			
Normal, mild, moderate, severe	32, 6, 3, 0	20, 5, 0, 1	0.958
Baseline intracranial artery stenosis Normal, mild, moderate, severe	35, 5, 1, 0	21, 3, 0, 2	0.223

## MRA examinations

Initial and second MRA examinations were performed with the identical equipment and acquisition technique. Images were acquired by a three-dimensional, Fourier transform time-of-flight method with a spoiled gradient-echo sequence with a 1.0 T MR system (Magnetom Impact, Siemens, Erlangen, Germany). Details of MRA acquisition and image reconstruction are described elsewhere (Uehara *et al.*, 1996). Two investigators (T.U., M.T.) who were blinded to all clinical information independently reviewed the MRA. When the judgment of the two readers was inconsistent, a decision was entrusted to a third reviewer (E.M.). Degree of stenosis was divided into five grades depending on the narrowness of the arteries, i.e. normal, mild, moderate, severe and occluded. Less than 25% reduction of an arterial diameter was graded normal, 25–49% reduction mild, 50–74% reduction moderate, 75–99% reduction severe and no opening as occluded (Uehara *et al.*, 1994, 1995). State of the extracranial portion of the internal carotid arteries (ICA) was evaluated at the bifurcation site by the carotid MRA and that of the intracranial arteries at the intracranial portion of the ICA, at the horizontal portion of the middle cerebral arteries (MCA), and at the basilar artery (BA) by the intracranial MRA. To measure the percentage stenosis of the extracranial portion of the ICA, we compared the diameter of maximal stenosis with that of the normal-appearing proximal ICA beyond the carotid bulb [North American Symptomatic Carotid Endarterectomy Trial (NASCET) Steering Committee, 1991]. As our previous study indicated that the subclassification between moderate and severe stenoses in the intracranial arteries based on degree of the reduction of the signal column width was not practical, moderate stenosis was defined as a more than 50% reduction of the signal column width without the discontinuity of the signal and severe stenosis as the discontinuity of the signal. This stenosis rating scheme has been validated in comparative studies with conventional angiography (Uehara *et al.*, 1994, 1995). The accuracy (sensitivity and specificity) was reported previously and was acceptably high; 100% of sensitivity and specificity for mild extracranial ICA stenosis and 100% of sensitivity and >85% specificity for mild stenosis of the intracranial ICA and MCA. The interrater reliability of the rating is as high as that of conventional angiography. In this study, 'atherosclerosis progression', implying both the exacerbation of pre-existing stenosis and the appearance of new stenotic lesions, was defined as an increase of more than one grade of the stenosis rating.