データマネージャーのための統計入門

富永　祐民　　愛知県がんセンター総長

優れた臨床試験の条件

皆さんはデータ処理・統計解析を直接担当することはないかもしれないが，これからお話することを聞いていただければ，データを集めたり送ったりするときに，いかにそれが大事なことかお分かりいただけると思う。アメリカに「Garbage in，garbage out」という諺がある。garbage とはゴミのことである。良い材料，きちんとしたデータがないと如何なる高性能のコンピュータ，優秀な統計ソフトを使っても，適格な答えが出ないということである。それゆえ，良いデータを集めるという皆さんの仕事は非常に重要だということになる。

臨床試験において治療効果や安全性を評価するためには次の 4 つの well が必要だと言われている。研究のデザインが良く（well designed），忠実にプロトコール通りに実行し（well done），適切に集計・解析し（well analyzed），それを妥当に解釈する（well interpreted），ということだ。これらを順に説明していく。

To evaluate the efficacy and side effects of treatment for cancer (or any other diseases), a

-- well designed,
-- well done,
-- well analyzed and
-- well interpreted

clinical trial is necessary.

優れたデザインの臨床試験

優れたデザインの臨床試験のためには次の 3 つが必須である。まず，無作為わりつけがなされていること，これは治療群間の比較性を確保するために必要なことである（ここでは，標準的治療法に対する有用性を検証するために行う第III相試験を想定している）。無作為わりつけは，目に見えない全てのバイアス（治療法以外の要因が群間で偏ることによって，得られた結果から治療法の効果を公平に評価する妨げになる要因）も全て無作為にわりつけてしまうという利点がある。2 番目は対照群（標準的治療群）を設ける，これは新しい治療法を評価する際の物差しになる。対照群は必ずしも無治療ということではなく，その時点で確立されている標準的な治療を対照群として，新しい治療法との比較をすることになる。3 番目は，適当な対象者数で行うことであり，多ければ多いほど良いということではない。これは統計学的な有意性をみるために必要な最低限の症例数であり，非常に重要だ。

優れたデザインの臨床試験

(1) 無作為わりつけ（Random allocation）
　　治療群間の比較性の確保

(2) 対照（標準）治療群の存在（Controlled）
　　治療効果判定の基準

(3) 適当な対象者数
　　統計学的な有意性

無作為わりつけの方法としてよく使われる方法には次の 3 つの方法がある。1 つは単純ランダム化法という方法で，乱数表を使って奇数か偶数かによって A 群，B 群どちらかにわりつける。この方法だと，例数が多くない場合には 2 つの群の例数がかなり偏ってしまう恐れがある。群間の例数の偏りを少なくするためにはブロックランダム化法が使われる。たとえば 4 例を単位としてバランスがとれるようにすることを考えると，A が 2 個，B が 2 個の並べ方は 6 通りあ

無作為わりつけの方法

(1) 単純ランダム化法（Simple randomization）
　　乱数 ⇒ 奇数 ⟶ A群
　　　　 ⇒ 偶数 ⟶ B群

(2) ブロックランダム化法（Block randomization）
　　治療群間の数のバランスをとる
　　|AABB|BBAA|ABAB|BABA|ABBA|BAAB|
　　①　②　③　④　⑤　⑥

(3) 融通性ランダム化法（Adaptive randomization）
　　A=B，A>B，B>A の 3 通りの封筒の山で調整

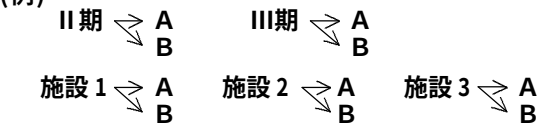
る。乱数表を使って 4 が出ればまず「BABA」とし，次に 2 が出れば「BBAA」と続け，1～6 以外の数字のときはパスして･･･，と必要な症例数分続けていく。これによって症例数のバランスをとりつつランダム化することができる。3 番目は融通性ランダム化法といって，これには色々な方法がある。Aが多くなるとBの方へ，Bが多くなるとAの法へわりつける。また，わりつけの際に考慮したい因子があるときにはそれらの因子に偏りが生じないように是正しながらわりつける方法もある。これはコンピューターを使わなければできない。

ブロックランダム化法では群間の症例数のバランスがとれるだけだが，層別化わりつけは，質のバランスもとるための方法である。たとえば，癌の臨床病期のⅡ期とⅢ期を対象として症例を集積する際に，それぞれの病期ごとにわりつける。それによって病期の偏りをなくすることができる。また，多施設共同試験の場合には施設間差が問題になることがあるので，できることなら施設ごとにわりつけることが望ましい。病期と施設を組み合わせてわりつけることもできるが，そういう場合は各施設の症例数がある程度多くないとうまくいかない。

層別化無作為わりつけ法
(Stratified randomization)

重要な因子（病期，施設など）のカテゴリー(サブグループ)別に無作為わりつけ

(例)



“治療群間の数と質のバランスを確保”

必要症例数の決め方

次に，適当な対象者数（サンプルサイズ）はどのように決めるかについてお話するが，それには統計学的有意性の検定についての基本的な理解が必要だ。試験によって得られた治療群間の差が偶然によるものか否かについて検定するに当り，２つの仮説を立てる。２群には差がないという仮説：帰無仮説と，２つの群には差があるという仮説：対立仮説である。ここで，得られた差が本当には差がない（帰無仮説が正しい）のに，偶然のいたずらで差があるように見える確率（α）を危険率という。危険率は通常，検定ではｐ値として示され，これは“取り過ぎのエラー”である。一方，本当には２群には差がある（対立仮説が正しい）のに，差がないとされてしまう確率（β）は“見落としのエラー”と言われる。βは，１－β：検出力として示されることも多く，これは２群の差の検出力の大きさを示す。ある群間差が得られたときのαとβは，症例数によって決まる。症例数が多いときはα，βともに小さくなるが，症例数が少ないときはα（取り過ぎのエラー）もβ（見落としのエラー）も大きくなる。通常，αは 5%以下，βは 20%以下となるように設定して，そのために必要な症例数を統計学的な数式により計算する。もし，非常に厳格な条件で試験を行いたいようなときには，αは 1%以下，βは 5%以下に設定する。

必要症例数（サンプルサイズ）

死亡率(生存率) A群 (B群)	B群 (A群)	[Ⅰ] α=0.05 1-β=0.8	[Ⅱ] α=0.01 1-β=0.95
0.10	0.80	8	17
0.20	0.70	17	36
0.30	0.60	44	99
0.40	0.50	390	880
0.45	0.50	1,560	3,550

A 群，B 群の条件をいくつか設定して計算した必要症例を示す。観察するのは死亡率でなくても，生存率でも，奏効率でもよい。［Ⅰ］は通常のα=5%，1－β=80%とした例数，［Ⅱ］はより厳格なα=1%，1－β=95%として計算した例数である。たとえば，A 群 10%，B 群 80%というように大きな群間差が想定できる場合（最上段の場合）は，1 群 8 例でよいが，２つの群がそれぞれ 45%と 50%というように小さな差しか想定できない場合（最下段の場合）には 1 群 1,560 例が必要となる。このような計算から，

通常の臨床試験では 1 群当り数十例から数百例が必要症例数として設定されることになる。

抗癌剤に限らないが, 臨床試験の各ステップにおける必要症例数の凡その目安を示した。第１相試験は安全性評価の試験であるので, 統計学的有意性は関係ない。各投与量で 3 例ずつで投与量を増量していき, MTD（有害事象が出た投与量）で数例追加して終了となるので，全体で 15 例から 20 例ぐらいとなる。第２相試験では厳密な統計学的な有意性は要求されないので通常 30 例以上が必要となる。第２相試験は前期と後期に分けられることが多いが，後期試験になると統計学的有意性も考慮することになるので，さらに多数が必要となる。第３相試験は有用性検証のための確認試験となるので統計学的な有意性が要求される。したがって，前述のように条件を設定して必要症例数を計算することになり，少なくとも 100 例以上，できれば 200～300 例以上で実施することになる。

抗癌剤評価のための臨床試験での必要症例数

第１相試験	各 dose 3 例 (sub MTD dose では数例追加)
第２相試験	少なくとも約 30 例以上
第３相試験	少なくとも約 100 例以上 できれば各群約 200～300 例以上

治療効果の評価のための統計手法

次に得られた結果の統計解析の話をする。治療効果の評価のための主要な統計手法を示したが, 今日はゴシック体で示したものについて話す。まず, クロス集計は奏効率, 副作用の出現率の検定などによく用いる。生命表法は生存率の解析には必須である。多変量解析法は, 治療群間の背景因子の偏りを補正して，治療法独自の効果を評価するために用いる。重回帰分析，ロジスティック型重回帰モデル，また Cox 重回帰型生命表モデル（Cox 比例ハザードモデル）と言われるものであるが，この Cox モデルは重回帰分析とロジスティックモデルを組み合わせたような手法である。

治療効果の評価のための統計手法

1．クロス集計　　単純クロス集計　2×2
　　多重分割表　　n(2×2)
2．生命表法
3．多変量解析法
　(1) 重回帰分析
　(2) 判別関数Ⅰ，Ⅱ
　(3) 数量化理論
　(4) ロジスティック型重回帰モデル
　(5) 主成分分析
4．Cox 重回帰型生命表モデル

たとえば, 対照治療群と新治療群の生存数と死亡数について, このようなデータが得られたとする（これは，有効・無効でもよい）。このような集計表をクロス集計表（この場合は 2×2 のクロス集計表という）これが偶然のバラツキによるものか, あるいはそうではなく, 治療効果の差によるものかを検定する。このようなときに最もよく使われるのは，χ²(カイ２乗)検定である。

	対照治療群	新治療群	計
生存	41 (46.6)	54 (63.5)	95
死亡	47 (53.4)	31 (36.5)	78
計	88 (100.0)	85 (100.0)	173

右のように 2×2 のそれぞれの数値（度数）a，b，c，dのデータがあったとする。検定のための統計量（χ²値）の計算式は次の 2 つがある。

	A群	B群	計
生存	a	b	a＋b
死亡	c	d	c＋d
計	a＋c	b＋d	N

$$\chi^2=\frac{(ad-bc)^2\times N}{(a+c)(b+d)(c+d)(a+b)} \quad \cdots式1 \qquad \chi^2c=\frac{(|ad-bc|-\frac{N}{2})^2\times N}{(a+c)(b+d)(c+d)(a+b)} \quad \cdots式2$$

式 1 の方は単純な χ^2 値の計算式である。これに対し式 2 は，連続補正を行った χ^2c 値（c は補正;corrected）を計算する式で，分子のところで N/2 を引くことによってやや控え目な数値が得られ，例数があまり多くないときにはこちらを用いた方が無難のようである。これらは電卓でも簡単に計算できる。式 1 あるいは式 2 で求められた χ^2 値が，もし 3.84 よりも大きければ 2 群間の死亡率の差が危険率 5%以下で有意であるといえる。さらに，もし 6.63 よりも大きければ危険率 1%以下で高度に有意差があるといえる。

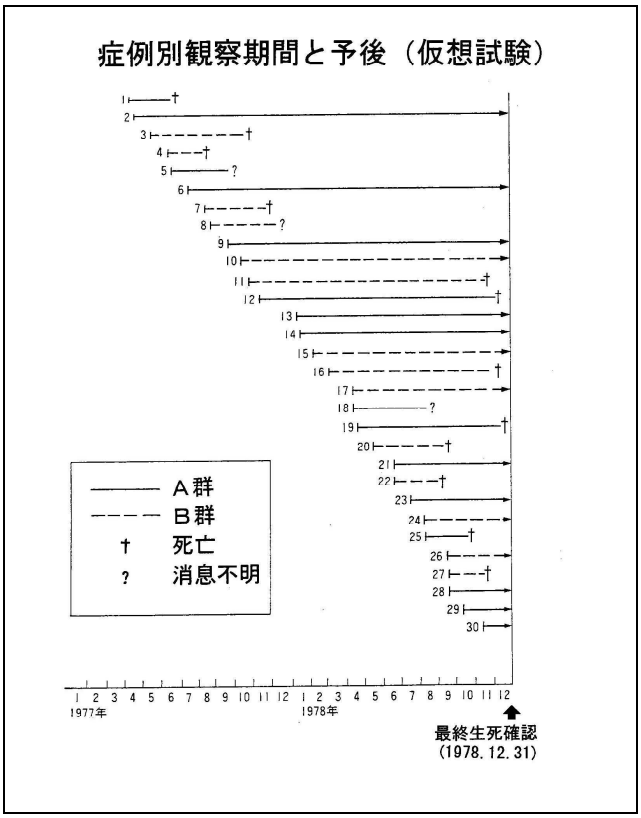
実際に前述のデータで，これを計算してみると，

$$\chi^2=\frac{(41\times31-54\times47)^2\times173}{88\times85\times78\times95}=5.01 \qquad \chi^2c=\frac{(\left|41\times31-54\times47\right|-\frac{173}{2})^2\times173}{88\times85\times78\times95}=4.35$$

となり，χ^2，χ^2c いずれも 3.84 より大きい。したがって，この 2 群の死亡率の差は統計学的に危険率 5%以下で有意であるといえる。

生存率の計算方法

次に，生存率の解析に用いる生命表法について説明する。生命表法は観察期間がマチマチな場合に使われる方法である。実際の臨床試験では，症例が一斉に登録されてスタートするのではなく，今日 1 例が登録されて，3 日後に 2 例目が登録され，･･･というように進んでいく。したがって，ある時点で解析しようとしたとき，症例によって登録後 1 ヵ月しか経っていない症例もあれば，2 年経過した症例もあるというように，症例によって観察期間が異なることになる。



このようなデータから，どのように生存率を計算するかというと，ある症例が最初の観察期間から i 番目の観察期間の終わりまで生存している確率 P_i は，次の式で得られる。

$$P_i = p_1\times p_2\times p_3\times\cdots\times p_i$$

この P_i は，累積生存率という。

これを計算するためには，まず，i 番目の観察期間の死亡率 q_i を計算する。

$$q_i=\frac{d_i}{l_i-\frac{1}{2}u_i-\frac{1}{2}w_i}$$

ここで，

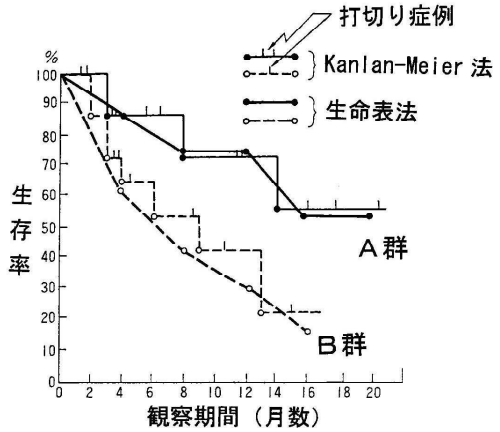
d_i = i 番目の観察期間の死亡者数
l_i = i 番目の観察期間の当初生存数
u_i = i 番目の期間中の脱落者数（プロトコールからの脱落および追跡不能者数）
w_i = i 番目の期間の途中で観察が中断している症例数（打切り症例）

先ほどの各期間の生存率 p_i は，$p_i=1-q_i$ により求まる。

前ページの図のデータからこの方法で求めたA群，B群の生存率をグラフにプロットすると，太線のようになる。この計算方法は元来，生命保険数理士が使っていたので生命保険数理法とも呼ばれる。あるいは，Cutler-Ederer 法と呼ばれることもある。

その後，Kaplan-Meier 法という方法が開発された。生命表法が観察期間ごとに計算するのに対し，Kaplan-Meier 法は1例ごとに生存率を計算する。そのため，症例数が少ない場合でも正確に生存率を計算することができる。ある程度症例数が多くなれば，いずれの方法でも生存率はほとんど変わらない数値が求まる。

Kaplan-Meier 法と普通の生命表法による生存率曲線の比較



2 群の生存率の差の検定法にはいくつかの方法がある。最もよく使われる方法は，ログランク検定と一般化 Wilcoxon 検定である。それぞれに特徴はあるが，どの方法で検定しても概ね同様の結果となる。生存率の群間差を検定した際には，どの方法を用いたかを明記しておく必要がある。

2 群の生存率の差の統計学的有意性の検定方法

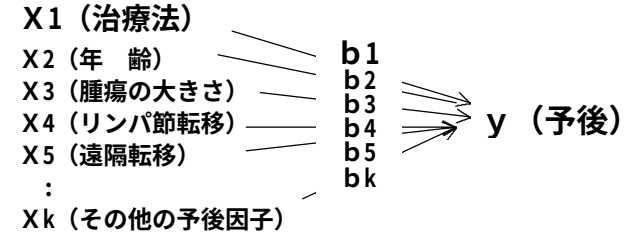
(1) 累積生存率 Pi の標準誤差 S.E.に基づく検定法
(2) Mantel-Haenszel 検定
(3) ログランク（logrank）検定
(4) 一般化 Wilcoxon 検定
(5) Cox-Mantel 検定

多変量解析の長所と限界

癌の臨床試験のように，背景因子が予後や効果に大きく影響を及ぼす可能性がある場合には，多変量解析が有用となる。多変量解析の利点としては，(1)症例数が少なくても同時に多数の因子を考慮して各因子独自の予後因子としての重みを推定しうる。(2)治療法も 1 つの予後因子とみなせば、治療群間の背景因子の偏りを統計学的に補正することができる。ということが挙げられる。ただし，多変量解析における統計学的補正はあくまで“補正手段”であり、無作為わりつけによる比較試験に置き代る方法ではないとうことに注意する必要がある。

癌の臨床試験における予後（生死）は，色々な因子によって規定されているが，多変量解析においては治療法も予後因子の 1 つとみなす。このような様々な因子が，それぞれ独自に並列で予後を規定していると仮定する。

治療法および各種の予後因子と予後との関係



（b1，b2，･･･bk は各因子と予後との関連度を示す）

これを数式で表す。yが予後（0；生存，1；死亡）を表し，X1 は治療法，X2～Xk は予後因子（背景因子）を表す。重回帰モデルというのは線型モデルと言われるもので，非常に単純な数式で表される。これをコンピューターで計算すると，治療法およびその他の予後因子についての係数，b1，b2～bk が求められる。これで得られたb1 は，他の因子の偏りを全て補正したうえでの，治療法独自の予後に対する重みを表すことになる。

重回帰分析モデルにおける予後と治療法および各種の予後因子との関係

$y = a + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k$

y：予後　　X1：治療法　　X2～Xk：予後因子
X2：年齢
X3：腫瘍の大きさ
⋮
Xk：その他の予後因子

Cox 重回帰型生命表モデル（比例ハザードモデル）は，数式は指数関数を含んで複雑にはなるが，考え方としては重回帰分析モデルと同様に，各因子独自の係数b1，b2～bk をコンピューターにより計算することができ，他の因子の影響を補正したうえでの，2 群間の生存率の差を見積ることができる。

$$\lambda(t;\underline{x}) = \exp(\underline{x}\underline{\beta})\lambda_0(t) \quad \cdots\text{式}1$$

$$\log_e\left(\frac{\lambda(t)}{\lambda_0(t)}\right) = b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2) + \cdots + b_k(x_k - \bar{x}_k) \quad \cdots\text{式}2$$

ここで，

$\lambda(t;\underline{x})$：瞬時 t における瞬間死亡率
$\underline{x}$：x_1，x_2，…，x_k；k 個の予後因子
$\underline{\beta}$：b_1，b_2，…，b_k；k 個の予後因子の係数
$\lambda_0(t)$：x_1，x_2，…，x_kのすべてが平均値のときの瞬間死亡率

生存率の統計的手法の選択は，同時に考慮すべき予後因子の数が少ないか多いか，観察期間が一定か症例によってマチマチかによって，使い分けることになる。

治療効果判定のための統計的手法の選択

同時に考慮すべき予後因子数	観察期間	統計的手法
少ない（2～3 個まで）	一定	2×2分割表（χ^2テスト） 多重分割表（M-Hテスト）
	不定	生命表法
多い（数個以上）	一定	多変量解析法 （重回帰分析 ロジスティック型重回帰モデル）
	不定	Cox 重回帰型生命表法

治療群間の偏りの防止と補正についての考え方としては，まず，(1)重要な予後因子(臨床病期、組織型など）で層別化してランドマイズするということが大事である。(2)無作為わりつけを行っても治療群間に偏りが生じることはあり得る。このようなときには，Cox 重回帰型モデルやロジスティック型重回帰モデルによって補正して解析する必要がある。(3)治療群間の予後因子(背景因子)の分布の差が統計学的に有意でなくても、予後因子(背景因子)の偏りの影響がみられることがある。病期分類などの非常に重要な予後因子は，統計学的に有意な群間の偏りがない程の僅かな偏りでも影響があり得るので補正因子に加えるべきである。(4)治療群間に予後因子(背景因子)について偏りがみられるときは、偏りを生じた因子のカテゴリー別に集計解析するか、他の適当な統計学的な補正（多変量解析）を行う必要がある。

多施設共同試験の際の注意点をまとめた。施設間の差は少ないほうがよく，特に特殊な技術，評価法を含む試験の場合は統一した手技・基準等で行うよう十分申し合わせて実施する必要がある。また，施設間の差の影響を少なくするためには，施設ごとの症例数ができるだけ多くなるようにした方がよい。

多施設試験実施上の問題点

1. 施設ごとにランドマイズして施設間格差の影響を少なくする（1 施設当りの症例数が多いとき）
2. 一定の症例数を確保しようとするとき、各施設ごとの症例数をできるだけ多くして施設数を少なくした方がよい
3. 施設間の診断・治療技術格差が少ない方がよい
4. 施設ごとの集計・解析結果は参考資料にとどめる
5. 研究報告は研究グループ名または全参加者の名前で行うことが望ましい

脱落・除外例の取扱い

ランドマイズドトライアル（無作為わりつけ試験）の場合は，できるだけ脱落例を解析対象症例から除外しないようにすべきである。これを「intent-to-treat analysis」の考え方という。必ず全登録症例を解析対象としなければならいということはないが，解析結果を示すときには全登録例，適格例，完全例はそれぞれ何例あったか明記する必要がある。

除外･脱落･判定不能例の取扱い（不適格･不完全例）

1. 除外･脱落の定義と取扱いは試験開始以前に決めておく
2. 除外･脱落をできるだけ少なくするようなデザインのプロトコールを考案する
3. ランドマイズドトライアルの場合はできるだけ(除外)脱落例を集計解析対象から除外しないようにする
   特に、副作用による治療中断･不完全治療例は集計解析対象から除外すべきでない
   - A 全登録症例
   - B 全適格症例（不適格例のみ除外）
   - C 適格完全例（解析可能例）

後層別解析の注意事項

解析対象症例全体としてみると治療群間に有意の差がみられないときに，特定の背景因子で層別して，どこかで効果（群間差）がみとめられないか探索することがある。これを後層別という。この後層別で特定のサブグループで有意の差がみられたときは，原則として参考所見にとどめておき，断定的な結論をくだすべきではなく，そのことを検証するための新たな臨床試験を行う必要がある。

統計学的有用性と臨床的有意性

これまでは統計学的な有意性だけを問題として話してきたが，最終的には，実際の生存率の差あるいは生存期間の延び（延命効果）を臨床的立場からその価値を判断して，総合的に得られた結果を考察する必要がある。

臨床的有意性についての判断基準

統計学的有意性の判断基準：$p < 0.05$ ($p < 0.01$)
臨床的　有意性の判断基準：？

抗癌剤(癌治療)における延命効果の判定基準案
- A ５年生存率の差＞5%
- B 50%生存期間　＞1.3 倍（最低＋50 日）

おわりに

今日お話した解析方法の中で，χ^2 検定と生命表法は電卓でも計算できる。重回帰分析や Cox 回帰分析は，以前は大型コンピューターを使わなければ出来なかったが，最近は解析ソフトを使えばパソコンでも手軽にできるようになっている。とはいえ，元々のデータの質が悪ければ，解析専門家がいくら頑張っても救いようがなく，適切な結論を導き出すことはできない。したがって，最初にも話したが，正確なデータを収集する皆さんの仕事は非常に重要だ。「Garbage in, garbage out」という言葉を覚えておいてほしい。

参考図書：治療効果判定のための実用統計学 －生命表法の解説と臨床試験の実際－（第 3 回改訂版）1991 年　富永祐民 著．蟹書房 発行／癌と化学療法社 販売