

団という意味でコホートとよばれる)を分類し、その後のエンドポイントの発生を調べることから、前向き研究(prospective study)ともよばれる(図1)。また、コホート研究は疾病発生を前向きに捉えることから、複数の疾病について曝露の影響を調べることができるという利点もある。しかし、その一方でコホート研究の実施には時間も費用もかかり、しかもそのほとんどが心筋梗塞を発生しない健康な人を追跡するために費やされる、という欠点もある(表1)。

喫煙と心筋梗塞発生との関係をコホート研究で調べた結果を示す(表2)。心筋梗塞を発生していない対象者2469人を研究開始時の喫煙状況で分類し(喫煙者1485人、非喫煙者984人)、10年間の追跡を行った結果、心筋梗塞の発生がそれぞれ86人、9人観察されたとする。喫煙群における心筋梗塞発生リスクは、 $86/1485=0.058$  (5.8%)であり、非喫煙群でのそれは、 $9/984=0.009$  (0.9%)である。ここで、リスクとは、あるイベント、特に好ましからざる結果が起きる確率のことである。この各群のリスクから、曝露(喫煙)効果の指標として、リスク差= $0.058-0.009=0.049$ 、リスク比= $0.058/0.009=6.44$ を求めることができる。ここで、この結果が喫煙状況をランダム化した研究から得られた場合には、リスク差0.049の解釈は、「100人が喫煙すると、喫煙しなかった場合に比べて心筋梗塞の発生が5人増える」となるのに対して、コホート研究では、喫煙状況を介入していないので、「喫

表1 コホート研究とケース・コントロール研究の長所と短所

	コホート研究	ケース・コントロール研究
観察の時間的順序と因果の方向	同じ	逆
まれな曝露	適している	適していない
まれな疾病	適していない	適している
疾病	複数調べられる	1つのみ
曝露	1つのみ	複数調べられる
費用と時間	かかる	かからない
サンプルサイズ	多い	少ない
情報の精度	高い	低い

表2 コホート研究の結果(仮想例)

喫煙の有無	心筋梗塞発生		合計
	あり	なし	
あり	86	1399	1485
なし	9	975	984
			2469

喫煙群の発生リスク= $86/1485=0.058$  (5.8%)  
 非喫煙群の発生リスク= $9/984=0.009$  (0.9%)  
 リスク差= $0.058-0.009=0.049$ 、  
 リスク比= $0.058/0.009=6.44$

煙群は非喫煙群に比べて心筋梗塞発生が100人あたり5人多い」となることに注意が必要である(喫煙群には、例えば、運動不足の人が多くいる可能性もあり、5人の心筋梗塞の増加は喫煙がすべての原因とはいきれない)。同様に、リスク比6.44の正確な解釈は、「喫煙群は非喫煙群に比べて心筋梗塞発生が約6.4倍多い」となる。

#### ケース・コントロール研究

ケース・コントロール研究は、因果の順序と研究実施の時間的順序が逆転していて、わかりにくい研究デザイン

であるが、その原理は実はコホート研究と同じである。喫煙による心筋梗塞発生リスクをケース・コントロール研究で調べることを考える。その研究デザインの概要は、「あるコホート集団内で、すでに心筋梗塞を発生した対象者(ケース)と心筋梗塞を発生していない対象者(コントロール)に対して、過去の喫煙状況、心筋梗塞発生に対するリスク因子を調べる」というものである。ケース・コントロール研究は、過去にさかのぼって曝露状況を調べることから、後ろ向き研究(retrospective study)ともよばれる(図2)。こ

で、ケース・コントロール研究におけるコントロールという用語は、「当該の疾患を発生していない人」という意味であり、コホート研究における「対照群」<sup>1)</sup>とは異なる意味で用いられることに注意が必要である。このように、ケース・コントロール研究は、コホート研究とは違ってすでに心筋梗塞を発生している人たちを対象とするので、時間も費用も節約できるという利点があるが、結果からさかのぼって原因を調べることに由来する様々なバイアスが入りやすいという欠点がある(表1)。

喫煙と心筋梗塞発生との関係をケース・コントロール研究で調べた結果を示す(表3)。この結果は、コホート研究の結果(表2)において、ケースのデータは全例使用し、コントロールのデータは5%ランダムにサンプルした結果である。コントロールに関して5%ランダムにサンプルしただけなので、コホート研究とほぼ同じ結果が得られるはずである。しかし、喫煙群、非喫煙群の心筋梗塞発生リスクは、それぞれ  $86/156=0.551$ ,  $9/58=0.155$ , リスク差=0.396, リスク比=3.55であり、コホート研究の結果とまったく異なる。これは、ほとんどすべてのケース・コントロール研究では、疾患を発生していないコントロールのデータ(あるいは、ケースのデータ)を全例調べることはできず、一部のサンプルを用いざるを得ないため(しかも、サンプリング割合も不明である)、喫煙群、非喫煙群の人数を固定できないからである(表3の行の合計を出すことには意味がなく、デザインで固定されるのは列

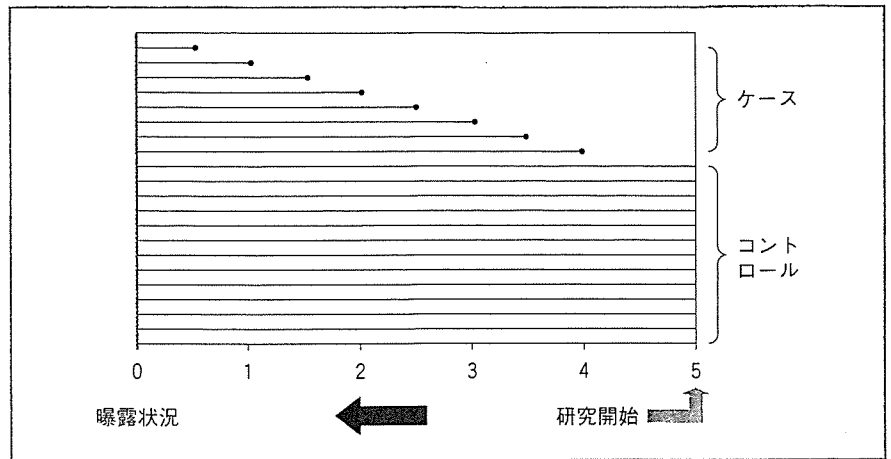


図2 ケース・コントロール研究の概要

図中の1本の線は対象者の追跡を示し、●はイベントの発生(ケース)を意味する

表3 ケース・コントロール研究の結果(仮想例)

喫煙の有無	心筋梗塞発生	
	あり	なし
あり	86	70
なし	9	49
合計	95	119

ケースでの曝露オッズ=86/9, コントロールでの曝露オッズ=70/49  
オッズ比=(86×49)/(70×9)=6.69

側のケースとコントロールの合計である)。したがって、ケース・コントロール研究では、各群で心筋梗塞を発生した人が何人いたかはわからず、疾患発生リスクを求めることはできない。それではケース・コントロール研究は何の役に立つのだろうか。その答えは、ケース・コントロール研究から求められるオッズ比が、コホート研究から求められるリスク比を近似するからである。オッズとは、あるイベントが起きる確率と起きない確率の比であり、

今の例では、ケースでの曝露オッズは86/9, コントロールでのそれは70/49, したがって、オッズ比=(86×49)/(9×70)=6.69となる。この曝露オッズ比は、コホート研究におけるリスク比と同じように解釈でき、「喫煙群は非喫煙群に比べて心筋梗塞発生が約6.7倍多い」となる。

ケース・コントロール研究から求められるオッズ比をコホート研究から求められるリスク比の代用として用いるためには条件が必要である。コホート

研究からのリスク比(表 2)は、 $(86/1485)/(9/984)=(86/9)/(1485/984)=6.33$ であり、ケースでの曝露オッズ(86/9)と全コホートでの曝露オッズ(1485/984)の比である。一方、ケース・コントロール研究からのオッズ比(表 3)は、 $(86/9)/(70/49)=6.69$ であり、ケースでの曝露オッズ(86/9)とコントロールでの曝露オッズ(70/49)の比である。明らかに、それぞれの分母の「全コホートでの曝露オッズ」と「コントロールでの曝露オッズ」がほぼ等しい場合に、ケース・コントロールオッズ比をコホートリスク比の代用として用いることができる。このことが成立するのは、もともとの全コホート集団と疾病を発生しなかったコントロール集団がほぼ同じ集団の場合、すなわち当該の疾病(イベント)がほとんど観察されない場合である。この条件は「まれな疾病の仮定」とよばれる。なお、前向き研究においても(疾病発生)オッズ比を求めることができるが、前向き研究ではリスク比やリスク差を用いて発生リスクを直接評価できるので、オッズ比を積極的に用いる理由はない。また、一般に前向き研究においてオッズ比を計算すると、リスク比より大きな値となることが多く、オッズ比は曝露効果を過大評価しがちでもある。したがって、オッズ比は発生リスクを直接求めることができないケース・コントロール研究においてのみリスク比の代用として使用すべきである。

### その他の研究デザイン

コホート研究とケース・コントロー

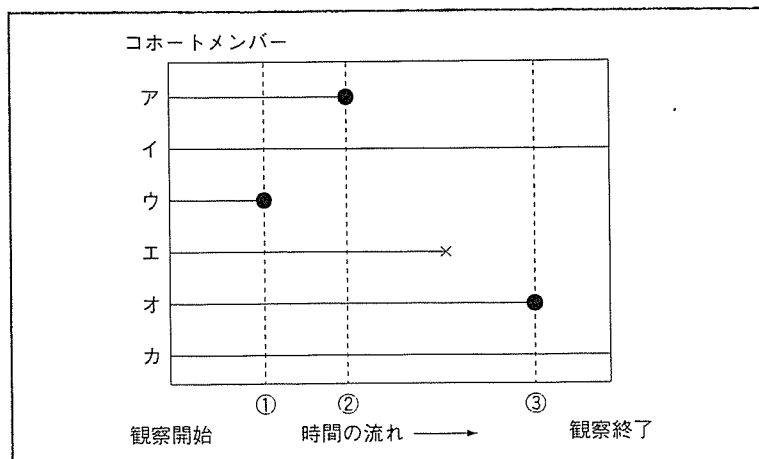


図 3 コホートメンバーの追跡状況

図中の 1 本の線が対象者の追跡を示す

●はイベントの発生(ケース)、×は観察打ち切りを意味する

ル研究の利点、欠点は相互に補完的である(表 1)。ケース・コントロール研究では、コントロール選択の際のバイアス、思い出しバイアスなどの統計解析では対処することのできない種々のバイアスが混入する可能性が高い。一方、コホート研究は、ケース・コントロール研究で問題となる種類のバイアスの影響は受けにくい、疾病がまれな場合には大規模コホートを長期間にわたって観察しなければならず、コストのほとんどが健常者の追跡に費やされてしまうという問題がある。最近、コホート研究を途中からケース・コントロール研究に転換するデザイン(コホート内ケース・コントロール研究)が提案され、実際の応用例も増えつつある。このコホート研究とケース・コントロール研究を混合したデザインは、あくまでもコホート研究の枠組みの中で行われるため、ケース・コントロー

ル研究に特有のバイアスを避けることができ、曝露や主要なリスク因子に関する情報は全ケース(あるいは、全ケースからのサンプル)とコントロールからのサンプルについてのみ収集・解析すればよいので、かなりのコストの減少が期待できる。コントロールの選択方法が古典的なケース・コントロール研究とは異なる代表的な 2 つのデザイン<sup>2)</sup>を以下で紹介する。

### 1 ネスティッド・ケース・コントロール研究

6 人のコホートメンバーの追跡を考える(図 3)。ネスティッド・ケース・コントロール研究では、コホート研究を実施しているので、例えば、図 3 の(ウ)が疾病を発生した時点①、(ア)が疾病を発生した時点②などの疾病発生時点の情報を利用してコントロールの選択を行う。コントロールの

選択方法にはいくつかのオプションがあるが、最も一般的なのが「ケースが発生した時点のリスク集団から1人またはそれ以上のコントロールをランダムに選択する」という時点マッチングとよばれる方法である。ケースグループとしては全ケースを用いることが多いが、全ケースからのサンプルでもかまわない。図3において、(ウ)が疾病を発生した時点①では、リスク集団は(ア, イ, ウ, エ, オ, カ)であり、(ア)が疾病を発生した時点②では、(ア, イ, エ, オ, カ), (オ)が疾病を発生した時点③では、(イ, オ, カ)がリスク集団なので、それぞれの時点でリスク集団からコントロールをランダムに選ぶことになる。

コントロールの選択の際には、将来疾病を発生する者やケース本人もケースのコントロールになりうることに注意しなければならない。時点①を例にとると、(ア)と(オ)は将来のケースで、(ウ)はケース本人であるが、いずれの対象者も(ウ)のコントロールになる可能性がある。また、同一の対象者が複数の時点でコントロールとして選択される可能性があり、(イ, オ, カ)は時点①, ②, ③すべてでコントロールとして選択されることも起こりうる。このようにケース発生時点でマッチして収集されたケースとコントロールの解析には、通常のマッチングを行った場合の解析方法(例えば、マンテル・ヘンテル法,あるいは条件付きロジスティック回帰など)をそのまま用いてオッズ比を推定し、信頼区間を求めることができる(したがって、

統計パッケージを用いて簡単に計算ができる)。しかも、この場合のオッズ比は正確に疾病発生率の比を推定しており、これには「まれな疾病の仮定」を必要としない。

## 2 ケース・コホート研究

前述のネステッド・ケース・コントロール研究では、ケースが発生した時点でコントロールを選択するため、特定のリスク因子と複数の疾病との関係を調べたい場合、調べる疾病の数だけコントロールグループが必要になる。しかし、コホート研究では特別な考慮をしなくても、1つのリスク因子が複数の疾病に影響するかどうかを調べることができる。コホート研究の枠組みの中でケース・コントロール研究を実施しているので、複数の疾病を対象とする場合でもコントロールグループは1つでいいはずである。このような考え方から提案されたのがケース・コホート研究である。

ケース・コホート研究では、コントロールは観察開始時のリスク集団、つまり全コホートメンバーからランダムに選択される(このコントロールグループをサブコホートとよぶ)。ケースグループは全ケースを用いても、ケースのサンプルを用いてもよい。明らかに、このデザインでもコントロールグループの中に将来疾病を発生する者が含まれる可能性があり、解析の際には同一人物がケースグループとコントロールグループに同時に属するということが起こり得る。図3の例では、ケースグループは(ア, ウ, オ)また

はそのサンプルで構成され、コントロールグループは全コホートメンバー(ア, イ, ウ, エ, オ, カ)からランダムに選択される。

このデザインのもとでは、ケースサンプルは古典的なケース・コントロール研究と同様に全ケースからのランダムサンプルであるが、コントロールサンプル(サブコホート)は全コホートからのランダムサンプルとなっている。このサブコホートのデータから曝露割合を推定することが可能なので、ケース・コホートデータから曝露オッズ比を求めると、正確にリスク比に一致することが知られている。また、観察期間を通して曝露割合が一定の場合、この曝露オッズ比は疾病発生率の比に一致することも知られている。このどちらの場合も、「疾病がまれである」という仮定とは無関係に成立する。オッズ比の信頼区間の計算は多少複雑であるが可能である。ただし、現在のところ汎用統計パッケージには組み込まれていない。

## おわりに

ランダム化比較試験では、実験であるがためにプロトコルの作成、インフォームド・コンセントの実施などが求められ、新医薬品市販前の臨床試験では法律で定められてきている。しかし、観察研究であっても研究の手順を明確にするためにプロトコルを作成すべきであるし、凍結保存された血液サンプルから遺伝子を調べる場合では、リスクが患者個人だけではなく患者の家系全体におよぶ可能性があることか

ら、臨床試験よりも厳格なインフォームド・コンセントが必要な観察研究すらある。したがって、観察研究であっても、実験研究と同様にプロトコルを作成するという原則は変わらない。科学的でできる限りバイアスの少ない疫

学研究を実施することは当然であるが、倫理的にも妥当な研究を実践していくことが不可欠である。

#### 文 献

1) 松山 裕：コントロール(対照群)の

考え方. Surgery Frontier 11 : 76-80, 2004

2) 佐藤俊哉：疫学研究のデザインと相対リスクの推定. 統計数理 40 : 173-184, 1992

## 3) 統計的仮説検定と治療効果の推定

東京大学大学院医学系研究科生物統計学

松山 裕

Yutaka Matsuyama

(助教授)

### Summary

多くの臨床研究では、治療効果の大きさを求めることに研究の主たる目的がある。そのためには、疾病の発生を定量的にとらえ、治療によって疾病がどの程度改善されるかをデータから求めなければならない。本稿では、疾病発生の指標と治療効果の指標について述べ、治療効果に対する統計的仮説検定の考え方を解説する。さらに、仮説検定と表裏一体の関係にある治療効果の推定と研究の精度を表す信頼区間について解説する。

### Key Words

治療効果の指標, 統計的仮説検定, P値, 95%信頼区間, 臨床的有意差

### 治療効果をはかる

#### 1 疾病発生の指標

疾病頻度のとらえ方には大きく有病 (prevalence) と発生 (incidence) の2つがある。前者の有病は、病気であるという状態に注目した指標で、「ある時点において、研究対象としている疾病に罹っていること」を指す。集団検診やスクリーニングで発見された疾病は検診時の有病状況であるし、剖検で確定診断された疾病は死亡時の有病状況である。疾病の有病数を把握することは、公衆衛生行政の観点からは重要な問題である。一方、後者の発生 (発症, 罹患) は、イベント (疾病, 死亡) が起きたかどうか注目した指標で、疾病の原因を調べるために重要な指標である。「研究開始時に研究対象としている疾病に罹っておらず、将来その疾病に罹る可能性がある」状態のことを「at risk」とよぶとすると、疾病発生とは、「at riskにある人を追跡して、新たに疾病に罹ること」を指す。

集団中の疾病発生の指標としては、割合 (proportion) で示す場合と率 (rate) で示す場合の2通りの考え方がある。割合とは、分子が分母に含まれる比であり、次元 (単位) をもたず、必ず0から1の間の値をとる量である。一方、率とは単位時間当たりの変化の速さを表す比であり、次元が存在し、0から無限大の値をとる量である。

簡単な例で両者の違いを示す (図1)。A, B, C, Dの4人からなるリスク集団 (population at risk) を1年間追跡した結果、AとDに対象疾病の発生がなく、BとCはそれぞれ3ヵ月目、9ヵ月目に疾病を発生したとする。このとき、疾病を発生したのは4人の集団の中でBとCの2人なので、疾病発生割合 (incidence proportion) は、 $2/4=0.5$  と計算できる。一方、疾病発生率 (incidence rate) は、観察期間中の疾病発生数をリスク集団の合計観察時間で割ることで得られる。この場合、リスク集団の合計観察時間は、 $1+0.25+0.75+1=3$  人年 (person-year) なの

で、「3人年あたり2」、あるいは「100人年あたり67」と計算できる。疾病発生率は、対象者がいつ疾病を発生したかわからないと求めることができず、割合と異なり時間の逆数の単位をもつ。なお、生存時間解析では、疾病発生率のことをハザードとよぶことが多い。このように疾病発生割合とは、研究期間中にリスク集団で発生した疾病の数を研究開始時のリスク集団の対象者数で割ったものである。一方、疾病発生率は、リスク集団で観察された人-時間1単位あたりの疾病発生数である。

一般に、比、割合、率という用語は混同して用いられている。例えば、生存時間解析<sup>1)</sup>で Kaplan-Meier 法により求めた生存率などの用語は、定義から明らかに生存割合である。生存率のように誤用であっても明らかに割合であるとわかる場合は問題ないが、疾病発生率、死亡率は割合の意味と率の意味が混在している。例えば、よく用いられる死亡率という用語は、リスク集団のなかで死亡した人数を問題としている場合(死亡割合)、死亡するまでの時間を問題としている場合(死亡率)の両方の意味が混在している。割合であるか率であるかによって、エンドポイントの定義、サンプルサイズ設計などの研究デザイン、および解析方法がことなるので、臨床研究を行う上では割合と率の明確な区別をすることが重要である。

## 2 治療効果の指標

仮想的な例として、肺がんを発症し

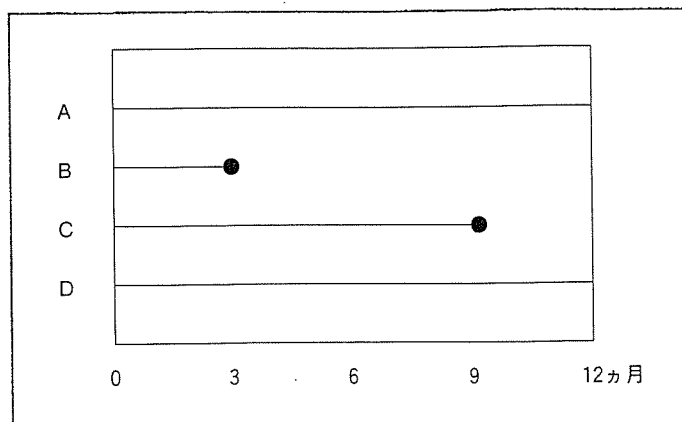


図1 4人のリスク集団の追跡例

●は疾病の発生を表す

表1 男性喫煙者 N 人における禁煙パンフレットの効果

禁煙 パンフレット	肺がん 発生数	対象者数	観察年	肺がん 発生割合	肺がん 発生率
送る	A	N	$T_1$	$A/N$	$A/T_1$
送らない	B	N	$T_0$	$B/N$	$B/T_0$

ていない男性喫煙者 N 人に対する禁煙パンフレットによる禁煙効果を調べること考えてみる(表1)。この N 人の男性に対して禁煙パンフレットを郵送した場合、その後の10年間に肺がんが A 人発生した(この場合の観察人-時間を  $T_1$ )とすると、肺がん発生割合、肺がん発生率は、それぞれ  $A/N$ 、 $A/T_1$  と求められる。一方、この同じ N 人に禁煙パンフレットを郵送しなかった場合、その後の10年間に肺がんが B 人発生した(この場合の観察人-時間を  $T_0$ )とすると、この場合の肺がん発生割合、肺がん発生率は、それぞれ  $B/N$ 、 $B/T_0$  と求まる。

表1の2つの状況は、「ひとつの集

団に対することなる治療状況」であり、その比較には妥当性(比較可能性)がある<sup>2)</sup>。この N 人に対する禁煙パンフレットの効果は、これら2つの結果の対比で表現され、治療によって疾病がどの程度改善されるかを定量的に表現した治療効果の指標にまとめることができる。

効果の指標には、大きく絶対的指標と相対的指標の2つがある。前者は、疾病発生の指標の差を取ったもので、指標として割合を用いた場合は、一般にリスク差(risk difference)、率を用いた場合は、疾病発生率の差(rate difference)とよばれる。表1では、以下の式から求めることができ、それぞれ、

この N 人に禁煙パンフレットが郵送されれば、「肺がんの発生が何パーセント減少(増加)するか」、「肺がんの発生が〇〇人あたり何人減少(増加)するか」を表す指標である。

$$\text{リスク差} : \frac{A}{N} - \frac{B}{N}$$

$$\text{疾病発生率の差} : \frac{A}{T_1} - \frac{B}{T_0}$$

一方、後者の相対的指標は、疾病発生の指標の比を取ったもので、指標として割合を用いた場合は、一般にリスク比(risk ratio)、率を用いた場合は、疾病発生率の比(rate ratio)とよばれる。表1では、以下の式から求めることができ、それぞれ、この N 人に禁煙パンフレットが郵送されれば、「肺がん発生割合が何分の一(何倍)になるか」、「肺がんの発生率が何分の一(何倍)になるか」を表す指標である。

$$\text{リスク比} : \frac{A}{N} \div \frac{B}{N}$$

$$\text{疾病発生率の比} : \frac{A}{T_1} \div \frac{B}{T_0}$$

多くの論文や学会発表などで、相対リスク(relative risk)という用語がよく用いられるが、これはリスク比、疾病発生率の比、オッズ比の総称である。上述のように、リスク比と疾病発生率の比は明らかに異なるので、相対リスクという用語は誤解を招く可能性があり、区別して使うことを薦める。

### 治療効果を検定する

禁煙パンフレットと肺がんの例で、リスク差を求めることに関心があるとす。前節で述べたように、禁煙パン

表2 禁煙パンフレットと肺がん

	肺がんの発生		
	あり	なし	合計
禁煙パンフレットあり	110 (0.22)	390	500
禁煙パンフレットなし	140 (0.28)	360	500
			1000

( )内は各グループの肺がん発生割合  
リスク差=0.22-0.28=-0.06

フレットの効果(治療効果)を調べるためには、N 人全員にパンフレットを送った場合の 10 年間の肺がん発生数と、「同じ」N 人にパンフレットを送らなかった場合の 10 年間の肺がん発生数を「同時に」比較しなければならない。しかしながら、これら 2 つの状況は、一方が観察されれば他方を観察することは不可能である。そこで、現実的なコントロールグループとして、ランダム化コントロール<sup>2)</sup>を採用したとする。すなわち、N 人の対象者をランダムに 2 つのグループに分け、半分には禁煙パンフレットを送り、残りの半分には送らなかったとする(表 2)。

禁煙パンフレットを郵送されたグループ、郵送されなかったグループそれぞれにおける肺がん発生割合(リスク、確率)は、110/500=0.22 (22%)、140/500=0.28 (28%)であり、禁煙パンフレットを郵送されたグループの方が肺がん発生割合が小さい。

この例で、われわれが調べたい仮説は、「研究に参加したすべての男性喫煙者に禁煙パンフレットを送らなかった場合よりも送った場合のほうが肺が

んの発生が減少する」である。しかしながら、この仮説は「研究に参加した対象者全体」についての仮説であり、ひとつの集団に対する 2 つの異なる状況の比較を問題としているので、直接検証することはできない。そこで、ランダム化の手続きにより選ばれたサンプルである禁煙パンフレットを「郵送されたグループ」と「郵送されなかったグループ」のそれぞれの肺がん発生割合を調べることによって、「対象者全体に対する仮説」が正しい(真)かどうかを検証することになる。

もし「対象者全体に対する仮説」が正しければ、サンプルである禁煙パンフレットを「郵送されたグループ」と「郵送されなかったグループ」の間でも前者の肺がん発生割合が小さくなることが予想される。実際に研究を実施した結果、パンフレット郵送グループのほうが肺がんの発生割合が小さかったとする(表 2)。予想通りの結果が得られたわけであるから、このことから「対象者全体についての仮説が正しい」と結論づけてよいであろうか(ここでは、統計的な誤差は無視する)。

仮説を簡単な形に表して問題を整理



してみる。

仮説 A：対象者全体について禁煙パンフレットが有効である

仮説 B：サンプルについて禁煙パンフレットが有効である

「仮説 A が正しいければ、仮説 B は正しい」という命題(命題 1)は、全体から部分を演繹的に予測したのであるから論理的に正しい命題である。一方、われわれには「サンプルを通して全体を推測する」という制約が課せられるので、データから推論できることは、「仮説 B が正しかったので、仮説 A は正しい」という命題 1 の逆命題である。一部の人に対して仮説が真だからといって全員に対して仮説が真とは限らないので、この逆命題は必ずしも真ではない。したがって、禁煙パンフレットを郵送されなかったグループに比べて郵送されたグループの肺がん発生割合が低かったとしても、われわれが研究で調べたい「対象者全体に対する仮説」を検証したことにはならない。

命題 1 が正しいとき、その命題から論理的に正しいといえるのはその対偶(仮説 B は誤っているので、仮説 A は誤っている)だけなので、もし「サンプルについて禁煙パンフレットが有効である」という仮説 B を否定できれば、「対象者全体について禁煙パンフレットが有効である」という仮説 A も誤っていることを証明できる。すなわち、仮説から演繹的に導かれる「予測」が否定されれば、その仮説が誤っていることを示すことはできる。このことを統計的仮説検定では利用している。ただし、われわれが検証したいこ

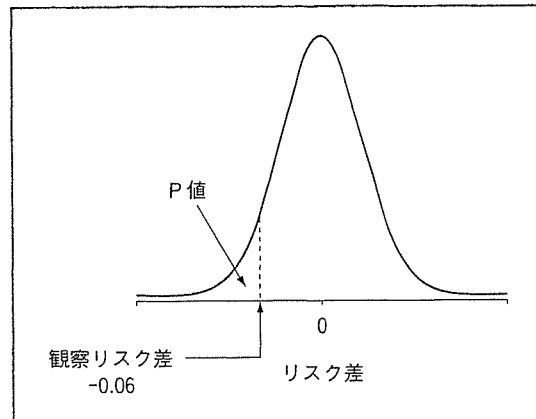


図 2 帰無仮説が正しいときのリスク差の分布

とは、仮説が誤っていることではなく仮説が正しいことであるので、「対象者全体について禁煙パンフレットが有効でなければ、サンプルについてもパンフレットは有効ではない」という命題(命題 2)を考えることにする。この命題 2 も否定的な意味ではあるが、全体から部分を演繹的に予測したものであるから論理的に正しい命題である。統計的仮説検定の手順をまとめると以下ようになる。

- 1) 「対象者全体について禁煙パンフレットは有効でない」という仮説を考える。この仮説は否定することが目的であるため、帰無仮説(null hypothesis)とよばれる。
- 2) 帰無仮説が正しければ「サンプルについても禁煙パンフレットは有効でない」と予想できる。
- 3) 実際に研究を行って「サンプル間で肺がん発生割合は等しい」という仮説を否定する証拠が見つかった場合、
- 4) 「対象者全体について禁煙パンフレットは有効でない」という帰無仮説

を否定できる。

表 2 の例を用いて具体的に上記の手順を考えてみる。

- 1) 「禁煙パンフレットと肺がんの発生は無関係」という帰無仮説を考える。
- 2) この無関係仮説が正しければ、禁煙パンフレットを「郵送されたグループ」と「郵送されなかったグループ」で肺がん発生割合は等しいはずである。
- 3) しかしながら、データには偶然のばらつきがあるため、発生割合がグループ間で完全に同じ値になるとは限らず、仮想的に研究を何回も繰り返せば、観察されるグループ間の肺がん発生割合の差(リスク差)はゼロを中心に左右対称にばらつくはずである(図 2)。観察されたりリスク差が、ゼロから十分離れていれば、仮説が間違っていると判断する。ただし、「ゼロから十分離れている」ということを、「リスク差が観察された値と同じか、それよりももっと仮説から離れた値をとる確率(P 値とよばれる：図 2 の点線より左側部分の面積)」の大きさと判断

する。

4)このP値が、事前に定めた値(有意水準とか $\alpha$ レベルとよばれる)よりも小さければ、「帰無仮説を棄却」し、リスク差はゼロではないと判断し、「禁煙パンフレットを郵送すれば肺がんの発生が減る」と考える。一方、P値が有意水準よりも大きければ、帰無仮説は棄却できず、「禁煙パンフレットと肺がんは無関係」かどうかは判断保留と考える。

検定結果が有意な場合には「差がある」といえるのに、有意でなかった場合には「差がない(帰無仮説が正しい)」といえないことに注意していただきたい。なぜなら、検定結果が有意でなかった場合は、命題2の逆命題(サンプルについて禁煙パンフレットは有効でなかったので、対象者全体についてパンフレットは有効でない)であり、これは必ずしも真ではないからである。

表2のデータにおいて、割合の差の検定(カイ二乗検定)の結果の両側P値は0.028であり、両側5%水準で有意である。ただし、検定結果が有意な場合、すべての人に対して効果があると解釈しがちであるが、それは誤った解釈である。「対象者全体について禁煙パンフレットと肺がんの発生は無関係」という帰無仮説の否定は、「すべての対象者について禁煙パンフレットが郵送されれば肺がんの発生が減る」ではなく、正しい解釈は「一部の男性喫煙者は、禁煙パンフレットが郵送されれば肺がんの発生が減る」である。

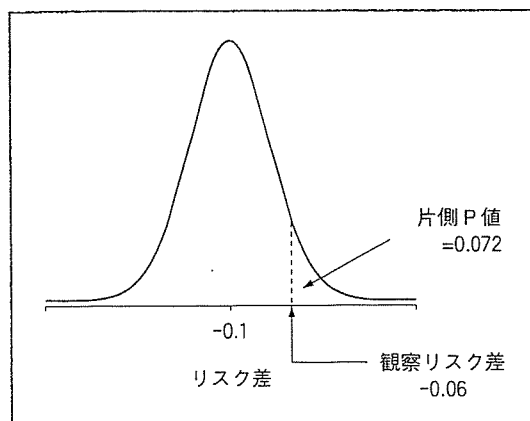


図3 真のリスク差が-0.1であった場合の仮説検定

### 治療効果の推定と95%信頼区間

前節で述べた統計的仮説検定は、禁煙パンフレットの効果(治療効果)の有無に関する二者択一の情報のみを与え、最も知りたい治療効果の大きさ「男性喫煙者に禁煙パンフレットを郵送することで、郵送しない場合に比べて肺がんの発生を何人予防できるか」については答えてくれない。この問いに答えるためには、治療効果の大きさを推定し、その信頼区間を求める必要がある。

表2のデータではランダム化コントロールを採用しているので、2グループ間には比較可能性がある(平均的には交絡はない)<sup>2)</sup>。したがって、禁煙パンフレットを「郵送されたグループ」の肺がん発生割合から「郵送されなかったグループ」の肺がん発生割合を引いた単純な(関連を表す)リスク差は、治療効果の指標に対するバイアスのない推定値となる。すなわち、単純なリスク差の推定値が-0.06であ

ることから、「この集団に対して禁煙パンフレットを郵送することによって、肺がんの発生が6%減少する、あるいは肺がんの発生が1000人あたり60人減る」と解釈できる。

このリスク差の値は、個人のリスク差の平均でもあり、また平均リスクの差でもあるので<sup>3)</sup>、治療効果の大きさの自然な推定値である。しかし、この推定値はサンプルに基づいて得られたものなので、必ずしも治療効果の真値ではない。このリスク差の推定値はある大きさ(標準誤差であらわす)をもってばらつくはずである。したがって、データから治療効果の推定値を求めるだけでは不十分で、「データと矛盾しない真のリスク差の範囲」(信頼区間とよばれる)を求める必要がある。

前節の仮説検定では、「禁煙パンフレットと肺がんの発生は無関係」、すなわち「真のリスク差=0」という帰無仮説を考えたが、真のリスク差がゼロ以外の場合の仮説検定を考えることもできる。例えば、「真のリスク差=

-0.1」という帰無仮説の検定を考えてみる。この検定の片側 P 値は 0.072 ( $>0.025$ )となり、両側 5%水準で有意ではない。つまり、「真のリスク差=-0.1」を「否定できない」ことになる(図3)。

いくつかの真のリスク差の値(-1から1まで)に対して同様のことを行うと、片側 P 値は 0.5 (真のリスク差=-0.06とした場合)から 0 (真のリスク差=1または-1とした場合)までの値をとる。このとき、「両側 5%水準で有意とならない真のリスク差の値」がデータと矛盾しない真のリスク差と考えることができ、その範囲がリスク差の 95%信頼区間である。したがって、真のリスク差の値を動かして検定し続けたときに、ちょうど片側 P 値が 0.025 となる場合の真のリスク差の値が信頼区間の下限/上限になる。表2の例で、そのような範囲を求めると-0.114 から-0.007 になる。

表2の例で、「リスク差=0」の検定結果は両側 5%水準で有意(P 値=0.028)であり、一方、リスク差の 95%信頼区間は-0.114 から-0.007 とその間にゼロを含んでいない。リスク差の 95%信頼区間を求めてその間にゼロが含まれていなければ、「治療効果なし(真のリスク差=0)」という帰無仮説は有意水準 5%で棄却されることになる。つまり、有意水準 5% (両側)の検定を行って有意であるということと、リスク差の 95%信頼区

間がゼロを含まない(リスク比の場合には1を含まない)こととは同じことである。

95%信頼区間を「その間に真の値を95%の確率で含んでいる」と解釈しがちであるが、これは誤った解釈である。上で述べたように仮説検定と信頼区間は表裏一体である。有意水準 5%の仮説検定では、「差がない」という帰無仮説が本当は正しくても100回中5回は間違えて「差がある」と判断してしまうことがある。1回の研究から得られた信頼区間が真の値を含む確率は1(含んでいる)か0(含んでいない)かのどちらかである。95%信頼区間とは、同じ規模の研究を繰り返して行ってそれぞれの場合について95%信頼区間を計算したとき、平均して95%は真の値を含むような区間のことである。

## おわりに

統計学的に「有意であった( $P<0.05$ )」、「有意でなかった(N.S.)」などのように検定結果(P 値)を用いて結果の議論をしている医学論文や学会発表をよくみかける。しかしながら、本来の仮説検定の意義は“decision making”であり、知識の積み重ねが要求される科学研究には仮説検定は不向きである。仮説検定の考え方は重要であるが、単に治療効果が「ある」か「ない」かよりも治療効果の大きさがどれくらいかを求めることのほうが、はるかに情報

量が多く、また医学研究の進歩にとって重要である。このためには、適切な治療効果の指標を定め、その大きさを推定し、その信頼区間を併記する必要がある。

対象者数が非常に多い場合は、わずかな治療効果(差)であっても「統計的に有意」となりうる。このわずかな差が仮に「統計学的に有意」であったとしても、個々の患者さんの治療においては臨床的に必ずしも重要であるとは限らない。一方、対象者数が少ないために、非常に大きな差であっても「統計学的に有意」とならない場合がある。もし、それが「臨床的に重要な差」であれば、単に「統計的な有意差がない」として無視すべきものではなく、さらなる検討の対象とされるべきである。「統計学的な有意差」と「臨床的な有意差」を明確に区別する必要があり、「臨床的に意味のある差」を検出しようとするような研究デザインを組むべきである。

## 文献

- 1) 松山 裕, 大庭幸治: 統計解析の実際. Surgery Frontier 11 : 412-419, 2004
- 2) 松山 裕: コントロール(対照群)の考え方. Surgery Frontier 11 : 76-80, 2004
- 3) Greenland S : Interpretation and choice of effect measures in epidemiologic analysis. American Journal of Epidemiology 125 : 761-768, 1987

## 4) サンプルサイズ的设计

東京大学大学院医学系研究科生物統計学

松山 裕

Yutaka Matsuyama  
(助教授)

### Summary

「研究者が考えている仮説が本当に正しい場合には肯定的な結論を、本当は誤っている場合には否定的な結論をきちんと導けること」がどのような研究を行う場合でも重要である。研究計画に不備がある研究を行って最終的に肯定的な結論も否定的な結論も得られなかったとしたら、研究参加者の協力を無駄にしたことになり、倫理的にも問題である。灰色の結論しか出せない原因のひとつに、サンプルサイズの問題がある。本稿では、サンプルサイズ設計の基本的な考え方とサンプルサイズ設計時の見積もりを間違えた場合の対策について解説する。

### Key Words

検出力,  $\alpha$ エラー,  $\beta$ エラー, 臨床的に意味のある差, 感度分析

### サンプルサイズ設計の必要性

サンプルサイズを設計するには、まず研究における主要なエンドポイント(評価項目)と統計解析の方法がはっきりと決まっていなければならない。この主要エンドポイントと用いる解析手法に基づいて、もし研究者が想定している治療効果の見積もり値が正しかったとしたら、その治療効果を正しく検出できる確率を設定し、研究に必要なサンプルサイズを計算する。「治療効果の見積もりが正しかった場合、その治療効果を正しく検出できる確率」を検出力(power)とよぶ。検出力が50%の研究では、治療効果の見積もりが本当に正しくても治療効果を検出できるチャンスは五分五分しかない。このような一か八かの研究に賭けることは危険が大きいき、検出力が20%、30%の研究では明確な結論が出ないことはほぼ確実である。

仮想的な例として、新しい禁煙プログラムを開発したとして、その効果を

ランダム化研究で評価することを考える(コントロール群は無介入とする)。この研究で何人の喫煙者に参加してもらえばよいだろうか。各群10人ずつの合計20人で実施した場合(ケース1)と各群500人ずつの合計1000人で実施した場合(ケース2)の結果を示す(表1)。この研究でのエンドポイントは、研究期間中に禁煙に成功したかどうかであり、(治療)効果は2グループ間の禁煙成功割合の差で表現される。「禁煙プログラムに効果がない」という帰無仮説の検定には、カイ二乗検定を用いるとする。ケース1とケース2における禁煙成功割合の差(禁煙プログラムの効果)の推定値とその95%信頼区間を示す(図1)。

ケース1では大きな差、ケース2では統計的に有意な差が観察されているので、どちらの研究も成功したかのようにみえるが、実はこの2つの研究はどちらも適切な研究とは言い難い。なぜなら、ケース1では禁煙プログラムに大きな効果(30%)がある可能

性は高いことはわかるが、効果がないことも否定できない(統計的には有意でない)。一方、ケース2では効果がないことは否定できる(統計的に有意である)、その効果の大きさ(7%)はわずかである。このような事態に陥らないためには、本当に禁煙プログラムに大きな効果がある場合には、確実に統計的に有意となって、逆に禁煙プログラムにたいした効果がない場合には、統計的に有意とはならないようなサンプルサイズが要求される。

### サンプルサイズ設計の基本的な考え方

通常サンプルサイズ設計は、「治療効果がない」という帰無仮説に対する統計的仮説検定に基づいて行われる。統計的にサンプルサイズを計算する際に必要なのは表2の4つであり<sup>1)</sup>、これらの値を研究開始前に決めなければならない。

「治療効果がない」という帰無仮説が正しい場合、群間差(2グループの禁煙成功割合の差)はゼロを中心に分布するはずである。この分布を用いて実際に研究を行って得られた群間差がゼロからどの程度ずれているかをp値で表現する。「治療効果がない」という仮説が正しい場合、p値は「群間差が観察された値以上にゼロから離れた値となる確率」であって、たとえp値が大きくても小さくてもこのままでは何も意味しない。非常に小さいp値、例えばp=0.0001が得られたとしても、「治療効果がない」という帰無仮説が正しいければ、10000回に1回し

表1 禁煙プログラムの効果

禁煙プログラム	ケース1			ケース2		
	成功	失敗	合計	成功	失敗	合計
実施	5	5	10	250	250	500
未実施	2	8	10	215	285	500
合計			20			1000

ケース1：禁煙成功割合の差=0.5-0.2=0.3(30%)、p=0.16(カイニ乗検定)  
 ケース2：禁煙成功割合の差=0.5-0.43=0.07(7%)、p=0.03(カイニ乗検定)

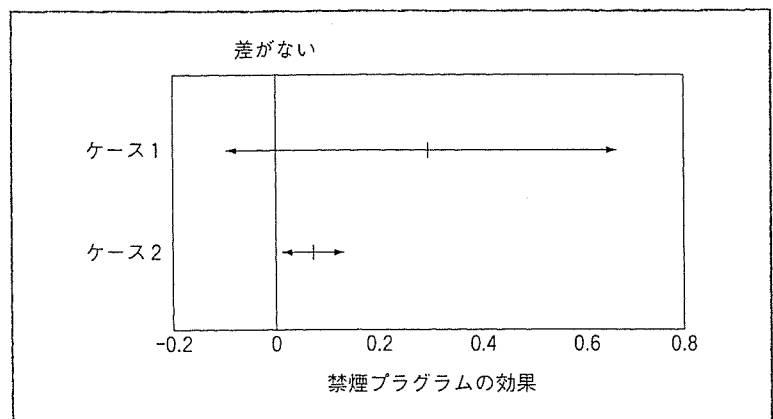


図1 禁煙成功割合の差の推定値と95%信頼区間

表2 サンプルサイズを計算するために決定しなければいけないもの

有意水準	本当は「治療効果がない」という仮説が正しいのに、誤って「治療効果がある」と判断してしまう可能性が、どの程度まで許容できるか
検出力	本当は「治療効果がある」という仮説が正しい場合に、正しく「治療効果がある」と判断できる可能性をどの程度まで確保するか
コントロール群での臨床イベント発生割合	禁煙プログラムの例では、無介入群における禁煙成功割合
治療効果の大きさ	試験群の結果がコントロール群の結果よりどの程度上回れば臨床的に意義があると考えられるか、その最小の値

か起こらないような非常にまれなことがたまたま起こってしまっただけである。しかし、われわれは治療効果の真実の大きさを知らないので、小さなp値が得られた場合、「本当は治療効果はなく、たまたま小さなp値が得られただけ」と考えるよりも、「禁煙プログラムには効果があるので、群間差がプラスの値になり、小さなp値が得られたのではないかと考えるほうが合理的である。

p値がどのくらい小さければ「治療効果がある」と判断するかの目安として、両側5%がよく用いられる。両側というのは、群間差がゼロから離れる方向としてプラス(介入群のほうが結果がよい)とマイナス(コントロール群のほうが結果がよい)の両方向を考慮することを意味する。両側5%は、「治療効果がある」と判断するための約束事であり、「もし本当に治療効果がない場合でも誤って治療効果ありと判断してしまう場合がある。そのような間違いが100回のうち5回起こる程度は我慢しよう」というのが5%の意味である。この値は、有意水準(あるいは、 $\alpha$ エラー、第1種の過誤)とよばれる。

有意水準とは逆に、「本当は治療効果がある場合に、誤って治療効果なしと判断してしまうことを許す程度」のことを $\beta$ エラー(あるいは、第2種の過誤)とよぶ。 $1-\beta$ が、「本当は治療効果がある場合に、正しく治療効果ありと判断できる確率」、すなわち検出力であり、通常80%や90%に設定される。

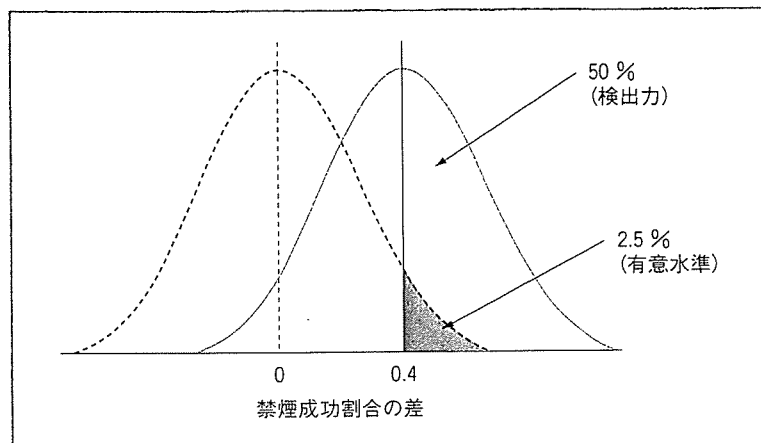


図2 ケース1における禁煙成功割合の差の分布

点線:「治療効果がない」という帰無仮説が正しい場合  
実線:「群間差=0.4」という仮説が正しい場合

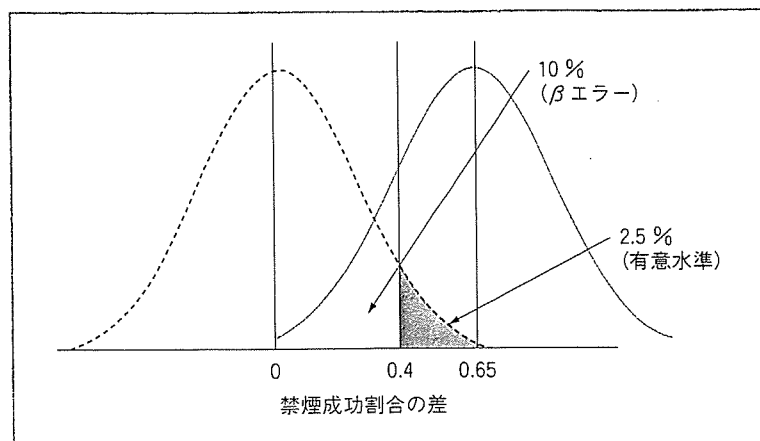


図3 ケース1における禁煙成功割合の差の分布

点線:「治療効果がない」という帰無仮説が正しい場合  
実線:「群間差=0.65」という仮説が正しい場合

コントロール群での臨床イベント発生割合(禁煙プログラムの例では、無介入群における禁煙成功割合)に関しては、過去の研究結果や文献例などから見積もることができる。最後の治療効果の大きさに関しては、試験群の結

果がコントロール群の結果よりどの程度上回れば「臨床的に意義」と考えられるかであり、統計的な問題ではなく研究を実施する目的に応じて決まる純粋に臨床的な問題である。

前述の禁煙プログラムのケース1

について考えてみる。10人ずつの合計20人で研究を行った場合、「治療効果がない」という帰無仮説が正しいときに2グループの禁煙成功割合の差がどのような分布にしたがうかは理論的に計算可能であり(図2の点線), 観察された群間差が0.4のときに片側p値がちょうど0.025となる。つまり, 10人ずつくらいの研究だと, 観察された群間差が40%以上であれば両側5%水準で「統計的に有意」となる。このことは, もし仮に「真の群間差が40%」であれば, 得られる群間差の分布は40%を中心に分布するはずなので(図2の実線), 仮説検定を行っても2回に1回しか「統計的に有意」とならないこと(検出力が50%)を意味する。

このケース1の例で, 真の群間差が65%の場合を考えてみる。この場合, 得られる群間差の分布は65%を中心に分布するはずなので(図3の実線), 10名ずつくらいの研究だと, 10%は統計的に有意にならない( $\beta$ エラー), すなわち「統計的に有意」といえるのは90%(検出力)である。このように, サンプルサイズ, 群間差, 検出力, 有意水準には密接な関係があり, サンプルサイズ, 群間差, 有意水準が決まれば, 検出力が計算可能である。したがって, 検出したい群間差, 有意水準, 検出力を決めれば, 必要なサンプルサイズを求めることができる。

禁煙プログラムの例で, 有意水準を両側5%, 検出力を90%, コントロール群での禁煙成功割合を10%, 禁煙プログラムの効果の大きさ(臨床

表3 介入群で期待される結果と検出力を変化させたときの必要サンプル

検出力	介入群の禁煙成功割合		
	50%	40%	30%
90%	52	84	164
80%	40	64	124
70%	32	52	98
60%	26	42	78
50%	22	34	62

サンプルサイズは2群の合計数

的に意味のある差の最小値)を10%と設定してみる。この条件で必要なサンプルサイズを計算すると, 各群26人, 全体で532人必要となる。

このように計算されたサンプルサイズが現実には収集可能な最大数よりも小さければ, このサンプルサイズで研究を実施することになる。しかし, 通常は, 「試験群の結果がコントロール群の結果よりどの程度上回れば臨床的に意義があると考えられるか, その最小の値」を用いてサンプルサイズを計算すると, とても参加が見込めない数になってしまうことが多い。収集できそうもない数になった場合には, 「試験群において期待できる結果」を過去の研究結果から見積もってサンプルサイズを計算することが多い。例えば, 「介入群での禁煙成功割合が50%くらいは期待できる」と考えれば, 有意水準両側5%, 検出力90%, コントロール群での禁煙成功割合10%の場合, 期待できる治療効果の大きさは40%であり, 必要サンプルサイズは各群26人, 全体で52人となる。

### 見積もり間違いへの対策

多くの研究において, 前述の「試験群において期待できる結果」を用いてサンプルサイズ設計を行っているのが現状である。しかし, そもそも「新しい治療を実施した場合に, 臨床的な結果がどの程度になるか」を知るために研究を実施するのであるから, 研究開始前にこの値を正確に見積もることは不可能である。通常は, 限られたほんの少数の対象者に対して研究者自身が実施した結果から推定するか, やはり少数の対象者に対して行われた文献報告などに頼らざるを得ない。このため, 試験群における結果の見積もりは間違っている可能性が高い。この見積もり間違いへの対策として, 研究計画の段階ではサンプルサイズの感度分析, 研究実施段階では中間解析が重要となる。

感度解析の考え方は, 試験群での結果に対する見積もりが不正確なのだから, その見積もりを変えることで必要なサンプルサイズがどの程度変化するか, あるいは検出力がどの程度変化す

るかを調べてサンプルサイズを決めるというものである<sup>2)</sup>。禁煙プログラムの例で、有意水準を両側5%、コントロール群での禁煙成功割合を10%とした場合に、介入群での期待禁煙成功割合を50%、40%、30%、検出力を90%、80%、70%、60%、50%に変化させたときの必要サンプルサイズ(合計)の変化を示す(表3)。

このような感度分析の結果を通して、「条件が悪くなった場合でも、ある程度の検出力を保証できるサンプルサイズ」を決定する必要がある。例えば、「治療効果に対する最悪のシナリオにおいて、検出力を最低60%(可能なら70%以上)に、治療効果が予想より少し悪いときには検出力を80%以上にする」という方針を考えてみる。この例では、最悪のケースである「介入群の禁煙成功割合が30%」のとき、検出力60%、70%では、必要サンプルサイズは、それぞれ78人と98人になる。一方、治療効果が予想より少し悪いケースである「介入群の禁煙成功割合が40%」のとき、検出力80%、90%では、必要サンプルサイズは、それぞれ64人と84人になる。したがって、この例における必要サンプルサイズを、84人から98人、あるいはきりのいいところで100人とすることが考えられる。このサンプルサイズでは、本当に禁煙プログラムに大

きな効果がある場合(介入群での禁煙成功割合が50%)には、確実に統計的に有意となるはずで、予想より多少治療効果が小さくなった場合でも(介入群での禁煙成功割合が30%)、ある程度の検出力が確保されている。さらに、禁煙プログラムに本当はたいした効果がない場合には(介入群での禁煙成功割合が20%や15%)、統計的に有意とはならないようなサンプルサイズとなっている。

### おわりに

サンプルサイズ設計はさまざまな見積り値に基づいており、不確実な要素が多い。しかし、だからといって適当な値をコンピュータプログラムに入力し、サンプルサイズを計算しさえすればよいということではない。サンプルサイズ設計を行う際に必ずしなければならないことは、潜在的な対象者が何名いるのかを把握することである。ランダム化臨床試験であれば、参加施設において過去1年間にプロトコル中の適格基準を満たす患者数がどれくらいかを調査しておく必要がある。この調査により、1年当たり最大何名の参加者を登録できるのか、あるいは予定している登録期間で最大何名収集可能か、サンプルサイズの実質的な上限をおえることができる。

感度分析では、いくつかの見積り

値を動かすことでサンプルサイズの変化を調べることができるが、サンプルサイズ設計でもうひとつ重要なのは対象者の脱落の考慮である。特に長期にわたる研究では、対象者に研究参加継続の意欲を高める工夫をすることが重要であるが、どうしても対象者の都合で研究が継続できないことがある。治療の中止や治療不遵守といったエンドポイントと関係した理由によるものは脱落ではなく、解析対象者に含めなければならないが、対象者の転居、不慮の事故といった治療やエンドポイントと無関係な理由による脱落は起こりうる。そのような脱落例を事前に見積もることができれば、その分を必要なサンプルサイズに上乘せすることができる。例えば、5%程度の脱落が予想されるのであれば、計算したサンプルサイズを1.05倍して最終的なサンプルサイズとすればよい。この脱落の見積もりなどもプロトコルに記載すべきである。

### 文献

- 1) Freidman LM, Furberg CD, DeMets DL: Fundamentals of Clinical Trials. 3rd edition. Springer-Verlag: 1998
- 2) 佐藤俊哉: サンプルサイズの設計: 駒澤 勉, 高木廣文, 佐藤俊哉(著): ヘルスサイエンスのための統計学. 45-56, 医歯薬出版株式会社, 東京, 1996



# 膵癌に対する新しい化学療法

上野秀樹\*・奥坂拓志\*\*・池田公史\*

## abstract

現在でも多くの膵癌は切除不能な状態で発見され、進行膵癌患者の予後は不良である。これらの進行膵癌に対しては、海外で行われた臨床試験の結果、ゲムシタビンの有用性が明らかにされ、進行膵癌に対する第一選択薬として現在広く認識されている。わが国でも2001年4月に膵癌に対するゲムシタビンの保険適用が承認され、膵癌に対して化学療法を行う機会が増加している。しかし、ゲムシタビン単剤の効果には限界があり、さらに優れた治療法を求めてゲムシタビンを中心とした多剤併用療法や分子標的薬剤を含む新しい抗癌剤の開発などが積極的に行われている。本稿ではそれらの新しい化学療法について、最近報告された臨床試験の成績を中心に解説する。

## I はじめに

わが国において膵癌で死亡する患者は年々増加しており、2001年には約19,000人に達している。これは日本における癌の死因の6.4%にあたり、肺、胃、大腸、肝に次いで第5位である。膵癌と診断された患者のうち切除可能な例は20%前後にすぎず、画像診断が進歩した現在においても多くの患者が進行膵癌の状態で見られ、その予後は不良である。

進行膵癌は脈管への浸潤などのために切除不能な局所進行膵癌と遠隔転移を伴った膵癌に分けられ、局所進行膵癌に対しては放射線化学療法または化学療法が行われ、遠隔転移を伴った膵癌に対しては化学療法が一般に行われている。ただし、全身状態が不良な例、黄疸を伴う例、著しい腹水を認める例などに対しては化学療法の効果は期待できないことから、緩和ケアを中心とした治療が選択される。

膵癌に対する化学療法は、進行膵癌に対する化学療法と術前または術後の補助化学療法に大別されるが、本稿では前者について最近の話題を中心に述べる。

## II 単剤療法

### 1) ゲムシタビン

膵癌に対しては従来フルオロウラシル (5-FU) を中心とした化学療法が行われてきたが、その治療成績は満足できるものではなかった。1997年、米国とカナダにおいて膵癌の初回化学療法例を対象としたゲムシタビンと5-FUの無作為化比較試験が行われ、生存期間中央値はゲムシタビン群が5.7カ月、5-FU群が4.4カ月と有意な ( $p=0.0025$ ) 差を認めた<sup>1)</sup>。また疼痛などの症状を緩和する効果もゲムシタビン群のほうが明らかに優れていた (24% vs 5%,  $p=0.0022$ )。これらの結果に基づきゲムシタビンが進行膵癌に対する第一選択薬と現在考えられており、

\* 国立がんセンター中央病院肝胆膵内科

\*\* 同 医長

表1 進行膵癌に対するゲムシタピン単剤療法の治療成績

報告者	報告年	対象	GEM投与量 (mg/m <sup>2</sup> )	症例数	奏効率 (%)	症状緩和 効果(%)	生存期間 中央値(月)	1年生存率 (%)
Casper	1994	初回化学療法例	800	44	11	NA	5.6	23
Carmicheal	1996	初回化学療法例	800	34	6	29	6.3	NA
Rothenberg	1996	5-FU無効例	1,000	63	11	27	3.9	4
Burris	1997	初回化学療法例	1,000	63	5	24	5.7	18
Crino	2001	初回化学療法例+他の化学療法無効例	1,000	33	12	NA	7.7	NA
Okada	2001	初回化学療法例	1,000	11	18	29	6.4	NA
Storniolo	1999	初回化学療法例+他の化学療法無効例	1,000	3,023	12	18	4.8	15

NA : not available, GEM : gemcitabine

報告者	報告年	抗癌剤	症例数	奏効率 (%)	生存期間 中央値(月)
Gebbia	1996	パクリタキセル	14	0	7.2
Rougier	2000	ドセタキセル	40	15	7.0
Okada	1999	ドセタキセル	21	0	3.9
Lenzi	2002	ドセタキセル	21	5	5.9
O'Relly	1996	topotecan	27	0	4.1
Scher	1996	topotecan	35	10	4.4
Sakata	1994	イリノテカン	35	11	NA
Wagener	1995	イリノテカン	34	9	5.2
Pazdur	1996	raltitrexed	42	5	NA
Miller	2000	pemetrexed	42	6	6.5
Cartwright	2002	カベシタピン	42	10	6.1
Okada	2002	ティーエスワン <sup>®</sup>	19	21	5.6
Cohen	2003	R115777	20	0	4.6
Bramhall	2001	marimastat	102	3	4.2
Moore	2003	BAY12-9566	138	1	3.7

表2  
進行膵癌に対する単剤療法の治療成績 (ゲムシタピン以外)

NA : not available

日本でも2001年4月に膵癌に対する保険適用が承認された。ゲムシタピンは代謝拮抗剤に分類される抗癌剤であり、細胞内で三リン酸化合物に代謝され、DNAの合成を阻害する。用法・用量は1,000mg/m<sup>2</sup>を30分かけて点滴静注し、これを週1回3週続けて投与した後、1週休薬する方法が一般的である。治療は病状の明らかな進行や重篤な副作用を認めない限り継続する。主な副作用は、骨髄抑制と悪心、食欲不振、倦怠などであるが、比較的軽度のことが多く、通常外来での治療が可能である。膵癌に対するゲムシタピン単剤の奏効率は5~15%程度であり、生存期間中央値は5~6カ月前後の報告が多い。ゲムシタピン投与により20~30%の患者で疼痛などの症状改善が報告されている。表1に進行膵癌に対するゲムシタピン単剤療法の治療成績を示す。

## 2) ゲムシタピン定速静注法

ゲムシタピンは、細胞内でdeoxycytidine kinaseによってリン酸化されて抗腫瘍効果を発揮する。ゲムシタピンがdeoxycytidine kinaseによって効率よ

くリン酸化されるよう投与方法の検討がなされ、10mg/m<sup>2</sup>/分の速度での投与が最適であることが示唆された。この投与方法を定速静注法 (fixed dose rate infusion) とよび、米国における第I相試験の結果、ゲムシタピンの定速静注法の推奨用量は1,500mg/m<sup>2</sup>/150分であることが報告された。Temperoらは、定速静注法 (1,500mg/m<sup>2</sup>/150分、週1回、3週投与1週休薬) と30分の点滴静注法 (2,200mg/m<sup>2</sup>/30分、2週に1回投与) との無作為化比較第II相試験を行い、生存期間中央値および1年生存率において定速静注法のほうが優れていたことを報告した<sup>2)</sup>。さらに現在、この定速静注法がゲムシタピンの標準的な投与方法になりうるかについて、米国で第III相試験が行われている。

## 3) 新規抗癌剤

最近報告された進行膵癌に対するゲムシタピン以外の抗癌剤による単剤療法の成績を表2に示す。ドセタキセルは開発当初比較的良好な奏効率が報告され注目を集めたが、その後の追試では良好な成績は

表3 進行膵癌に対する多剤併用療法の治療成績 (第II相試験)

報告者	報告年	治療法	症例数	奏効率 (%)	生存期間 中央値(月)	1年生存率 (%)
Berlin	2000	GEM + 5-FU (bolus injection)	36	14	4.4	8.6
Hidalgo	1999	GEM + 5-FU (continuous infusion)	26	19	10.3	40
Wein	2002	GEM + 5-FU (high dose infusion)	23	9	8.3	30
Colucci	1999	GEM + シスプラチン	52	31	NA	NA
Heinemann	2000	GEM + シスプラチン	41	11	8.2	27
Philip	2001	GEM + シスプラチン	42	26	7.1	19
Cascinu	1999	GEM + ドセタキセル	18	6	5.4	NA
Ryan	2002	GEM + ドセタキセル	34	18	8.9	29
Rocha Lima	2002	GEM + イリノテカン	45	20	5.7	27
Stathopoulos	2003	GEM + イリノテカン	60	25	7.0	23
Louvet	2002	GEM + oxaliplatin	64	31	9.2	36
Alberts	2003	GEM + oxaliplatin	47	11	6.2	NA
Neri	2002	GEM + エビルピシン	44	25	10.9	23
Hess	2003	GEM + カペシタビン	27	19	6.4	33
Scheithauer	2003	GEM + カペシタビン	41	17	9.5	NA
Kralidis	2003	GEM + raltitrexed	25	12	6.2	12
Kindler	2002	GEM + pemetrexed	42	15	6.5	29
Safran	2001	GEM + トラスツズマブ	21	22	NA	NA
Abbruzzese	2001	GEM + cetuximab	41	12	6.8	33
Kindler	2003	GEM + bevacizumab	21	38	NA	54
Hecht	2003	GEM + ONYX-015 (intratumoral injection)	21	10	7.5	29

NA : not available, GEM : gemcitabine

得られなかった。イリノテカンやtopotecanなどの topoisomerase I 阻害剤も膵癌に対して試みられてきたが、有効性は明らかでない。最近注目されている抗癌剤のひとつに、経口フッ化ピリミジン系抗癌剤であるカペシタビンやティーエスワン®が挙げられる。カペシタビンは5-FUのプロドラッグで、現在欧米を中心に臨床試験が行われており、膵癌に対する奏効率は10%であったが、24%に症状の改善が認められた<sup>3)</sup>。ティーエスワン®はわが国で開発されたテガフルとギメラシル、オテラシルカリウムの配合剤で、進行膵癌を対象とした前期第II相試験で21% (4/19) の奏効率が示され<sup>4)</sup>、後期第II相試験がわが国で進行中である。分子標的薬剤は現在最も注目されている抗癌剤のひとつであり、最近さまざまな作用機序をもった薬剤が膵癌に対して試みられている。marimastatやBAY12-9566は、細胞外matrixを分解する酵素であるmatrix metalloproteinaseを阻害し、癌の発育を抑制することが期待されて開発された薬剤である。また、膵癌にK-rasの突然変異が高率にみられることに着目し、rasタンパクの機能発現に必要なfarnesyl化を抑制するfarnesyl transferase阻害剤 (R115777) を用いた治療が試み

られている。さらに、上皮成長因子受容体 (epidermal growth factor receptor : EGFR) のモノクローナル抗体であるcetuximabや選択的EGFRチロシンキナーゼ阻害剤のOSI-774などの分子標的薬剤についても臨床試験が進んでいる。これらのうち、marimastatとBAY12-9566は、それぞれゲムシタビン単剤との間で第III相試験が行われたが、いずれの試験においてもゲムシタビン単剤を上回る成績は得られていない<sup>5), 6)</sup>。

### III 多剤併用療法

単剤で良好な抗腫瘍効果を示したゲムシタビンは、他の抗癌剤との併用によりさらに治療成績が向上することが期待されており、現在多くの臨床試験が世界中で進行している (表3, 4)。併用する薬剤は、単剤でも膵癌に対してある程度の効果が認められている5-FUやシスプラチンなど既存の抗癌剤や、分子標的治療薬などの新規抗癌剤が中心となっている。

現在までに最も多く報告されている併用薬は5-FUであり、bolus injection, continuous infusion,

報告者	報告年	治療法	症例数	奏効率 (%)	p値	生存期間中央値(月)	p値
Berlin	2002	GEM + 5-FU	160	7	NS	6.7	0.09
		GEM	162	6		5.4	
Colucci	2002	GEM + シスプラチン	53	26	0.02	7.5	0.43
		GEM	44	9		5	
Heinemann	2003	GEM + シスプラチン	98	10	NS	7.6	0.12
		GEM	100	8		6	
Louvet	2003	GEM + oxaliplatin	157	26	0.03	NA	NA
		GEM	156	16		NA	
Rocha Lima	2003	GEM + イリノテカン	173	16	<0.001	6.3	0.79
		GEM	169	4		6.6	
Van Cutsem	2002	GEM + R115777	341	6	NS	6.4	0.75
		GEM	347	8		6.1	
Bramhall	2002	GEM + marimasat	120	11	NS	5.5	0.95
		GEM	119	16		5.5	

表4  
進行膵癌に対する多剤併用療法とゲムシタビン単剤療法の比較試験（第Ⅲ相試験）  
NA：not available,  
GEM：gemcitabine

high dose infusionなどさまざまな方法で試みられてきた。第Ⅱ相試験では比較的良好な奏効率と生存期間が報告されたが、Berlinら<sup>7)</sup>が行ったゲムシタビンと5-FU bolus injectionとの併用療法vsゲムシタビン単剤療法の比較試験では、生存期間中央値が6.7か月vs 5.4か月と有意差を認めなかった(p=0.09)。

シスプラチンは、ゲムシタビンとの併用により相乗効果を示し、他の癌種において良好な成績が報告されていることから、膵癌に対しても併用効果が期待された。Colucciら<sup>8)</sup>は、weeklyのシスプラチン(25mg/m<sup>2</sup>)を、Heinemannら<sup>9)</sup>は2週に1回のシスプラチン(50mg/m<sup>2</sup>)をゲムシタビンと併用し、それぞれゲムシタビン単剤と比較試験を行った。両試験とも併用群のほうが無増悪生存期間は優れていたが、生存期間の有意な延長は認めなかった。

イリノテカンも膵癌に対し期待された抗癌剤であるが、Rocha Limaら<sup>10)</sup>の報告ではイリノテカン(100mg/m<sup>2</sup>)とゲムシタビン併用群とゲムシタビン単剤群の間で、生存期間の差は認めなかった。

Louvetら<sup>11)</sup>は、第三世代の白金化合物であるoxaliplatin(100mg/m<sup>2</sup>)と定速静注法のゲムシタビン(1,000mg/m<sup>2</sup>を100分間で投与)の併用とゲムシタビンの通常投与方法を比較し、奏効率と無増悪生存期間は併用群のほうが有意に優れていたことを報告している。現在生存期間に関する解析が進行中であり、結果が期待されている。その他、ドセタキセルやカペシタビン、raltitrexed, pemetrexedなどの新しい抗癌剤とゲムシタビンの併用療法に関する

臨床試験が現在進行中である。

分子標的薬剤では、matrix metalloproteinase阻害剤であるmarimastatや、farnesyl transferase阻害剤であるR115777とゲムシタビンの併用療法がゲムシタビン単剤療法と第Ⅲ相試験にて比較されたが、生存期間の差を認めなかった<sup>12), 13)</sup>。その他、EGFR阻害剤のcetuximab、HER2阻害剤のトラスツズマブ、VEGF阻害剤のbevacizumabなどの分子標的薬剤とゲムシタビンの併用療法が積極的に行われており、今後の結果が注目される。

## IV おわりに

ゲムシタビンの登場により、膵癌に対する化学療法は大きな変貌を遂げてきた。2003年には米国臨床腫瘍学会(ASCO)で膵癌の化学療法に関する大規模な比較試験について3演題報告され、分子標的薬剤を含む新規抗癌剤の臨床試験も活発に報告されている。膵癌患者の予後は今なお不良であり、近い将来に、より優れた治療法が開発されることを期待したい。

### 参考文献

- 1) Burris HA 3rd, Moore MJ, Andersen J, et al: Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. J Clin Oncol 15: 2403~2413. 1997