

## 10、介入研究のストップ—2つのSequential Designによるサンプルサイズの減少への挑戦

介入研究は、介入による影響が明確になった時点で研究の進行をストップさせることが、倫理上またコスト上、要求される。どのような時に研究をストップするかは、あらかじめ研究プロトコルの段階で決定されていなければならない。研究ストップの概要をここに述べておく(Lee 1992)。これはいわゆる最終段階の臨床試験における問題である。

介入研究の研究目的には、大まかに分けて主に二つある。即ち、ベストな治療法(介入法)を選択する目的と、治療(介入)による影響の程度をできるだけ正確に推定する目的である。研究をストップすることが倫理的であるがゆえに行われるのは、前者が介入研究の目的である場合であり、後者が介入研究の目的の場合は暗に研究は計画の終わりまでストップされないことを意味する(Lee 1992)。しかし、後者が目的でも、あまりにも介入による影響が明白な場合は研究計画は途中で終了することが考慮されるべきである。

次に、ここでは、Wald(1947)により発達させられた古典的Sequential 法と Armitage(1975)により論じられた臨床試験へのSequential Designの適用について簡潔に述べる。Open Sequential DesignとClosed Sequential Designについて簡潔に述べる。

### 1) Open Sequential Design

この古典的なsequential手順においては、患者がペアで研究に参入する；一人はランダムに治療Aを割り当てられ、もう1は治療Bを割り当てられ、治療がどのように割り当てられたかは、この時それぞれのペアについて記録される。いくつかの症例において、患者は研究に引き続いて(sequentially)参加する。そして、それぞれの患者は、2つの治療を、連続して受ける、例えば2週間というふうに。治療の順番はランダムであり治療は患者に区別されないようになっている。第二の期間の終わりにおいて、それぞれの患者はAとBどちらか良かった方の治療

が与えられる。これが、一致しなかったペア(untied pair)として認識される。どちらかの症例において、データはAかBに関してのシリーズになる。もし二つの治療がそれぞれの患者において同じ結果をもたらしたのなら、これは、一致したペア(tied pair)と認識される。分析は、全体として、一致しなかったペアに基づく。もし二つの治療等しく有効であるなら、AとBを希望する数が等しくなることが期待される。この検定は、この時、患者数に等しいもしくは患者のペアに等しいnの二項分布に基づく。そして $p=1/2$ である。表は、典型的なOpen Sequential Designの例である。2つの薬AとBが1つのペアに与えられたと想定しよう。一列目はペア番号、二列目はペア間でどちらが良かったか、を示している。A(x)で表す好みの累積数は、三行目に、比較するために四行目には境目(critical values)が示されている。両側のcritical valuesは5%の有意水準で二項分布から与えられる。この例においては、nはペアの数である。critical values(a,b)はこのように得られる。

$$P(x \leq a | n, p=1/2) \leq 0.025$$

そして、

$$P(x \geq b | n, p=1/2) \leq 0.025$$

この試験は、 $x \leq a$ もしくは $x \geq b$ (critical region)の時に終わる。 $x=10$ がcritical regionに落ちる時、 $n=12$ まで、結果は有意でなかった。従って、12番目のペアの後に、調査者は、治療間に有意な違いがあり、試験が終わったと結論した。

明らかに、結論がなしえて試験が終わるまでに、大きな患者数が必要になる。試験が開始される前に必要とされる患者数を管理する方法はない。

Open Sequential Designの表

患者ペア数	どちらが良いか	累積のA(x)が好みの数	Critical Values
1	A	1	—
2	A	2	—
3	A	3	—
4	B	3	—
5	A	4	—
6	A	5	(0,6)
7	A	6	(0,7)
8	A	7	(0,8)
9	B	7	(1,8)
10	A	8	(1,9)
11	A	9	(1,10)
12	A	10	(2,10)

2) Closed Sequential Design

これはArmitage(1975)により示されたデザインである。Open DesignとこのDesignとの主な違いは、このデザインは、調査者が2つの治療法に違いがないという結論をすることができる。

図3にClosed Sequential Designの例を示した。境界の点は、2項分布もしくは一致しないペアの場合は正規近似の表から得ることができる。次々と得られる結果は、原点から出発したジグザグ線で示される。それぞれがAを好むときには、線は右上に1つの固まりで描かれて小さな四角の斜線を形作る。それぞれがBを好むときには、線は右下に1つの固まりで描かれて小さな四角の斜線を形作る。

結果は、境界線の1つが横切られるまで入力され続ける。Closed Sequential Designの図4の実線は下の表から描かれ、治療Aが良いという方向で有意な違いが示されている。しかし、第2の図の破線は、2つの治療の間にさほど違いはなく、試験は終了させるべきである。

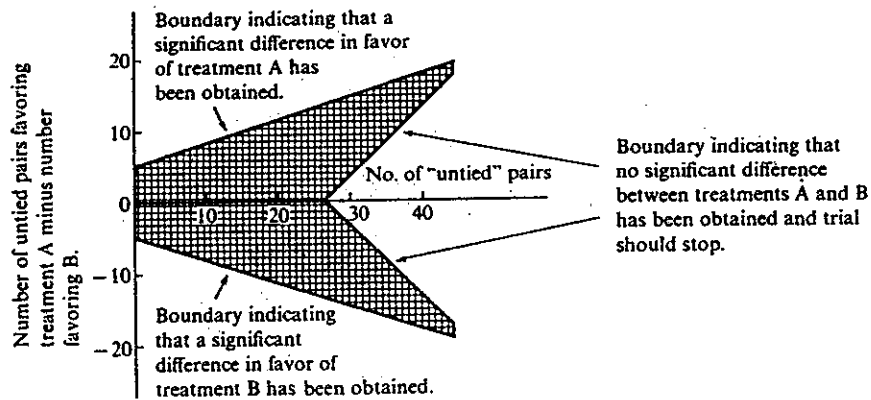


図3 Closed Sequential Design

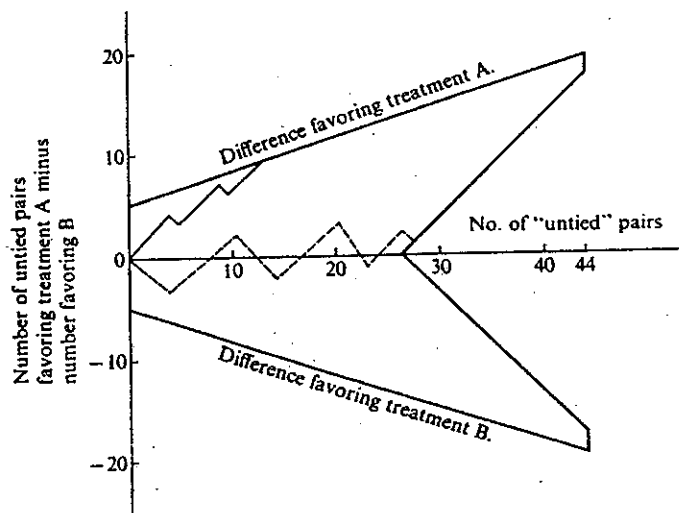


図4 Closed Sequential Designで起こりうる結果

Johnson & Pearce(1990)は、このデザインを多嚢胞性卵巣病の患者に排卵を誘発するために使用した。標準の治療は、clomipheneの投与であったが、これは40-50%の妊娠に自然流産を起こす可能性がある。著者らは、clomipheneで排卵を導入するか、buserelinで下垂体を抑制しその後純粋卵胞刺激ホルモンで排卵を導入する

か、のどちらかの後に、過去に自然流産経験のある多嚢胞性卵巣病48人の女性で、自然流産の率を比較した。患者のペアはお互いにコインをトスして治療が割り当てられた。方向性は、ペアの両方が妊娠しそのうち一人が流産した症例として定義された。20症例の参加の後、11はbuserelinが良いと観察され、clomiphenが良かったのは2例であった。7例に関してはどちらが良いかは決まらなかった。buserelinが良いことを示す境界線が切られて、結果は0.01レベルで有意であった。

境界線を築くために二項分布を使うことに加えて、他の定型化を使うことができる。Montaner(1990)は三角検定が使われた例を示した。これは、三角の受容域からなる。この検定は、関心のある結果が二種類の時に、対数オッズ比により測定された二つの治療法の違いに基づく。調査者は、sequential な手順が使われるかどうかをあらかじめ決定しなければならない。長所と短所は次の通りである。

#### 1) 長所

- a. サンプルサイズは固定標本手順より平均的に小さい。
- b. もし、1つの治療法がその他の治療法より有効であるという強い初期の証拠がある時、早期の決定に到達することができる。
- c. ほとんどの事例においては、sequential な試験の方が簡単に試行できる。

#### 2) 短所

- a. 試験を継続するかどうかの決定がすぐにできるほど、患者からの結果が利用できない可能性がある。例えば、反応が生存時間であるなら、結果が決定できる前に、数ヶ月もしくは数年が経過してしまう可能性がある。
- b. だいたいサンプルサイズが固定標本試験より小さいながら、その保証はない。いくつかの状況では、sequential な試験もたくさんの数の患者を要する。
- c. 最初の決定は少数の患者からなされるので、この治療の有効性は小さな患者群から推定されたのに過ぎない可能性がある。

## 11、固定標本試験(fixed-sample trials)におけるサンプルサイズの計算

サンプルサイズの計算には、 $\alpha$ エラー(第1種の過誤)の水準の設定、 $\beta$ エラー(第2種の過誤)の水準の設定、介入群の期待損失割合もしくは期待生存割合、非介入群の期待損失割合もしくは期待生存割合(従って、介入群と非介入群の期待損失割合もしくは期待生存割合は異なっている必要があり、両群の期待損失割合もしくは期待生存割合の差や比と片方の群の期待損失割合もしくは期待生存割合でも良いことになる)、と両群の数の比が与えられる必要である。両群の損失割合もしくは生存割合で介入による影響を推定するのではなく、両群の与える結果が連続量で示される場合には、さらに両群の差の平均と分散等の情報が必要である。

サンプルサイズの計算は、Clinical Trials, a practical approach (Pocock 1983)、Clinical Trials, design, conduct, and analysis (Meinert 1986)、無作為割付臨床試験(浜島1993)Statistical methods for survival data analysis; second edition (Lee 1992)等のテキストに詳しく記載されているのでここでは省略する。

あらかじめ意味のある研究結果を得ることができるサンプルサイズを予測しておくことには、介入群もしくは非介入群に入ることのデメリットが大きいときは、介入研究の参加者をできるだけ小さくするという倫理的意味がある。

## 12、分析

分析に先立って、生存分析(survival analysis)の基礎を修得しておくことは、応用可能性を広げるために役に立つと考えられる。ここでは、生存分析の基礎である連続確率分布について解説する。

### 基本的考え方—連続確率分布(Lee 1992)

生存時間は、故障・死亡・反応・再燃・ある病気の発症などのような一定の事象(損失)までの時間を測定するデータである。生存時間の分布は通常3つの関数によって記述される。即ち、(1)生存関数 $S(t)$ 、(2)確率密度関数 $f(t)$ 、(3)ハザー

ド関数 $h(t)$ 、である。これらの3つの関数は数学的に同等である。すなわち、これらの内の一つでも導き出されれば、あとの2つも求めることができる。Tを生存時間とすると、Tは上記の等しい関数で特徴づけることが可能である。それぞれの関数を、介入群と非介入群それぞれにおいて求めることにより、介入による影響(非介入群に比較して比較群の差もしくは比として求めることができる)を推定することができる。

(1)生存関数の定義：この関数は $S(t)$ により、 $t$ 時より長生きした人の確率として定義される。

$$\begin{aligned} S(t) &= P(t\text{時より長生きした人}) \\ &= P(T > t) \end{aligned}$$

Tの累積分布関数 $F(t)$ の定義から、

$$\begin{aligned} S(t) &= 1 - P(t\text{時より前に死んだ[事象を起こした]人}) \\ &= 1 - F(t) \end{aligned}$$

後に述べるセンサリングがないとき、

$$\begin{aligned} S(t) &= 1 && \text{For } t=0 \\ S(t) &= 0 && \text{For } t=\infty \end{aligned}$$

$S(t)$ は累積生存率とも呼ばれ、 $S(t)$ のグラフは生存曲線と呼ばれる。生存関数もしくは生存曲線は、50パーセンタイルや、その他のパーセンタイル(例えば25や75)を求めるのに使われ、二群もしくはそれ以上の群の比較に使われる。平均は、分布の中心的な傾向を描出するのに使われるが、中央値の方が通常よく使われる。なぜなら長いもしくは短い生存時間の小数の人は、平均生存期間では不均等に長いか短くなってしまふからである(数が少ないとき極端な値によって平均が引っ張られ、全体の特徴を表せないこと)。

実際には、センサードがないとき、生存関数は $t$ 時より長い生存者の占める割合で推定される。

$$S(t) = (t\text{時より長生きした患者の数}) / (\text{総患者数})$$

## (2) 確率密度関数(もしくは密度関数)の定義

$\Delta t$  の間隔を持つ  $t$  時と  $t + \Delta t$  時の間に事象が生じる確率の極限として、生存時間  $T$  が持つ関数。以下のように定義される。

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P\{\text{間隔}(t \text{ 時}, t + \Delta t \text{ 時})\text{において個人が死ぬ}\}}{\Delta t}$$

1. この  $f(t)$  のグラフは、負にはならない関数である：

$$\begin{aligned} f(t) &\geq 0 && \text{すべての } t \geq 0 \\ &= 0 && t < 0 \text{ の時} \end{aligned}$$

2. 密度曲線と時間軸( $t$ 軸)に囲まれる面積は1である。実際に、もしセンサードした症例がない場合は、確率密度関数  $f(t)$  は、単位間隔において死亡した(事象を起こした)症例の占める割合として推定される。

$$f(t) = \frac{\text{t 時において始まる時間間隔の中で死亡した患者の数}}{(\text{総患者数}) \times (\text{時間間隔の幅})}$$

$S(t)$  の推定と同様に、センサード観察が存在するときは、上式は適用できない。いずれかの時間間隔において死亡した個人の占める割合や、最も死亡頻度が高いピークは密度関数から見つけることができる。密度関数は、無条件損失率 (unconditional failure rate) として知られる。

## (3) ハザード関数

生存時間  $T$  のハザード関数  $h(t)$  は、条件付きの損失率 (conditional failure rate) を与える。これは、個人がごく小さな時間間隔の始まりの時点まで生き残っていることを仮定してその時間間隔の間の損失確率として定義される。

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P\{\text{間隔}(t \text{ 時}, t + \Delta t \text{ 時})\text{において、t歳まで生きた個人が死ぬ}\}}{\Delta t}$$



ハザード関数は、累積分布関数 $F(t)$ と確率密度関数 $f(t)$ で次のように定義することもできる：

$$h(t) = f(t) / \{1 - F(t)\}$$

このハザード関数は、瞬間損失率(instantaneous failure rate)、死亡率(force of morbidity)、条件付き死亡率(conditional mortality rate)、年齢別損失率(age-specific failure rate)とも呼ばれる。これは、ある個人のある年齢の関数としての損失の傾向の測定値である。その意味で $\Delta t \times h(t)$ という量は、年齢 $t$ 歳の人が、 $t$ から $t + \Delta t$ までの短い時間で死ぬ期待値である。それ故、ハザード関数は、老化過程の間の単位時間あたりの損失リスクを与える。これは生存データ分析において重要な役割を果たす。

実際、センサード症例がないときに、ハザード関数は、単位時間間隔において死亡する患者の割合として推定される。その患者たちは、その時間間隔の開始時には生き残っていた者たちである。

$$h(t) = \frac{t \text{ 時において始まる間隔において死ぬ患者の数}}{(t \text{ 時において生き残っていた患者の数}) \times (\text{時間間隔の幅})}$$

$$= \frac{\text{その時間間隔において単位時間あたりに死んだ患者の数}}{t \text{ 時において生存している患者の数}}$$

保険統計士は通常、その時間間隔における単位時間に死亡した患者の数を、その間隔の中央点における生存者の平均数で割って、その時間間隔の平均ハザード率として使う：

$$h(t) = \frac{\text{その時間間隔における単位時間あたりに死亡した患者の数}}{(t \text{ 時において生きている患者の数}) - 1/2 (\text{その時間間隔における死亡者数})}$$

この保険統計士式の推定法は、上式より高いハザード率を与える。

### 1 3、データ管理(Hertge 1998)

実験的研究は、患者(臨床試験)、ボランティア(フィールド試験)、もしくは共同体(communities)全体(共同体試験)に、ランダムに曝露を割り当てる。臨床試験と共同体試験についてはここでは述べない。フィールド試験には、予防であろうとスクリーニングであろうと、コホート研究と同じフィールドでの作業を含むだけでなく、募集やランダム化という特有の作業もある。参加者は、対象者にかなりの要求をするいずれの試験の利点についても納得している必要がある。もしその試験が特定の人口集団に的を絞っているのなら、この的が対象者を集めることになる可能性のある母集団を示していることもある。もし名前と電話番号もしくは住所のリストが利用できれば、研究スタッフが個別に前向きの募集方法が行えるので、募集は最も簡単である。例として、食事と類腺腫性 adenomatous ポリプの再発との関係を調べるための試験において、調査者は参加した消化器科医からの紹介や参加した内視鏡サービス機関の医学記録の総覧によって、参加が可能な者を定めた例がある。

募集の可能な者のリストがないときには、調査者は、公的な掲示、ニュース、広告によって募集しなければならない。重要人物と住民組織もしくは医学組織が関心を高めることによってスポンサーになってくれたり支持してくれることを、しばしば調査者は要請することになる。ほとんどの募集は、対象者自身(例えば、無料医学検査)と社会に対して試験の価値を説得することによって成功する。試験への参加にかかる駐車料金もしくは他の少額のコストを払い戻す試験もある。もし研究による時間的拘束もしくは身体への負担が大きければ(例えば何度も採血をする)、金銭的補償が必要になるかもしれない。Hartgeらの経験上では、25ドル(現状のドル)以下の金銭的な誘導では、参加者を著しく増加させることはできない。一方、やや多額を支払えば、参加に影響を与えることができる。100ドルを越える支払いは珍しいが、参加者に多くを要求する研究では必要かもしれない。

ひとたび参加を募集すれば、対象者は乱数表(紙上もしくはコンピューター上

の)を使って試験のどれかの群にランダムに割り当てられる。ランダム化は実験群間の参加者の類似性を保証しないので、調査者は主な臨床結果の予知因子を測定するために研究の起点での質問票を使うべきである。調査者はまた、追跡期間の終わりまで参加者の居所についての情報を集める(例えば、友人の名前と住所)。多くの介入試験では、質問票を郵送し、返事がない人には電話をかけることによって、毎年もしくは他の定期的な追跡を必要とする。もし臨床結果もしくは曝露についての新しいデータが必要ではなくても、住所については現時点のものを把握しておく必要がある。

#### 1 4、追跡tracingと追跡調査follow-up(Hertge 1998より改変)

いくつかの研究では、対象者との接触は必要ではなく、死亡した人の死亡日時と場所を含む生存状態の決定のみが必要となるので、この場合は死亡診断書を得ることになる。合衆国では主な生存状態に関するデータ源としては、the National Death Indexや、Social Security Administration files, Veterans Administration filesがあるが、日本においては手続き上、手に入れることはなかなか困難である。一軒一軒遺族宅を回るか、県庁を通して厚生省統計情報部に申請する必要がある。費用と研究カバーの範囲はこれらの点で異なってくる。

もし研究対象者が臨床結果や曝露について質問票で質問されるならば(積極的追跡調査active follow-up)、彼らはまず居場所を確かめられなければならない。氏名、住所、電話番号、子供に関しては両親の氏名、配偶者の氏名、雇用主についての最近の精確な詳細情報を用いて、調査者は最新の住所に郵送するか対象者に電話することにより調査を始めるだろう。その様な詳細な情報がなければ、調査者は対象者の居場所を捜し当てるためにいろいろな方法を使う。もし多くの対象者が死んでいれば、生存状態についての調査は最初のステップとなるだろう。

死亡したかどうか分からない対象者については様々な方法で追跡できる。市町村の住民台帳や合衆国郵便収集サービスは対象者の居所を突き止めるために

使うことができる。クレジットの事務所にある記録から回答者が生きているかどうか分かるし(例えばクレジットの使用日時から)、住所も提供されうる。運転免許証の記録では、対象者が最後の更新時には生きていたことが分かり、住所をつかむことができる。自動車登録と電話リストは生存状態を示さない。なぜなら、家族が死亡者の名前を取り除かずにこれらを使うことができるからである。医療保険支給当局の記録には、メディケアの資格のある人の名前や、年齢、住所が含まれている。

分かっている最新の住所の日付けが古ければ古いほど、追跡は難しくなる。一方多くのコホートメンバーが死亡しているならば、追跡は比較的簡単である。特に国営死亡索引が始まった1979年以降は簡単である。追跡のための手続きはコホートの構成によって違う。合衆国における職業コホート研究では、調査者はしばしば社会保障番号を知っている。これは死亡やクレジットの事務所、他の記録と照合する際の助けになる。医療コホートでは、データの確認はしばしばまばらで、比較的不正確であるので、コホートの大部分を見つけ出すために多くの情報源が必要であろう。女性のコホートの追跡(例えば姓名の変更から)や生まれる前あるいは誕生時に曝露されたコホートの追跡においては、さらなる問題が生じる。

積極的な追跡を必要とする研究に関しては、対象者を追跡し、臨床結果データを収集することが通常相前後してなされる。質問票を郵送するか、電話でインタビューするか、もしくは対面してインタビューするかに関する手順は、ケース・コントロール研究における手順と類似している。1つの典型的な手順は、追跡質問票を一度送り、それが返ってくるのを数週間待ち、非回答者にもう一度郵送し、それから残りの非回答者に電話することである。第2回目の郵送および電話は、どちらもほとんどの研究において回答を著しく増加させる。

縦断研究では繰り返しの追跡調査を必要とし、特別な問題を呈する。ある環境下では、1回目での非回答者に対しては2回目で近づくことができる。これらの継続した接触により、最初の追跡よりも多大な回答数に到達することができる。

なぜなら、参加したくない対象者は通常最初の時点で拒否するからである。一方、もしコホートにおいて何度も追跡調査がなされるならば、さらなる質問に回答する動機が揺らぐかも知れない。追跡調査毎に研究の成果と進行状況を記述した手紙を送ることは、対象者が参加する意志を持ち続ける助けとなる。いくつかの研究では、一度対象者が拒否すると彼らはその後の追跡調査では接触されない。それぞれの追跡調査において例え良い回答率であっても、非回答者に対して何度も送り続けることは下に示すように累積回答率を下げるからである。

#### 縦断研究の繰り返しの追跡調査における回答率

1986年	78%回答	22%非回答	→ 接触なし
1989年	92%回答 <sup>a</sup>	8%非回答	
1993年	93%回答	7%非回答	42%回答 <sup>a</sup> 58%非回答
1993における累積			
	$(.78 \times .92 \times .93) +$		$.22 + (.78 \times .92 \times .07) +$
	$(.78 \times .08 \times .42) = 69\%$		$(.78 \times .08 \times .58) = 31\%$

#### 15. 死亡と疾病発生の確認と詳細(Hertge 1998より改変)

研究ではしばしば、追跡段階での質問票で回答者もしくは彼らの近親者により報告された臨床結果を、他の情報源からのデータで確認することが要求される。例えば死亡の事実が分かれば、死亡の原因は死亡診断書から収集できる。

合衆国では州によっても異なるが、州から死亡診断書を入手するには、詳しい申請書が必要で、6から12ヶ月かかる。支持している制度化された検討委員会が是認する文書や研究プロトコールが必要な州もある。対象者のリストは通常、死亡診断書の番号が分かっている者、番号は分からないが死亡日が分かっている者、そして死亡日が分からない者にグループ分けされている。

重大な病気も追加記録を得ることによって確認され詳述できる。腫瘍登録によってカバーされた地域の住民のがんに関しては、その登録によりしばしば対象者もしくは近親者により報告されたがんを確認することができる。入院を要するような他の状態に関しては、患者からの許可を得て記録を病院に請求することが可能である。合衆国のいくつかの病院では60日以上経過したold書式については許可を与えないので、タイミングが重要である。わが国では、きちんとした書面で研究目的と研究者の所属が明確であれば、たいがいカルテが残っている限りは、協力してもらえる。病院のカルテの維持管理料もままならぬ時勢なので、カルテの保存期限(5年)が過ぎると破棄する病院が多い。時にマイクロフィルム化して保存してある病院もある。

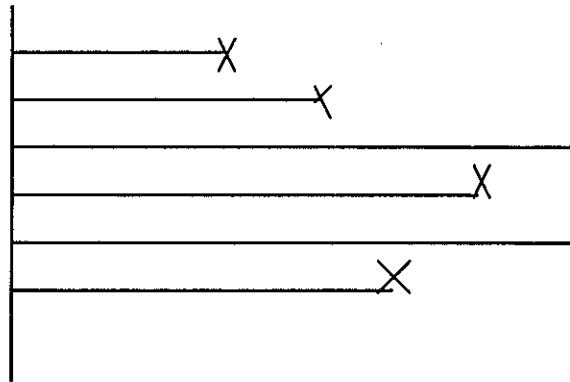
## 1.6、センサードと不適切例の取り扱い

観察期間が終わっても事象を迎えていない例や、観察期間中に事象を迎えずに行方不明や研究目的以外の死因の例をどうするかが問題となる。これらは、センサード症例(censored case)やセンサード時間(censored time)と呼ばれる。以下の三種類がある(Lee1992)。いずれにしても、センサードを取り扱うには、事象以外の出来事で観察が終わるまでの期間を考慮に入れて計算する(例：罹患率における人年計算や比例ハザードモデル等)。

### 1) I型センサーリング

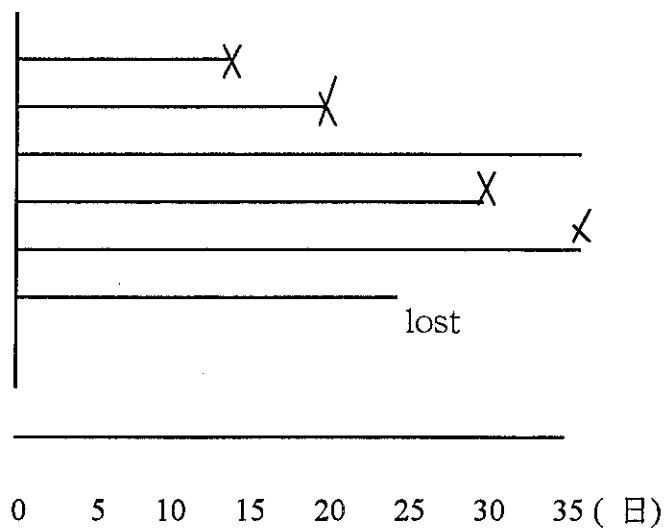
動物実験で一定の期間観察し、生き残ったものは殺す例である。exact or uncensored observations：観察期間内に事象が起こった例、censored observations：観察期間内に事象が起こらなかった例、このタイプのセンサーリングでは、観

察期間中に、目的外の損失(accidental loss)はなく、観察期間の終わりにセンサードした例は、センサード観察が、研究機関の長さに等しくなる。



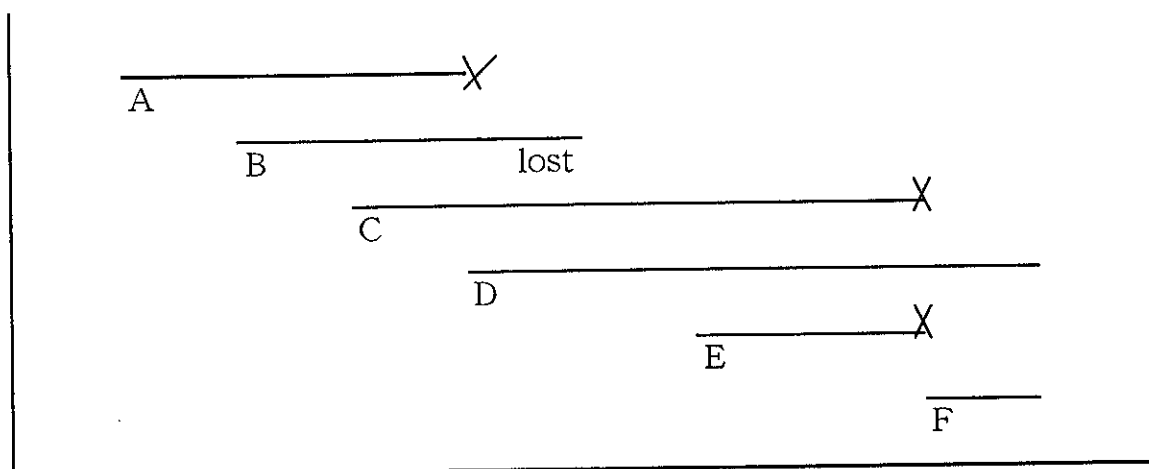
## 2) II型センサーリング

動物実験で動物の決まった割合、例えば100例中80例が死ぬまで観察をやめないこと。動物実験で動物の決まった割合、例えば100例中80例が死ぬまで観察をやめないこと。例の場合では、6匹中、4匹が癌を発生してから実験を終わると決めている。



### 3) III型センサーリング

たいていの臨床研究では、観察期間が一定で、患者は期間内のいろいろな時点で研究対象となってくる。何人かは研究の終了前に死ぬので、正確な生存時間が分かっている。その他は終了前にいなく(withdraw)なり、見失う(例ではB)。また、終了時まで生き残っている人もいる(例ではDとF)。“LOST”患者には、最後に接触したまでを生存時間に、生きている人は、対象となったときから終了前の最後の接触までを生存時間にする。次の例は白血病の緩解期間を表している。



通常、生存分布に使われるのは、指数分布、ワイブル分布、ログノーマル分布とガンマ分布である。もしモデルを用いることが時間や金がかかったり、いいモデルがない時はノンパラが良いであろう。

## 17、介入研究で用いる主な分析方法

### 1) 介入の描出

介入後の観察開始を可視的に描出する方法には、Kaplan-Meier法(積極限法)とActuarial法(保険統計士式)がある。詳しい描出法の解説はここでは省略するが、前者は、事象が起こるたびに生存確率を計算する方法で、後者は一定期間毎に生存確率を計算する方法である。両者共に、数値としても表されるが、同時に



しばしば折れ線様のグラフとしても表される。同じ方法で描いた、非介入群のグラフと介入群のグラフを並べて標記すると、介入による影響が非常に明瞭に見ることができる。ただし、すでに述べたように、ランダムに割り付けた介入研究でも、交絡バイアスから自由であるとは限らないので、その時は交絡要因を調整する必要がある、後に述べるように、主に、層別分析(制限・マッチングを含む)と数理モデルを利用した方法(多変量解析)が用いられることとなる。

## 2) 介入研究におけるマッチング

介入研究は、要するにランダム化したコホート研究と考えられるので、ランダム化もしくはマッチングに関連したコホート研究と観察によるコホート研究に関して論じたRothman & Greenlandの議論(1998)を紹介しておく。これらの議論を、知っておくことは、ランダム化したりマッチングしたりして研究する上で、重要な参考になると考えられる

### (1)コホート研究におけるマッチング

一定の比で曝露対象者に非曝露者をマッチングすることは、粗リスク差と粗リスク比の交絡を防ぎ得る。なぜなら、追跡開始時点でのそのようなマッチングは研究対象者間の曝露とマッチング要因の関連を防ぐからである。このような利点にも関わらず、マッチされたコホート研究は珍しい。その主の理由はたぶん、大きなコホートのマッチングは非常に高くつくからであろう。コホート研究は通常、症例対照研究より多くの対象者を必要とし、マッチングは通常、時間のかかる過程である。このコスト効率の悪さを改善するために、事象が起こったマッチドセットの、マッチされていない交絡要因についてのデータだけを集めることが行われることが可能であるが、この方法が実際に行われることはまれである。

マッチされたコホート研究がまれな理由は、コホートのマッチングは必ずしもマッチング要因の調整のための必要性が無くなることがない可能性があるからである。もし曝露とマッチング要因が疾病のリスクもしくはセンサーリング(競合リスクや追跡不能)に影響するのなら、まっちゃんぶによって作られた元々の

バランスが、分析のために利用する人と人年に拡張できないであろう。これはマッチングが曝露マッチング要因の関連が、元々の人数のカウント数のみを防ぐことになるからである。つまり、曝露の影響とマッチング要因の影響が、曝露と残された人と観察された人年との関連を作る可能性があるということである。もし純粋にカウントデータとリスクのみが検証され、センサリングが生じないとしても、マッチングに使われる何らかのリスクファクターの調整は、リスク差とリスク比の間の妥当な標準偏差の推定値を得るのに必要になるであろう。対照的に、症例対照研究は、(リスクよりはむしろ)曝露と関連したマッチング要因の調整を必要とする。

## (2) コホート研究におけるマッチングと効率

マッチングは、しばしば、影響の指標の標準偏差を減少させることによってコホート研究における統計学的効率を改善することができるが、そのような利益は、もし曝露がランダム化されていないときには保証されない。この点において、コホートの観察研究は、ランダム化試験より症例対照研究に似ている。

観察コホート研究とランダム化コホート研究の間の相違を理解するために、それぞれのデザインにおけるマッチングの計画案を比較しよう。ランダム化試験では、マッチングはしばしばブロック化と呼ばれる。そして治療割付のための計画案である。ペアのブロック化においてはマッチング要因(ブロック化)に関する同じ値を持つ対象者のペアがランダム化され、1つは研究治療に、もう1つは対照治療に割り当てられる。そのような計画案は、例外もあり得るが、ほとんどいつも対応するブロック化されないデザインと比較して、統計的により正確な(効率の良い)影響の推定値が得られる。

非実験的コホート研究においては、マッチング要因は、治療の割り当てよりも対象者選択に関する計画案の一つであると言える。たぶんほとんどの通常のコホートマッチング計画において、マッチング要因分布が曝露コホートにおける分布と同じになるように非曝露対象者が選ばれる。この計画案は、個人もしくは頻度マッチングによってとり行われる可能性がある。例えば、調査者が追跡

のために曝露コホートを同定することを想定しよう。そして、このコホートの、年齢と性の分布が記録される。この時、それぞれの年齢-性層の中では、等しい数の非曝露対象者が追跡されるために選択されるだろう。

観察的なコホート研究のマッチングは単純なものかもしれないが、その効率に関する意味合いは単純なものではない。ランダム化された実験理論からの昔からの議論は、リスク要因のマッチされたランダム化(ブロック化)は、研究結果が連続量の時に、影響の推定値の精度を改善するであろうことを意味している。つまり、影響は平均の差として測定され、結果のランダムなバラツキは、結果に独立した誤差項を付け加えることにより表されることができ。しかし、これらの議論は疫学のコホート研究に繰り越すことはできない。なぜならマッチする選択は、全体の研究コホートの共編量の分布を変えるからである。一方、マッチされたランダム化ではこれは起こらない。結果が離散値の時も、古典的な議論は崩壊する。なぜなら、この場合、それぞれの曝露水準内での結果の平均(期待)値に、結果のバラツキが依存するからである。従って、観察コホート研究において、例えば(症例対照研究と違って)、それがバイアスを招来しないとしても、マッチングは、時には効率を害する可能性がある。

### 3) 層別分析

交絡要因になっていると考えられる要因に従って、層に分割し、それぞれの層で分析を行う方法である。それぞれの層の結果を、重み付け平均等で結合することもあり得る。マッチングはそれぞれの層が1対1もしくは1対2以上の層からなる層別分析と考えればよい。制限は、ある交絡要因を持つ(もしくは持たない)層だけを分析する方法である。この場合、一般化の解釈のためには、可能であればその交絡要因を持たない(もしくは持つ)層についても解析しておくと便利である。層別分析の詳細は、成書を参照されたい(Rothman 1998)。

### 4) 多変量解析

交絡要因となっている可能性のある変数をモデルに組み込んだ多変量解析を行い、その変数を調整した曝露による影響を推定する方法である。通常、ロジス

ティック回帰分析やコックスの比例ハザードモデルがよく用いられる。具体的な方法論は、成書を参照されたい(浜島1993)。

## 18、研究プロトコールの重要性

疫学研究は、多くの人を巻き込んで行われることが多いこと、再調査が困難な場合があること、コスト・時間がかかること、などの理由で、事前に研究プロトコールを入念に作成し、複数の人に事前に検討してもらっておくことが重要である。とりわけ介入研究においては、historicalなデータを利用する観察研究とは違い、参加者に対して、事前に研究参加のためのインフォームドコンセントを行わなければならない。インフォームドコンセントを行うためには、参加者に研究の概要を説明しなければならない。また協力者に対して研究の全体像を説明しなければならない。くわえて、比較的多くかかる研究費用を獲得するために詳細な研究プロトコールを事前に作成しなければならない。しかも、様々なレベルの読者に向けて、異なったプロトコールを作成しなければならない。

プロトコールには、研究の目的、これまでに行われた関連した研究のレビュー、対象者の選定、インフォームドコンセントの行い方、脱落者・拒否者の取り扱い、試行期間の行い方、等が詳細に書かれていなければならない。また曝露の割り当て・ランダム化の仕方も必要である。比較群にプラセボを使うのかこれまでの治療法を使うのか、何種類の曝露をどのように組み合わせるのかも必要である。

分析は、intention-to-treat分析かそれとも別の分析を使うのか、つまり脱落者等をどのように取り扱うのかについて、あらかじめ決めておかなければならない。また、研究をいつうち切るのかについても明確に決めておかなければならない。なぜなら既に述べたように、曝露群と比較群の間にあまりにも明瞭な差が、研究期間中に生じた時に、それを続けることは倫理的ではないからである。