

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 44, 175-188.

Xie, F. and Paik, M. C. (1997). Multiple imputation methods for the missing covariates in generalized estimating equation. *Biometrics* 53, 1538-1546.

付 録

A 重み付き推定量の漸近分散

欠測メカニズムが MAR のもとで、式 (8) における回帰パラメータ β の重み付き推定を考える。 β の解は、以下の重み付き推定方程式を解くことで得られる。

$$U(\beta, \hat{\gamma}) \equiv \sum D_i^T(x_i; \beta) V_i^{-1}(\hat{\alpha}) w_i(\hat{\gamma}) [y_i - g(x_i; \beta)] = \sum U_i(\beta, \hat{\gamma}) = 0$$

ただし、 $g(x_i; \beta) = E(y_{it} | x_i)$, $D_i(x_i; \beta) = \partial g(x_i; \beta) / \partial \beta^T$, $V_i(\hat{\alpha})$ はその要素が (9) 式から計算される作業共分散行列、 $w_i(\hat{\gamma})$ は重み行列 (対象者が各時点で観察される累積確率の逆数) である。

(10) 式のような重みを推定するためのモデルが正しければ、上記の重み付き推定方程式から得られる β の解 $\hat{\beta}$ は一意に定まり、 $U(\beta_{true}, \hat{\gamma})$ と $(\hat{\beta} - \beta_{true})$ は漸近的に平均がゼロで、分散共分散行列がそれぞれ、 C , $(\hat{\Gamma}^{-1})C(\hat{\Gamma}^{-1})^T$ の正規分布に従うことが示されている (Robins et al., 1995)。これらの分散共分散行列の推定値は、 $\hat{\Gamma} = \partial U(\hat{\beta}, \hat{\gamma}) / \partial \hat{\beta}^T$, $\hat{C} = \hat{A} - \hat{B}\hat{\Omega}\hat{B}^T$ から求めることができる。ただし、 $\hat{A} = \sum U_i(\hat{\beta}, \hat{\gamma}) U_i(\hat{\beta}, \hat{\gamma})^T$, $\hat{B} = \partial U(\hat{\beta}, \hat{\gamma}) / \partial \hat{\gamma}^T$, $\hat{\Omega}$ は重みを推定するための式 (10) のモデルに対する尤度関数から計算される観察情報行列の推定値である。

上記の β に対する漸近分散は、 $(\hat{\Gamma}^{-1})\hat{A}(\hat{\Gamma}^{-1})^T - (\hat{\Gamma}^{-1})\hat{B}\hat{\Omega}\hat{B}^T(\hat{\Gamma}^{-1})^T$ と表現できる。この分散の第 1 項 $(\hat{\Gamma}^{-1})\hat{A}(\hat{\Gamma}^{-1})^T$ は、重みの推定誤差を考慮していない通常のプロバスト分散 (Liang and Zeger, 1986) である。したがって、重みを考慮した分散のほうが通常のプロバスト分散よりも少なくとも効率が良いことがわかる。

B パターン混合モデルにおける周辺平均 $\sum_d p^{(d)} \theta_i^{(d)}$ の漸近分散

ある時点における (脱落パターンに関して平均した) 周辺平均 $\hat{\theta}_t = \sum_d p^{(d)} \theta_i^{(d)}$ の分散は、 $\pi = (p^{(d)}, \dots, p^{(D)})^T$, $\delta = (\theta_i^{(1)}, \dots, \theta_i^{(D)})^T$ とすると、 $\hat{\theta}_t = \sum_d p^{(d)} \theta_i^{(d)} = \pi^T \delta$ なので、デルタ法により、

$$\text{Var}(\hat{\theta}_t) = \hat{\pi}^T \text{Var}(\hat{\delta}) \hat{\pi} + \hat{\delta}^T \text{Var}(\hat{\pi}) \hat{\delta}$$

となる。ただし、 $\text{Var}(\hat{\pi}) = \text{diag}(\hat{\pi}) - \hat{\pi}\hat{\pi}^T$ である。

一般に、任意の関数 $h(\hat{\delta}, \hat{\pi})$ に対する漸近分散は、 $\text{Var}[h(\hat{\delta}, \hat{\pi})] = \Omega \begin{pmatrix} \text{Var}(\hat{\delta}) & 0 \\ 0 & \text{Var}(\hat{\pi}) \end{pmatrix} \Omega^T$ で

ある。ただし、 $\Omega = \begin{pmatrix} \frac{\partial}{\partial \delta} h(\hat{\delta}, \hat{\pi}) & \frac{\partial}{\partial \pi} h(\hat{\delta}, \hat{\pi}) \\ \frac{\partial}{\partial \delta} h(\hat{\delta}, \hat{\pi}) & \frac{\partial}{\partial \pi} h(\hat{\delta}, \hat{\pi}) \end{pmatrix}$ である。

別刷

Surgery Frontier

メヂカルビュー社

2) 統計解析の実際

東京大学大学院医学系研究科生物統計学

松山 裕・大庭 幸治

Yutaka Matsuyama Koji Oba
(助教授)

Summary

医学データの統計解析を行う際に、統計専門家以外のユーザーにとって最も悩ましい問題は、多くの統計手法が存在し、統計ソフトを用いれば簡単にそれらが利用できる状態にある一方で、どの手法を選択すればよいかの指針がはっきりしないことであろう。統計手法の選択を決める要因は複数考えられ、①研究の型(データ収集方法)、②変数の数と興味のある関連、③目的および証明したい仮説、④想定できる前提、⑤変数の型・分布・データの質、⑥事前情報の質と量、などさまざまなものがある。本稿では、結果変数(エンドポイント)が生存時間データの場合における統計解析手法(率の算出、カプラン・マイヤ法、ログ・ランク検定、マンテル・ヘンツェル法、Cox 回帰)について紹介する。

Key Words

発生率、生存時間解析、Kaplan-Meier 法、log-rank 検定、交絡調整、Cox 回帰

はじめに

医学研究でよく利用される統計手法を表にまとめた(表 1)。結果変数(エンドポイント)の型の観点から分類したのが表 1 である¹⁾。本稿では、誌面の都合上、あるイベント発生までの時間を結果変数とする場合の統計解析(生存時間解析と総称される)について述べる。生存時間解析を含むその他の手法については、医学統計の教科書²⁾を参照いただきたい。

率の算出

特定の個人が研究開始時に研究対象としている疾病にかかっておらず、将来その疾病にかかる可能性があり、かつ研究者による積極的な追跡を受けている状態を「at risk」とよぶ。この「at risk」の状態にある対象者が新たに疾病にかかった場合、それを疾病発生(incidence)という。コホート研究(前向き研究: prospective study)では「at risk」状態にある対象者の集団

(リスク集団)から研究期間中に起きた疾病発生数を数え上げる。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分数」であり、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、「単位時間あたりの変化を表す比」であり、単位が存在し、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 ヶ月目、9 ヶ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。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

表1 医学研究でよく用いられる統計手法

解析の目的	結果データの型		
	二値データ	連続データ	生存時間データ
分布の記述	頻度集計 分割表	ヒストグラム 平均, 標準偏差 散布図, 相関	カプラン・マイヤ法
単純な群比較	カイ二乗検定 リスクの推定	t検定 Wilcoxon 検定	ログ・ランク検定 発生率の推定
層別解析	マンテル・ヘンツェル法 調整リスクの推定	分散分析	マンテル・ヘンツェル法
回帰モデル	ロジスティック回帰 調整リスクの推定	重回帰分析 分散分析	Cox 回帰 調整発生率比の推定

のとなる。研究開始後には新たに対象者を登録しないコホートを「登録に関して閉じたコホート」とよぶが、疾病発生割合はこの閉じたコホートでなければ定義することができない。

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。図1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-years) 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。このような疾病発生率の算出方法は、人-時間法 (person-time method) とよばれ、対象者がいつ疾病を発生したかが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。疾病発生率は閉じたコホートでも、あるいは研究開始後の対象者の登録、転出などを許す開いたコホートのいずれであっても定義可能であり、コホート研究で特に関心のある疾病発生の指標である。なお、

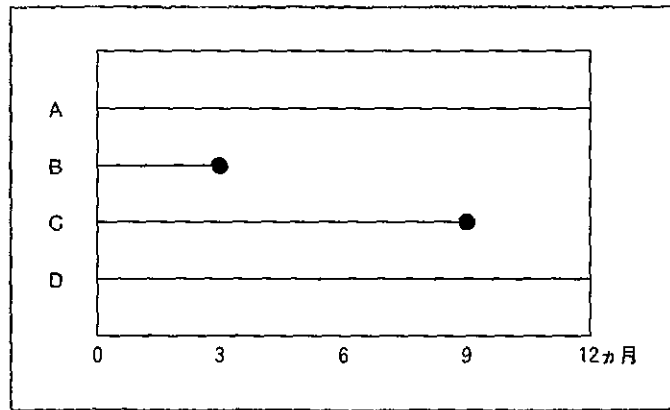


図1 4人のリスク集団の追跡例

●が疾病の発生を表す。

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。

乳がんに対する内分泌療法剤であるタモキシフェンの服用と二次がんの関係を調べるために行われたヒストリカル・コホート研究の際に得られた乳がん再発に関するデータの一例を示す (表2)。全国9施設の乳がん患者全例のうち1982年から1990年までの間

に治癒切除手術が実施された一側性の原発女性乳がん患者が対象である。乳がん術後にタモキシフェンを服用したグループの再発率は、1,000人年あたり10 (手術時のリンパ節転移なし)、49 (手術時のリンパ節転移あり) であり、タモキシフェン非服用グループでのそれらは、それぞれ14, 52である。

表2 乳がん術後のタモキシフェン服用と乳がん再発率(リンパ節転移の有無別)

リンパ節転移	タモキシフェン服用	再発数	対象者数	観察人年	再発割合	再発率
なし	あり	96	1334	9713.8	0.072	0.010
	なし	171	1592	12599.3	0.107	0.014
あり	あり	368	1215	7514.1	0.303	0.049
	なし	253	760	4862.2	0.333	0.052

time-to-event データの解析

発生率を疾病発生の指標として用いる場合、対象者が研究で関心のある疾病を発生したかどうかだけでなく、研究参加時点から疾病発生までの時間を問題としなければならない。心筋梗塞などの生活習慣病の発症、あるいは究極的には死亡をエンドポイントとした場合には、エンドポイント発生までの時間を長引かせる治療法は臨床的に意義がある。疾病を発症するまでの時間、死亡するまでの時間、再発するまでの時間といった何らかのイベントが起きるまでの時間(time-to-event)をエンドポイントとした場合のデータ解析は一般に生存時間解析とよばれる⁹⁾。

生存時間解析では、イベント発生までの時間をグループ間で比較することが目的となるが、対象者全員に対してイベント発生までの時間を観察できることはまれである。ほとんどの研究では、研究終了時点で関心のあるイベントをまだ発生していない対象者、追跡途中でイベントの発生状況を確認できなくなる対象者、関心のあるイベントの発

生以外の理由で死亡する対象者などが存在する。このような関心のあるイベントの発生までの時間を定義できない対象者は打ち切り(censoring)とよばれる。特に、追跡途中でイベントの発生状況を確認できなくなる対象者のことを追跡不能(loss to follow-up)、関心のあるイベントの発生以外の理由で死亡する対象者のことを競合リスク(competing risk)による打ち切りとよぶ。打ち切りを受けた対象者は、研究参加時点から打ち切りを受けるまで「at risk」の状態にあり、それらのイベント発生時点は「打ち切り時点以降である」という情報をもっているの、解析対象から除くことはできない。

通常の生存時間解析においては打ち切りデータが存在するために、イベント発生までの時間を直接比較することはできない。しかしながら、「単位時間あたりに何例イベントが起きるか」というイベント発生の強さの指標である発生率の計算・比較は可能である。打ち切りを受けた対象者に関しては、打ち切りを受けるまでの時間を観察時間と考え、人-時間法の分母を「打ち

切りを含めた合計観察時間」とすればよい。ただし、どのようなタイプの打ち切りでも許されるわけではない。通常の生存時間解析で想定している打ち切りは、関心のあるイベントとは無関係な理由による打ち切りである。例えば、研究終了時点でまだイベントを発生していない場合やイベントとは関係のない理由による転居(追跡不能)、参加拒否、死亡(例えば、偶然の交通事故)などである。逆に、打ち切りが(その後に潜在的に観察される)イベント発生に関して何らかの関係をもっている場合は情報のある打ち切り(informative censoring)とよばれる。例えば、心筋梗塞発生をエンドポイントとしている場合に、コレステロールが高くなってきた対象者ほど打ち切りを受けやすい場合や危険因子がよく似ている疾病(例えば、脳梗塞など)による死亡などである。情報のある打ち切りに対しては、単純に打ち切り例として解析を行うと結果にバイアスが生じることが知られている。最近は、そのようなバイアスを補正する解析方法が提案されつつあるが、汎用的で広く

表3 タモキシフェン服用による乳がん再発率の抑制効果(リンパ節転移の有無別)

リンパ節転移	再発率の比(95%信頼区間)	再発率の差(95%信頼区間)
なし	0.71 (0.57, 0.93)	-0.0036 (-0.007, -0.001)
あり	0.94 (0.80, 1.10)	-0.0031 (-0.011, -0.005)

受け入れられている解析方法は存在しないのが現状である。情報のある打ち切りを否定できない場合には、イベントの定義を複数行い、当該の事象を打ち切りとみなす場合とイベントとみなす場合の2通りの解析を行い、結果の整合性をみるのが現実的な対処法である。

表2のデータに対するタモキシフェン服用の再発抑制効果を示す(表3)。リンパ節転移なしのサブグループでは、タモキシフェン服用グループの方が再発率が低く(再発率を約3/4倍、あるいは10,000人年観察すると再発が36人減少)、その差は統計的に有意である(再発率比の信頼区間が

1を含まない、あるいは差の信頼区間が0を含まない)。リンパ節転移ありのサブグループでもタモキシフェンを服用した方が再発率は小さいが、その差は統計的に有意ではない。なお、発生率比、発生率差の95%信頼区間の計算方法は疫学の教科書⁷⁰⁾を参照していただきたい。

人-時間法による率の算出・比較は理解しやすく計算も簡単であるが、その方法が妥当であるためには、「発生率が観察期間を通して一定である」という条件が必要である。この条件が成立している場合、人-時間法による発生率の逆数は、「合計観察時間/発生数」なので、イベント発生までの平均

的な時間として解釈できる。したがって、「2グループ間で発生率が等しい」ことは「2グループ間の平均的なイベント発生時間が等しい」ことを意味し、打ち切りを含むイベント発生までの時間の比較を行っていることになる。

しかしながら、人-時間法に必要な条件「発生率は観察期間を通して変化しない」は研究開始から1年たっても5年たっても発生率は同じであることを要求しており、かなり厳しい仮定である。この仮定が満たされないと思われる場合には、観察期間を発生率が一定とみなせる期間に区切って比較を行う必要がある。一方、上の条件を緩めて「発生率は時間とともに変化してもかまわない」とすると、発生率は時々刻々と変化するので、もはや人-時間法のような安定した推定値を得ることはできなくなる。しかし、時間によって変化する発生率自体を推定することは困難であっても発生率が累積して観察される発生曲線(累積発生率)

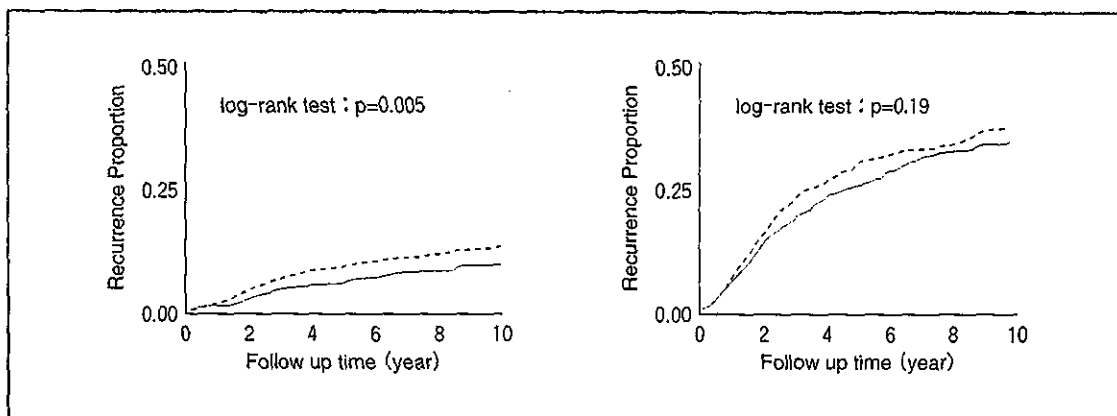


図2 乳がん再発曲線(実線:タモキシフェン服用グループ, 点線:非服用グループ)

左図:リンパ節転移なし, 右図:リンパ節転移あり

表4 乳がん術後のタモキシフェン服用と乳がん再発率(リンパ節転移の有無と閉経状況で層別)

リンパ節転移	閉経状況	タモキシフェン服用	再発数	観察人年	再発率	再発率比 (95%信頼区間)
なし	前	あり	48	4992.7	0.0096	0.78
		なし	80	6455.6	0.0124	(0.54, 1.11)
	後	あり	48	4721.1	0.0102	0.69
		なし	91	6143.7	0.0148	(0.48, 0.97)
あり	前	あり	205	4213.6	0.0487	1.03
		なし	130	2758.5	0.0471	(0.83, 1.29)
	後	あり	163	3300.5	0.0494	0.84
		なし	123	2103.7	0.0585	(0.67, 1.07)

を推定することは比較的容易である。イベント発生までの時間の分布やイベント発生率に特に強い仮定を必要としないイベント発生曲線を求める方法が Kaplan-Meier 法 (Kaplan-Meier method) である⁵⁰⁾。表2のデータに対する乳がん再発曲線を示す(図2)。リンパ節転移のあり・なしにかかわらず、タモキシフェン服用グループの方が観察期間を通して常に再発割合が低いことがわかる。

時間によって発生率が変化してもかまわないが、「2グループ間の発生率は等しい」という仮説を検定する手法がログランク検定(log-rank test)である⁵¹⁾。図2のデータに対しては、リンパ節転移なしのグループでは、ログランク検定のp値は0.005であり、リンパ節転移ありのグループでのそれは0.19である。2グループ間での発生率が等しいかどうかの検定手法としては、ログランク検定のほかにもウィルコクソン検定を拡張した一般化ウィルコクソン検定を用いることもできる。しか

し、一般化ウィルコクソン検定は研究開始直後に死亡率が異なると検定結果が有意になりやすい性質がある。このため、早期打ち切り例が多数存在すると、もし打ち切り例をもっと追跡すれば差がなくなるような場合であっても誤って差があるという結果になる可能性があり、注意が必要である。また、Kaplan-Meier 曲線を比較する際にどうしても研究期間の最後の方に目がいてしまいがちであるが、後半部分では「at risk」の状態にある対象者が少なくなっていることが多いので、生存曲線の推定も不安定である。数名のイベント発生で曲線の形状がかなり変化する場合があるので、「研究の後半で差が開いている」という解釈には注意が必要である。そのような場合、生存曲線の横軸に適当な時点ごとの「at risk」の人数を併記して、注意をうながしておくことが重要である。

交絡の調整・制御

表2のデータにおける「リンパ節

転移の有無」のような変数は、交絡因子(confounding factor)とよばれる⁵²⁾。交絡因子とは、当該のイベントに対するリスク因子で、かつ比較する群間でその人数が異なるような変数である。交絡因子は原因と結果の間の因果関係をゆがめるので、バイアスのない治療効果を求めるためにはその影響を調整する必要がある。交絡の調整は、デザインでも統計解析でも可能である。前者の方法には、対象者の限定、マッチング、ランダム化の3つの方法が存在する⁵³⁾。ここでは、後者の統計解析による方法について述べる。

統計解析による交絡調整方法には、層別解析によるものと回帰モデルによるものの2つが存在する⁵⁴⁾。層別解析(stratified analysis)とは、各層内では交絡の影響が無視できるように交絡因子でデータを十分に層別し、層ごとの効果の指標を要約する方法である。例えば、リンパ節転移なしで閉経前のグループでは、全員がリンパ節転移なしで閉経前なので、これらの要因による

交絡は生じない(表4)。この層別したデータに対して、「治療によって疾病発生率は変化しない(疾病発生率の比が1)」という帰無仮説の検定は、マンテル・ヘンツェル検定(Mantel-Haenszel test)で調べることができる⁷⁾⁸⁾。マンテル・ヘンツェル検定は、層ごとのカイ二乗検定を総合したような検定である。このデータの場合、カイ二乗値は4.03、p値は0.04であり、「リンパ節転移の有無と閉経状況の影響を調整すると乳がん術後にタモキシフェンを服用することにより再発率が変化する」ことがわかる。

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)のみを用いて結果の議論をしている医学論文をよくみかける。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方が、はるかに情報量が多い。このためには、適切な効果の指標(リスク差、リスク比、ハザード比など)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。

表4のデータに対して、マンテル・ヘンツェル発生率比を求めると、0.87(95%信頼区間:0.76, 0.99)となり、リンパ節転移の有無と閉経状況の影響を調整すると乳がん術後にタモキシフェンを服用することにより再発率が約5/6倍になることがわかる。

マンテル・ヘンツェル推定量は、比較的計算が簡便で、全体のサンプル数が多ければ各層内のサンプル数は少なくても妥当な結果を導くなど統計的な性質も優れているので、臨床・疫学研究で最もよく用いられる解析手法である⁷⁾⁸⁾。

なお、各層ごとの効果の指標だけを求めて、それらの結果を要約しない方法を層別解析とよんでいる論文をよくみかけるが、それは「サブグループ解析」であって、層別解析ではない。サブグループ解析とは、リンパ節転移なしで閉経前といった特定のサブグループに対する結果を調べる方法で、事前に計画されていないサブグループでの効果の大きさを事後的に推定するためには一般に非常に大きなサンプルサイズを必要とする。

層別解析による交絡の調整方法は、数学的な仮定も少なくわかりやすいという利点をもっているが、多くの交絡因子で層別すると、層内の対象者数が少なくなったり、時には比較する2グループの一方に対象者がいる層がなくなったりということが生じる。そのため、効果の指標の推定値が不安定になる、あるいは効果の指標の推定自体ができなくなるという問題が存在する。回帰モデルに基づく方法とは、リンパ節転移や閉経状況などの交絡因子と乳がん再発との間に、数学的な強い仮定を設ける方法である。例えば、乳がん再発率の対数が術後からの時間に関係した効果、リンパ節転移の効果、閉経状況の効果の和で決まると仮定するモデルが比例ハザードモデル(Cox回

帰)とよばれ、医学・疫学で頻繁に用いられるモデルである⁹⁾¹⁰⁾。数式表現すると、ある対象者に対する疾病発生率(ハザードともよばれる)、 $h(t|x, z)$ に対する以下のようなモデルである。ただし、タモキシフェンの服用の有無を $x(x=1$ ならあり、 $x=0$ ならなし)で表し、交絡因子(ここでは、リンパ節転移の有無、閉経状況、ステージの3つを考える)をそれぞれ、 z_1, z_2, z_3 で表すとする。

$$\log h(t|x, z) = \log a(t) + bx + c_1z_1 + c_2z_2 + c_3z_3$$

このモデルにおいて、 $x=0, z=0$ の場合の疾病発生率 $a(t)$ は、ベースラインの疾病発生率(あるいは、ベースラインハザード)とよばれ、時間とともに変化してもかまわない。しかしながら、このモデルでは、治療変数や交絡因子の効果は時間によらず一定(比例ハザード性とよばれる)と仮定している。この条件のもとで治療や交絡因子の効果を表す回帰パラメータ(b, c_1, c_2, c_3)をデータから求める。回帰パラメータの推定には部分尤度に基づいた複雑な計算が必要であるが、統計解析ソフトを利用すれば容易に実行可能である⁹⁾。得られた結果の解釈は、例えば、 $\exp(b)$ はリンパ節転移、閉経状況、ステージといった交絡因子の影響を調整したタモキシフェン服用の再発率比(incidence rate ratio)を表している。

タモキシフェンデータに対するCox回帰の結果を示す(表5)。交絡因子の

影響を調整すると、タモキシフェンの服用は再発率を 0.83 倍 (95%信頼区間: 0.73, 0.95) にすることがわかる。この結果は、リンパ節転移と閉経状況のみを調整したマンテル・ヘンツェル発生率比の結果とほぼ同じである。また、交絡因子に関しては、リンパ節転移の有無とステージが統計的に有意な独立したリスク因子となっている (それぞれ、再発率を 3.7 倍, 1.6 倍にする)。

回帰モデルによる交絡の制御方法は、多くの交絡因子を一度に調整できるという利点をもっているが、仮定されたモデルが正しくなければ結果にバイアスを生じる。モデルはあくまでモデルであり、すべてのモデルは誤っていると考える方がよいので、強い仮定を必要とするモデルを用いた解析に過度に依存することは危険である。実際のデータ解析の場面では、単純な層別解析の結果とモデルによる解析結果の整合性を検討しておくことが重要である。

また、実際のデータ解析において交絡の調整となると、ある要因が交絡因子かどうかを判断する必要が生じる。交絡因子の見極めのためによく用いられる方法は、治療グループと対照グループの間で背景要因や共変量が t 検

定やカイ二乗検定で有意であれば交絡因子と判断しようという方法である。多くの論文でこのような方法をみかけるが、この方法は誤りである。なぜなら、統計的仮説検定とは、部分から全体を推論するために必要な演繹的論理であり、帰無仮説が否定されれば「帰無仮説は誤っている」と主張できるが、帰無仮説が否定できないときには「帰無仮説が正しい」とは積極的に主張できないからである。例えば、閉経状況が交絡因子かどうかを判断するために、「タモキシフェン服用グループと非服用グループで閉経前の人の割合が同じである」という帰無仮説を考えて、割合の差の検定 (カイ二乗検定と同じ) を行い、5%水準で有意差があれば「2グループ間で閉経前の人の割合が異なる」ことはわかっていても、5%水準で有意差がなければ「2グループ間で閉経前の人の割合が同じ」、つまり「交絡はない」とはいえず、「交絡があるとはいえなかった」という判断になる。交絡は存在するかどうかは問題なのではなく、その程度が問題である。ある要因が交絡因子かどうかを見極める際には、まずその要因が交絡因子の必要条件 (交絡因子はリスク因子である、交絡因子は比較する集団間でその

分布が異なっている、交絡因子は原因と結果の間の変数ではない) を満たしていなければならないが、その見極めには臨床的・疫学的な常識、知識などの事前情報・過去の研究成果がなによりも重要である。

おわりに

臨床研究における生物統計学の貢献は、本質的にバラツキを伴うデータ、およびその曖昧さ・不確実さの中で得られる結果に対する信憑性を確保・評価することである。これらの信憑性を支える柱が、研究計画・データ管理・統計解析の3つである。いくら素晴らしい統計解析手法を用いたとしても、研究計画とデータ管理がいい加減であれば、そこから得られる結論は弱いことが予想される。したがって、妥当で効率のよい研究デザインはどのようなものであり、プロトコルをどのように開発するか、調査票をどのように設計するかなどの研究計画の問題、モニタリング、データ入力、標準化とデータベース管理をどのように行うか、そしてそれらの品質保証をどうするかなどのデータ管理の問題をまずクリアすることが大事であることも忘れていただきたい。

表 5 Cox 回帰の結果

変数	推定値	標準誤差	再発率の比	95%信頼区間
タモキシフェン (あり v.s. なし)	-0.181	0.068	0.83	0.73, 0.95
リンパ節転移 (あり v.s. なし)	1.300	0.076	3.67	3.16, 4.26
閉経状況 (後 v.s. 前)	0.084	0.067	1.09	0.95, 1.24
ステージ	0.452	0.056	1.57	1.41, 1.76

文 献

- 1) 大橋靖雄：統計処理の落とし穴.
Diabetes Journal 24 (4) : 15-20,
1996
- 2) Altman DG (木船義久, 佐久間昭
訳) : 医学研究における実用統計学.
サイエンティスト社, 東京, 1999
- 3) Armitage P and Bery G (椿美智子,
椿 広計訳) : 医学研究のための統
計的方法. サイエンティスト社, 東
京, 2001
- 4) Matsuyama Y, Tominaga T, Nomura
Y, et al : Second cancers after adju-
vant tamoxifen therapy for breast
cancer in Japan. Annals of Oncology
11 : 1537-1543, 2000
- 5) Clark TG, Bradburn MJ, Love SB,
Altman DG : Survival analysis Part
I : basic concepts and first analyses.
British Journal of Cancer 89 : 232-
238, 2003
- 6) 大橋靖雄, 浜田知久馬 : 生存時間解
析— SAS による生物統計—. 東京大
学出版会, 1995
- 7) Rothman KJ and Greenland, S :
Modern Epidemiology, 2 nd ed. Lip-
pincott-Raven, PA, 115-134, 1998
- 8) 宮原英夫, 丹後俊郎(編) : 医学統計
学ハンドブック : 16 章 疫学, 442-
473, 朝倉書店, 1995
- 9) 松山 裕 : 生物統計学の基礎的考え
方. Surgery Frontier 10 : 415-421,
2003

別刷

Surgery Frontier

メディカルレビュー社

1) コントロール(対照群)の考え方

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

ある治療法の効果を評価するためには、別の治療法との比較が不可欠である。単に、ある治療法を実施する前と実施した後の結果を比較するだけでは、得られる結論は弱く、より適切なコントロール群の結果と比較する必要がある。本稿では、科学的な研究においてある要因に関する効果を調べる際に、なぜコントロールが必要なのか、コントロールにはどのような種類があるのかについて述べる。

Key Words

治療効果、コントロール、比較の可能性、ランダム化、マスク化

はじめに

臨床研究や疫学研究などの医学研究は、実験研究(experimental study)と観察研究(observational study)の2つのタイプに大きく分けることができる。新薬候補物質を標準薬と比較するランダム化研究は、実験研究の代表的なものであり、コホート研究やケース・コントロール研究などの多くの疫学研究は、観察研究の代表的なものである。この2つの研究方法の最も大きな違いは、前者では研究者自身はその研究で最も関心のある要因を人為的に操作するのに対して、後者ではそのような操作が行われない点である。

臨床医学系学会では Evidence Based Medicine (EBM) という言葉が大流行している。EBM とは、目の前の患者の問題点を一定の手順で定型化し、主に文献検索と抽出された文献の批判的吟味により過去の「証拠・根拠」を点検し、そこから有効な情報を引き出し、目の前の患者に対して実践することで

ある。最も高い証拠を提供するのが実験研究、とくに長期大規模ランダム化試験そして複数の実験研究を統計的に併合するメタ・アナリシスとされている。しかしながら、すべての研究のゴールは、研究目的である仮説について科学的に妥当な証拠を得ることにあり、観察研究から得られる証拠の質は、もし正しく計画された実験が可能であったならば得られたであろう証拠の質と同じであることを期待している。

本稿では、実験研究・観察研究のいずれにおいても科学的に妥当な結論を得るために要求されるいくつかの要件のうち、最も基本的な要件であるコントロール(対照群)の必要性について述べる。

なぜコントロールが必要なのか？

ある特定の個人(以下、Aさん)に関して、「アスピリンを飲むことで頭痛が治るかどうか」について考えてみる。もしAさんがアスピリンを飲んで2時間以内に頭痛が治ったとする。この

Surgery Frontier 11(1): 76-80, 2004

事実(観察)だけによって、アスピリンはAさんの頭痛を治すのに効果があつたといえるだろうか。

Aさんが何も薬を飲まなくても2時間後に頭痛が治っていたかもしれないので、その答えは否である。すなわち、Aさんに対するアスピリンの治療効果を調べるためには、「アスピリンを飲んだAさんが2時間後に頭痛が治った」という事実はそれ自体だけでは意味をなさず、以下の2つの状況を「同時に」知る必要がある。

状況1：Aさんが2003年12月1日にアスピリンを「飲んだ場合」に2時間以内に頭痛が治るか
どうか

状況2：Aさんが2003年12月1日にアスピリンを「飲まない場合」に2時間以内に頭痛が治るか
どうか

この2つの状況の結果がわかれば、Aさん個人に対してアスピリンの効果があるかどうかを調べることができる(表1)。状況1でAさんの頭痛が治り、かつ状況2では治らなければ、Aさん個人にとってはアスピリンの効果があると判断できる。しかし、どちらの状況でも頭痛が治れば、Aさんにとってはアスピリンの効果なしということに

なる。また、どちらの状況でも頭痛が治らなくてもAさんにアスピリンの効果なしということになる。これら2つの状況は、一方が観察されれば他方は絶対に観察することができないので、反事実的(counterfactual)と呼ばれる。

薬剤の評価に限らず科学的な因果推論を行うためには、人為的に操作可能な要因(アスピリン)を個人(Aさん)に対して加えた場合と加えない場合の反事実的な結果を比較しなければならない。この原理が、科学的な研究においてある要因に関する効果を調べるためにコントロールを必要とする根拠となっている。

表1に示した反事実的な結果は、実際に受けた治療によらず概念的には存在するものの同時に観察することはできない。したがって、データからは検証不能な仮定(例えば、アスピリンを飲まなかったこと以外はAさんとまったく同じ他人Bさんが存在するなど)をおかない限り、個人に対する治療効果を調べることはできない。しかしながら、このように「効果」というものを定義することによって、因果関係を調べるための2つの重要な要件を導くことができる。

1. 因果関係を調べる要因は、人為的

に操作可能な要因でなければならない

2. 人為的に操作可能な要因の使用以外の条件が全て同一のコントロールが理想のコントロールである

なぜランダム化?

個人に対して2つの治療を同時に実施することはできないので、個人に対する治療効果を調べることはできないとしても、個人の集まりである集団に対する治療効果(平均的な治療効果)ぐらいは調べることはできないだろうか。

次の仮想的な臨床研究の例を通して考えてみる。「高脂血症に対する治療法として食事療法のみよりも、食事療法に抗高脂血症薬を加えたほうが心筋梗塞の発症を減少させることができるかどうか」という仮説を調べるために、6人の高脂血症患者さんを対象とした臨床研究を行うことを考えた。試験治療(抗高脂血症薬+食事療法)グループとコントロール治療(食事療法のみ)グループ(以下、対照治療)の2グループ間で、試験開始から5年後の心筋梗塞発症割合を比較することをこの研究における主な評価項目(エンドポイント)とする。

前節で述べた治療効果の定義に従うと、この仮想的な臨床研究における6人の対象者全体での治療効果を調べるためには、やはり6人全員に試験治療を行った場合の5年間の心筋梗塞発症数と、「同じ」6人に対照治療を行った場合の5年間の心筋梗塞発症数を「同時に」比較しなければならない(図1)。

表1 ある個人に対するアスピリンの効果

		飲まない	
		治る	治らない
飲む	治る	なし	効果あり
	治らない	負の効果	なし

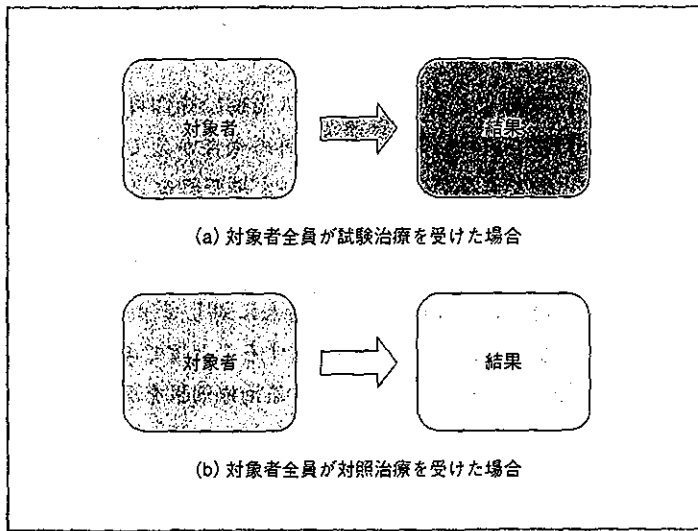


図1 臨床試験で知りたいこと
状況(a)が事実であれば、状況(b)は反事実(逆も同様)

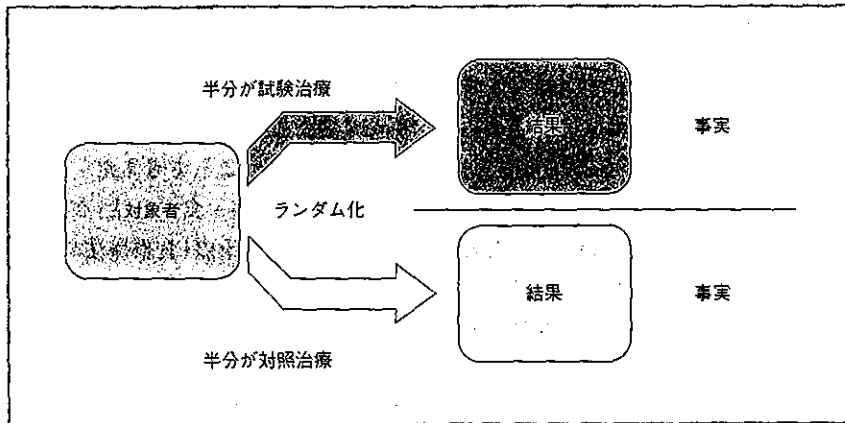


図2 ランダム化の意義

残念ながら、本質的にバラツキを伴うヒトを対象とした研究においては、すべての要因を制御できることはまれであり、現状の知識ではわからない未知の要因も存在するのが常態である。

集団に対しても理想のコントロールグループを観察することはできないが、例えば、コインを投げたりサイコロを振ったりという純粋に偶然の要素のみに基づいて対象者を2つのグループに半数ずつ分けるとどうであろうか(図2)。今の仮想例では、対象者が6人と少ないが、もし対象者数が増えれば、グループ分けに恣意性が入らず、性、年齢、総コレステロール値などの心筋梗塞発症に関わる既知の要因のみならず、生活習慣などの(われわれの現在の知識ではわからない)未知の要因をも含めて、平均的には心筋梗塞発症リスクが均一な2つのグループを作り上げることができる。すなわち、このランダム化という操作で治療に割り付けられた半分の患者さんの結果(心筋梗塞の発症割合)は、図1(a)の状況の代用として、また、ランダム化という操作で対照治療に割り付けられた半分の患者さんの心筋梗塞発症割合は、図1(b)の状況の代用として用いることができ、それら現実に観察可能な量を比較することで集団に対する治療効果を調べることができる(図2)。

ランダム化コントロールを採用したとすると、試験治療、対照治療グループそれぞれの対象者に割り付けられた内容を5年間守ってもらい、それぞれの心筋梗塞の発症割合を比較することで、「対象者全員に試験治療を行っ

つまり、理想的な研究(理想のコントロールを用いた研究)を行うためには、「その研究で関心のある結果(心筋梗塞の発症)に関わるすべての要因のうち、操作可能なひとつの要因(抗高脂血症

薬を与えるかどうか)のみがグループ間で異なる状況」を作り上げる必要がある。この目的を達成するためには、結果に影響をおよぼす可能性のあるすべての要因を制御する必要があるが、

た場合の5年間の心筋梗塞発症数と、同一の対象者に対照治療を行った場合の5年間の心筋梗塞発症数の比較」を平均的な比較可能性という観点からバイアスなく行うことができる。さらに、この治療法のランダム化という「強い」人為的な操作を行うことにより、治療効果の検定や推定といった統計計算の基礎も与えられる。

コントロールの種類

理想のコントロールグループを作成することは実際にはできないが、試験治療の効果を調べるための現実のコントロールグループの設定方法には、前節で述べたランダム化コントロールも含めて以下の4つの方法がある(表2)⁴⁾。

- ①ランダム化コントロール
- ②同時コントロール
- ③既存コントロール
- ④コントロールなし

いずれの方法も、研究者自身による人為的な操作により試験治療を行うかどうかを決定することができるが、どの方法が「理想のコントロール」に一番近いであろうか。理想的なコントロールグループは実際には存在しないが、このような研究が実施できたと仮定し、そのような研究を「比較可能性(comparability)がある」と呼ぶことにすると、①から④へ番号が大きくなるにしたがって、この比較可能性(バイアスのひとつ)を保証することが難しくなる(表2)。

①のランダム化コントロールの場合でも、上の仮想例のように対象者数が6人と少なければ、偶然の要素により

表2 コントロールのレベル⁴⁾

コントロールグループ	設定の仕方	バイアス
ランダム化コントロール	試験治療と対照治療をランダムに割り付ける	測定できない特徴も平均的には同一にできるが、偶然のバラツキによるバイアスは起こりうる
同時コントロール	試験治療を実施したグループと同時期に対照治療を実施したグループ	測定できる特徴は同一であるかどうか確認できるが、測定できない特徴はどうしようもない
既存コントロール	過去に対照治療を実施したグループの既存のデータ	記録に残っていた特徴は同一であるかどうか確認できるが、記録されなかった特徴はどうしようもない
コントロールグループなし	「対照治療で30%が改善する」といった研究者の常識(暗黙のコントロール)	バイアスがないとまったく保証できない

一方のグループに総コレステロール値の高い対象者が集まってしまうこともあり、比較の妥当性が損なわれる可能性がある。コホート研究やケース・コントロール研究などの疫学研究では②の同時コントロールを採用していることになるが、その場合、データとして測定したリスク要因に関しては比較グループが同一であることを示せるものの、未測定あるいは未知の要因に関しては同一であるかどうかを検討できない。さらに、③のヒストリカル研究では、①や②の場合と異なり、対照群に関してはこれからデータを測定するわけではないので、さらに比較の妥当性が損なわれる。最後に、④のコントロールなしは、「何もしなければ状態も変化しない」というコントロールを採用していることに相当する。物理や化学の分野では、「何もしなければ状

態も変化しない」ことが多いであろうから、このタイプのコントロールを用いて因果関係を示してきたが、個人差の大きい医学領域では④のコントロールなしで比較の妥当性を確保することは難しい。

プラセボとマスク化

前述の仮想的な臨床研究の例における、ある特定の個人が試験治療(抗高脂血症薬+食事療法)を受けた場合と対照治療を受けた場合(食事療法のみ)の比較では、厳密にいうと、抗高脂血症薬が食事療法に比べて心筋梗塞発症に効果があるかどうかを調べているのではなく、「食事療法に加えた抗高脂血症薬の服用」が心筋梗塞の発症に効果があるかどうかを調べていることになる。すなわち、「抗高脂血症薬を服用すること」という心理的効果を

含めた比較になっている。抗高脂血症薬がこの心理的効果以上の効果があることを証明するためには、「ある個人が食事療法に加えて抗高脂血症薬を服用した場合」と「その個人が食事療法に加えて高脂血症に対する有効成分を含まない薬もどきを服用した場合」を比較しなければならない。この不活性な物質だけで作った「薬もどき」のことがプラセボと呼ばれる。

対照治療としてプラセボを使用したプラセボ対照比較試験は、対照治療グループの対象者に医療上のメリットがなく、倫理的に問題となることがある。前述の仮想的な臨床研究の例では、対象者全員に食事療法を実施しているので、食事療法だけでもコレステロールが十分コントロール可能な患者さんだけを対象とするように試験の選択基準が設定されていれば、プラセボの使用は倫理的にも許されるであろう。このように標準的治療に上乘せの効果を期待する場合、プラセボの使用に倫理的な問題は生じないかもしれないが、その他の状況でプラセボの使用が倫理的かどうかは、無治療という選択が許されるかどうかにか依存する⁹⁾。

無治療が許されない場合には、プラセボを用いることは非倫理的であり、「証明された治療法」を対照治療として用いた研究を行わなければならない¹⁰⁾。この場合は実薬コントロールと呼ばれる。

前述の仮想的な臨床研究の例において、対照治療を「食事療法+プラセボ」とすることによって、薬の服用という心理的効果を排除でき、比較可能性が

保たれたが、対象者に抗高脂血症薬を服用しているかプラセボを服用しているかがわかってしまうと、試験治療グループに再び心理的効果が入ってしまう。このため、対象者にはどちらを服用しているのかわからないようにする工夫がなされる。これがマスク化(Masking)である。また、プラセボを用いることによって、対象者だけでなく医師に対してもどちらを服用しているかをマスク化することもある(二重マスク化)。対象者が抗高脂血症薬かプラセボのどちらを服用しているのかを医師が知ってしまうと、その対象者に対する医師の評価、管理にバイアスが入る恐れがある。医師に対してもマスク化することにより、これらのバイアスを避けることができる。ただし、治療の種類によっては、マスク化が不可能、あるいはすべきではない状況も存在する。例えば、投与直後の嘔気のようなその治療に特有の副作用が存在する場合には、マスク化が実際には有効ではなかったり、血小板減少が予想される薬剤を投与する場合には、血液検査、血小板輸血の準備のためにどちらの治療群かを前もって知っておく必要があったりする。これらの場合であっても、患者さんにだけはマスク化を行う、最終的に評価を行う医師にはマスク化を行う、などマスク化のレベルを変えて、可能な限りマスク化することを心がけてほしい。

おわりに

コントロールの選択の仕方としては、表2の①から順に実施可能性を考え、

どうしても実施できない場合に次のレベルのコントロールを選択すべきである。単に実施が容易だという理由だけから③や④を選択することは避けるべきで、より妥当なコントロールである①や②を採用できない理由、③や④のコントロールでも比較の妥当性が十分確保できるかどうかを示しておく必要がある。また、たとえ①のランダム化コントロールを採用したとしても、その研究の質が非常に低ければ(例えば、治療のコンプライアンスが悪い、脱落例が多いなど)、得られる結果は妥当なものにはなり得ず、②の同時コントロールを採用した良質な疫学研究の方が得られる結果の信憑性が高いことも忘れてはならない。

文献

- 1) Evidence Based Medicine Working Group : Evidence-based medicine. JAMA 268 : 2420-2425, 1992
- 2) Rothman KJ, Greenland S : Modern Epidemiology, 2nd ed. Lippincott-Raven, PA, pp. 67-78, 1998
- 3) 佐藤俊哉 : 治療のランダム割り付けと治療効果の検定. 医学のあゆみ 173 : 779-784, 1995
- 4) 佐藤俊哉 : Pコントロール. これからの臨床試験 : 橋広計, 藤田利治, 佐藤俊哉編. 朝倉書店, 東京, pp. 21-33, 1999
- 5) Altman DG : Practical Statistics for Medical Research. Chapman & Hall, London, pp. 440-476, 1991
- 6) Rothman KJ, Michels KB : The continuing unethical use of placebo controls. N Engl J M 331 : 394-398, 1994

カレントセラピー

別刷

月刊カレントセラピー [別刷] 2004 VOL.22 NO.6 **6**月号

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 ($p=0.0025$) 差を認められた¹⁾。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 ²)	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'Reilly	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	8	6.5
Cartwright	2002	カベシタピン	42	10	6.1
Okada	2002	ティーエスワン®	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/m²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/m²/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/m²/150分であることが報告された。Temperoらは、定速静注法 (1,500mg/m²/150分、週1回、3週投与1週休薬) と30分の点滴静注法 (2,200mg/m²/30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²⁾。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 + エピルビシン	44	25	10.9	23
Hess	2003	GEM + カベシタビン	27	19	6.4	33
Scheithauer	2003	GEM + カベシタビン	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ベシタビンやティーエスワン®が挙げられる。カベシタビンは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³⁾。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⁴⁾、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^{5), 6)}。

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,