特集

化学療法剤の臨床評価

Clinical Evaluation of Chemotherapeutica

砂 原 茂 一*
SIGEICHI SUNAHARA, M.D.

化学療法と臨床試験

薬剤の効果と毒性を最終的に評価するためには臨床試験によらなくてはならない。いわゆる controlled clinical trial —— 対照試験である。これは British Research Council が1947年SM(ストレプトマイシン)の効果と安静療法に比較したのにはじまるけれども日本では1957年私達が同じく結核化学療法の研究にこれを初めて用いた。

イギリスやアメリカでは治療報告の70%は対照試験の形をとっているが日本ではわずかに10%である。そして各種薬剤のうち化学療法剤(抗結核剤をのぞく)は対照試験の行なわれることの少ない分野である。それは1つには試験管内実験及び動物試験で治療効果がきわめて明確にとらえられると思われているからであり、2つには起因菌や臨床症状を追及することによって比較的簡単に薬効の評価が可能であると考えているかららしい。

果たしてそのように考えていいものであろうか。

thiocarlide はマウスやモルモットの実験結核症に著明な効果があることは疑えないのだが、私達の臨床試験の結果はその臨床的有用性を完全に否定した。

ウサギにおける実験的結核空洞の成立をimuranが美事に抑制した(小川・山村)。ところがヒトの肺結核空洞の治療をImmanによつて促進しようとする試みは失敗に帰した(砂原ら)。

ブドウ球菌の場合にしてもSmith株のような特定の菌株を用いて動物の皮下膿瘍を起こした場合

と、たとえばヒトにおける肺のブドウ球菌感染症を同一視することはできないのであろう。

試験管内実験や動物実験は臨床効果の評価のための不可欠の前提ではあるが十分な条件ではない。

それなら個々の臨床家が自分の直接取扱った少数、あるいは多数の患者の治療経験にもとづいて薬剤の臨床的有用性を決定することは正しいこととして許されるであろうか。

ethambutolによる肺結核再治療成績の報告を手もとの雑誌類の中から手あたり次第引抜いてみると有効率は0%から91%の間にバラつく。これでは、多数の抗結核剤の中にどのようにこの薬剤を位置づけていいのかわからない。

INHにはいくつかの誘導体があるが、1966年に私が調べたところでは北海道と九州が誘導体の使用がきわめて多かつた。これらの地方にINH誘導体の方がINHそのものよりも有効な患者が多数に存在するということはきわめて考えにくいことである(私達の対照試験ではINHとINH誘導体の間に効果の差はない)。主観的な臨床評価法によつて引きおこされたかたよりとしか考えようがない。

化学療法剤の場合、試験管内抗菌力や感染動物の治療試験だけでもとづいて効果を判断するのは危険なのである。

化学療法にも正しく設計された臨床試験が不可欠なのである。

臨床試験のそなえなくてはならぬ条件

臨床試験のもたなくてはならない条件は

1. 比較
2. 同時性

* 国立療養所東京病院(清瀬市竹丘3-1-1)
Director of Tokyo National Chest Hospital
[Jan. 11, 1972 受付]
第179回順天堂医学会学術集会(Nov. 20, 1971)

3. 無作為割りつけ

4. 客観性

である。

まず2つの群（または、それ以上の数の群）の間で比較を行なう形をとらなくてはならない。治療をしないもの、あるいは現在標準的に行なわれている治療を施したものと比較するのではなければ新しい治療（この場合は新薬）の値打ちを定めることができない。結核の場合なども今日では化学療法の影響下にない患者の経過が多くの臨床家によつて忘れられているけれども、SMが導入された時代、安静のみと、SM治療とを比較した成績をふりかえつて見てはじめてSMの懸け値のない実力が判定されるのである。すなわち安静だけでも結核菌が陰性化したものが少なくなかつたのである。したがつて対照群との比較がどうしても必要となる。

その比較は同時的ではなくてはならない。古い病歴をもち出していわゆる historic control としてはならない。流行病の場合はその流行毎に病気の重さの異なることはよく知られているが、肺結核のような慢性感染色の場合でも年代によつて病気の重さは同じでない。私達が対照試験をはじめた1950年代はNTA分類の高度進展型、ことに学研F型という化学療法にきわめて反応しにくい病型のものが大変多かつたが、1970年代に近づくとも高度進展が少なくなりF型はほとんどなくなつたのである。つまり治りやすくなつたから今日の結核化学療法の臨床効果が昔の症例よりも大きくてもそれは薬が強力になつたからではなくて客種がよくなつたのである。したがつて比較は同時的ではなくてはならない。

大変大切なことは比較される2つの群の分け方が無作為であることである。臨床試験は物理や化学の実験とはちがつて均一な材料を用いて実験することはできない。来るものは拒まずで一応すべての患者を受けいれねばならない。しかし実験感度をますために層別することはしばしば行なわれる。それにしても対象のバラつきを全く克服し、地均しすることはできない。私達にできることは

比較する2つの群のバラツキの程度をひとしくすること、つまり群全体として comparable な2群をつくることである。対象のバラツキとしては病気の重さ、病気の型のバラツキもあるし、定量的にはなかなか把えることのむづかしい病気に対する抵抗力のバラツキがあり、さらに薬剤代謝のパターンならびに速度の個体的バラツキがある。代謝についていえば、たとえばINHには迅速代謝型、中間型、遅延代謝型の3型があり、私達の研究によるとこれは2つの対立遺伝子によつて支配されており、一卵性双生児ではINH投与後の血中濃度は符節を合するが如くに一致する。一方rifampicinにも迅速代謝型と遅延代謝型があり、同一人を何回も測定すると高い血中濃度を示すものはつねに高く、低いものはつねに低い。しかしこの場合はINHの場合のような明確な遺伝法則を設定することはできない。身長や血圧とひとしく複数の遺伝子の組合せによる支配の結果であらう。

このような個体差の組合わせはきわめて複雑であるし、なかにはまだ私達が知ることができないが、化学療法の効果につい影響を与える因子も存在するはずである。したがつて既知、未知の因子をふくめてバラツキの程度をひとしくする、comparable な2群をつくり上げるには確率論にしたがうしかない。つまり無作為割りつけである。

したがつて臨床評価には無作為割りつけをした2群の一方を対照群、一方を試験群として前向きに同時比較をすることが不可欠の条件である。

なおINHの場合はその代謝に個体差があるばかりでなく私達の研究によると著しい人種差があり、アジア地域では北から南へ、アイヌからインド人まで、遅延遺伝子頻度が次第にふえていくのである。アイヌ人はエスキモーについて迅速代謝型が多いが、南に下るにしたがつて次第にヘリインドに至ると白人、黒人なみとなるのである。

なお臨床効果の評価にさいしては客観性を保つために特別の努力がなされねばならない。痛みなどは偽薬でも30%は治癒するのである。主観性を

排除する仕組みとしては二重盲検法がある。化学療法の場合は客観性の高い評価基準がある場合が多いから必ずしも二重盲検を必要としないが、主観的な症状の軽減などを目安とするときは二重盲検が必要となるし、副作用の客観的な評価にもどうしても必要である。

臨床試験の問題点 2～3

主として、私達の10数年手がけてきた結核化学療法の対照試験から引き出すことのできた問題点の2～3を検討することにする。他の感染症の化学療法の場合にも多少は参考になるであろう。

1. 評価の物尺

肺結核症に対する化学療法の効果はいろいろな物尺を用いて測ることができる。X線写真の変化、菌所見の推移、赤沈や一般症状の経過あるいはこれらの組合せなどである。このうち再現性の最も高いのは喀痰中の結核菌の培養である。たとえば200例の肺結核患者をあつめこれを無作為に2つに分割し、両群に同じ化学療法を行なうと6カ月間の培養所見の経過はほとんど完全に一致する。この場合空洞の推移を物差とするとそれほどよくは一致しない。

X線所見はあとまで保存することができるし、客観性が高いように思われるけれども再現性はそれほど高くないのである。pattern認識だからやむをえないことも知れないが、熟練したX線読影の専門家をあつめてもX線写真の質的な読みは相当大幅にバラつくのである。体温、赤沈、体重などは一層不確かな物差である。

肺結核症は結核菌によつて引き起こされる病気であるから、その原因である結核菌の消長を物差とするのが合理的であることは理解しやすい。臨床評価の物尺としては、

1. 病気の本質に密着していること
2. 測定に安定性があること
3. 定量的であること
4. 感度の高いもの

をえらぶ必要がある。化学療法の場合は病原菌の消長がもつとも信頼性が高いのである。

たとえば結核菌の場合菌所見だけではなく、X線所見(空洞影と浸潤影)、臨床症状など多くの因子を組合わせた総合指数 (composite index) を作り上げると病気の姿を多角的に表現できるから治療効果の物尺としてもつとも適しているように考えられがちだが、そのようにするとかえつて能率が悪いのである。菌所見の物尺としての鋭敏な感度が他の因子に水増しされて切れ味が悪くなるのである。

2. 層別の意味

同じ細菌によつて引き起こされる感染症であつても宿主側と細菌側との複雑な条件の組み合わせによつて文字通り千差万別である。肺結核症の場合を例にとると排出する菌量には Gaffky 10号、8号というようなものから1、2集落の培養が時々見出される程度のものもあるし、たしかに結核症にはちがいが無いがどうしても結核菌の見つからない場合もある。X線像もいろいろであるから、ある程度まとめて取扱いうるようには各種の分類法が提案されている。広く行なわれているのはNTA分類、学会分類、学研分類などである。NTA分類の高度進展は化学療法に反応しにくい、学研分類は化学療法への反応を目標に工夫された分類だから特に利用度が高い。基本型のC型、空洞のKx、Ky、Kz型などは化学療法の影響を受けにくい型であり、もつとも頑固に化学療法に抵抗するのはF型である。さらに薬剤、たとえばINHの代謝の個体差も化学療法の効果に無関係ではない。このようなすでに私達に分かつている因子のほかには特異的あるいは非特異的な抵抗性の違いの存在も考慮に入れなくてはならない。

このような層別因子——背景因子は薬効評価にさいし重要な意味をもっている。

第1に無作為に分割した2つの群——治療群と対照群が果して予期通り comparable であるかどうかの検定に用いられる。割りつけが厳格であつて、標本数が十分多ければ、化学療法の効果につよく相関する因子について comparable な群ができ上つているはずである。このことを実さいに確かめてから臨床試験を出発させなくてはならない。

第2に層別は実験感度を高めるために必要である。たとえばNTA分類の軽度進展のような軽い、治りやすい（化学療法を用いなくても安静療法だけで治るような）層を用いると化学療法剤の実力を評価できない。INH+Sulfa剤というような弱い治療方式でも100%近く結核菌が陰性化してしまうからSM+INH+PASのような強力な治療方式との間の差がとらえられない。つまり臨床試験の感度が低くなるのである。肺結核の化学療法の場合であれば、むしろ高度進展というような治りにくい層を用いる方が感度の高い実験になるのである。それぞれの化学療法剤について対照（偽薬又は標準薬）との間の差が出きただけハッキリととえられるような層を探りあてて実験計画を組むことが必要なのである。そうでないと効果の小さい化学療法剤を不当に高く評価することになったり逆に強力な化学療法剤の効果を見のがすことになるのである。

このように層別化した群について臨床試験を行なうことが大切なことではあるが心しなくてはならないことは、ある特定の層について得られた結論を不用意に拡大してはならないということである。たとえば肺結核の軽度進展という層について行なった試験の結論を、化学療法に反応しにくい高度進展をふくむ肺結核一般に及ぼして、その化学療法剤がすべての肺結核症について高い治癒率を期待できるかの如き主張をしてはならないのである。

臨床試験を行なう場合はそれから引き出される結論の及ぶ範囲、及ばない範囲をはつきりと区別してかからなくてはならない。

3. 脱落の問題

ことに慢性の感染症で長期の化学療法を必要とする場合には脱落の防止と脱落例の取扱いが臨床評価の重要なポイントとなる。

一年なら一年、化学療法剤を経験し得た症例だけについて奏効率を算出することがしばしば行なわれるが、一般的にいえばその化学療法剤に反応しにくいあるいは副作用のはなはだしい症例が脱落し、一年後に残っているのはほとんど優等性ば

かりである。2つの治療方式を比較する場合など、脱落は劣った治療方式の方に有利に働くのである。したがって最もきびしい脱落処理の仕方では試験終了後の成功率は化学療法剤Aでは90%、Bでは60%であつたとすると、優れた成績を示したA群からの脱落はすべて失敗例、劣った成績を示したB群からの脱落はすべて成功例であつたと仮定してもAがBよりも優れた結果になったときだけAがBに優ると判定するやり方である。

もちろん、もつとも大切なことは脱落率をできるだけ小さくするような実験計量をあらかじめ組むことである。1月位の治療試験なら脱落を5%以内位に食いとめたい。又比較する2群の間で脱落理由にかたよりのないことが望ましいし、又脱落直前の時点での臨床状況にもかたよりのないかどうかを確かめておくべきである。たとえば肺結核の化学療法の場合A群からの脱落者は大部分菌が陰性化した症例であるのにB群からの脱落は大部分菌陽性例であつたというような場合はB群は脱落によつて不当に有利な影響を受けていることになる。このような点を正しく吟味して全体的な比較評価が行なわれねばならないのである。

したがって臨床試験の報告には必ず脱落についての検討の項目がふくまれていなくてはならないのである。

4. 服用方法の臨床試験

ふつう臨床試験はある薬剤の一般的な評価を定めるために行なわれる。しかしその評価は層別の仕方に大きく左右されることはすでにのべた通りである。それだけではなく同じ化学療法剤であっても一日量、一回量、服用間隔などによつても治療効果が影響をうける。じつさいには臨床試験第2期までの瀬踏み試験（pilot trial）で用量や用法についての一応の見当をつけておいて、その情報の上に立つて第3期試験——controlled trialが行なわれるものである。しかし適正な用量、用法を定めるためにはその目的で独立した対照試験を行なうことが必要になつてくる場合がある。

肺結核の化学療法の中核の位置を占めているINHについては、私達は4mg/kgから18mg/kg

にわたる1日量のちがいが治療効果にどのような影響を及ぼすかという試験、同じ1日量を早朝1回に服用した場合と2回以上に分服した場合の比較、INHを毎日投与した場合と週2回投与した場合（この場合は普通量を用いる方式と大量用いる方式を含む）の比較などをcontrolled trialの形で行なっている。その結論は 1)毎日投与の場合INHを増量しても臨床効果は増さない（日本人にはINHの迅速代謝型が多いことを考えるとこの結論は重要である）。2)INHは1日量を分服するよりも早朝1回投与の方が効果が大きい。3)普通量の週2回法は毎日法に劣るが大量の間歇

法は普通量毎日法にひとしい。

一般の化学療法の場合はこのような厳密な、細かい対照試験が行なわれていないので平均的な投与法による対照試験を1回行なって、あとはその薬品の pharmacokinetics から推理をして服用量、服用方法の違いによる臨床効果の動きを推定しているにすぎない。多くの場合はそれでも一応間にあうわけだけれども、重要な薬剤あるいは問題の多い薬剤の場合はINHの場合のように「使用法の臨床試験」を行なうことが望ましいのである。