

(1)安全性 / 有効性 司会：谷田憲俊（兵庫医科大学第4内科）

浜：今日の全体のセッションの説明をいたします。1日目の昨日は、まずは反省ということで、薬害の歴史を振り返り、午後の薬剤疫学の課題で日本ではどのようなことに取り組んでいけばよいのかを議論いたしました。そしてヒックスさんの講演で、どのようなことが外国では議論され、進行しているのか、今後どのような方向で世界は動いていくのかについてお話がありました。

今日2日目は、いよいよそれぞれの医薬品が良いものか悪いものかを検討、学んでいただくためのガイドになるようなこととお話したいと思います。明日は専門的なセッションを用意していますが、今日は医師・薬剤師だけでなく一般の方々もかなり来ておられますので、よく分かるように、相当かみ砕いてお話申し上げます。従って、医師・薬剤師の方々には少し物足りないことがあるかもしれませんが、そういう方々は是非、明日のセッションにも参加していただいてもっと詳しく検討する方法を知っていただくとよいのではないかと思います。

まず、Gでは安全性と有効性をどういうふうに評価したらよいのかをお話します。次にHは、個々の薬剤を実際に評価してみたら、これは良いとは言えない、あるいは悪い、といった実例をお話します。午後のIでは、日本では新薬として登場したが、世界的に見ると実はそうではない、それが非常に高い値段で売られているということが最近分かってきましたが、そういう経済的な問題に焦点を当てて、薬の良い悪いに取り組んでいきたい。

次にJ、実は分科会の一つに予定していましたが、非常に参加のご希望が多いので全体会議にいたしました、患者への情報提供の今後の課題です。次のKは個々の分科会という

ことで参加者のみなさまはそれぞれに別れていただきます。最初に分科会の希望を聞きましたが、それほど偏りはありませんので、それにとらわれることなく好きな分科会に参加して下さって結構です。

遠慮なくご質問いただけたらと思います。質問の時間がないときは、昼休みとかにでも講師を捕まえて、個人的にどんどん質問していただいて結構です。よろしくお願ひいたします。

谷田：おはようございます。兵庫医大の谷田と申します。このGの最初のタイトルは、よい薬、悪い薬の見分け方ですが、考えてみれば変ですよ。薬によいものと悪いものがあるなんて。しかし、残念ながらそれを話題にしなくてはならないというのが、報道や昨日のセッションでお分かりのことと思います。Gは、かなり専門的なこととなります。なるべく講師の方々は、かみ砕いてお話しただくようお願いいたします。まず、最初に、「よい臨床試験、無作為化比較試験とは」、東京医科歯科大学の津谷さん、よろしくお願ひいたします。

よい臨床試験とは つたにきいちろう 津谷喜一郎



東京医科歯科大学、研究者

もともと漢方薬や鍼灸など東洋医学の臨床評価の方法論に関心があり、日本での臨床試験の開拓者の一人である佐久間昭氏に学んだ。「二重盲検法ではなくその基本にある無作為化が大事なんだ」と佐久間氏に言われてから20年になる。臨床試験の分野というのは当時も今も人不足で、東洋医学だけでなく近代医学の評価、さらにはそれらのシステムティック・レビューを行うコクラン共同計画にまで活動が広がった。

クエスチョンにどう答えるか？

図-1はノルウェーで93年にあったカンファレンスの表紙です。But will it work, Doctor? (先生、それは本当に効くんですか?)とあります。これは医者の立場に立ってみると、正しくきちんと答えるのはなかなかむずかしい質問です。「嫌な」質問だと思えます。

臨床の場で使える実際の情報というのは図-2のようなモザイクのパターンになっています。これは漢方薬を中心に描いたものですが、西洋薬も同じです。情報が混乱している。そういった状況で、先程の「それが効くんですか?」と

いう問いに対して答える一番強力なものが臨床試験です。つまり、リサーチクエスチョンに対して答えを与えるのが臨床試験。その中で基本的なものが、ランダム化比較試験あるいは無作為化比較試験と呼ばれるものです。

臨床試験とは

臨床試験 (clinical trial: CT) とは何か。4つの要素から成り立っています。第1に「評価を目的に」、治療ではないのです。第2に「ヒトを用いて」行う、動物に対する臨床試験もありますが、基本的にはヒトを対象としたものです。第3に



図-1 1993年ノルウェーでのカンファレンスの表紙

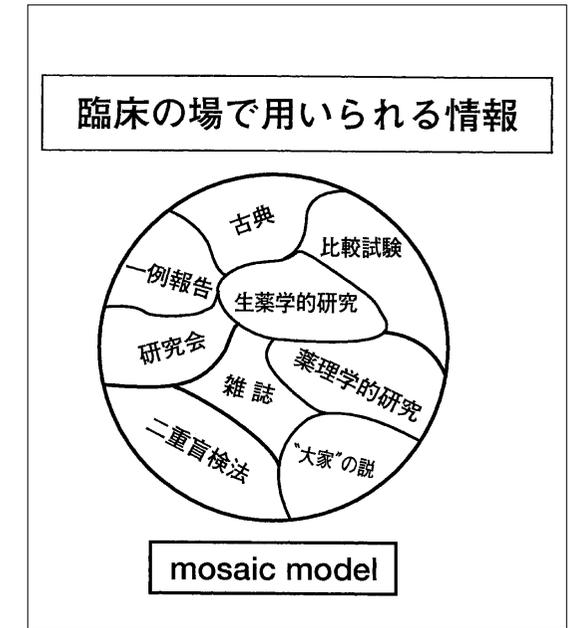


図-2 臨床の場で用いられる情報

医薬品評価の方法 その(1) (G・H)

「意図的に開始される」、ズルズルと始めるのではなくある意図をもって始めるのです。第4に「科学的実験」です。

例えば、中国からお土産で漢方薬を買ったからちょっと使ってみましょう、効いたからつぎの患者さんにも使ってみましょう、というのは臨床試験ではありません。科学的実験であるためには、計画書、プロトコルがないといけません。

臨床試験に近い言葉に、「人体実験」や「生体実験」があります。非常にネガティブなイメージがありますが、基本的には臨床試験と同じです。ただ、その目的とするところが、寒冷実験や感染実験などの残虐な実験など、社会的に誤った、あるいは意味のない臨床試験である場合に、これらの用語は使われることが多いものです。

また、「治験」という言葉があります。これは法的には臨床試験の内で厚生省に承認申請することを目的としたもの、行政的な定義です。しかし、治験のことを「治療経験」だと思っている医師もいますね。昔の論文を読むと「
の10人に使った治験」というタイトルが多くみつかります。今でもそういう使い方をする人がいます。しかし、意図的に評価を目的として開始した実験ではないですから、今という臨床試験ではない。さらに「治療実験」の略と思っている人もいます。これは臨床試験に近いですが、臨床試験は治療だけでなく、予防、リハビリ、健診など広い分野で行われます。

各1つの臨床試験には、それぞれ薬物などの有効性や安全性を見ようという目的がありますが、臨床試験すべてに関する第1のゴールというのはバイアス〔系統的なエラー〕を小さくすること。これがゼロにできれば一番いいわけですが、小さくすること。第2のゴールはバラツキ〔ランダムなエラー〕を小さくすることです。

ときどき単に薬の投与の前後で比較すればいいのではないと言われる。基本的にそれで悪くはありません。しかし、そこにはいくつかの問題点があります。第1に、多くの病気は患者間

でバラつく。第2に、多くの患者は自然に治る。第3に、1人の患者の中でよくなったり悪くなったり動きます。第4に、平均への回帰。もともと患者さんというのは悪いときに医療機関に来ますから、2回目というのは良くなっていることが多い。これを平均への回帰と言います。第5に、食事、仕事、ストレスなど他の要因で動く。第6に、前に行った治療が持ち越しされる、キャリア・オーバー効果と言います。こうした問題があるということで、やはりコントロール(対照)を置いた臨床試験が必要であることとなります。

RCTとは

ときどきランダムマイズド・コントロール・トライアル(RCT)の構造を理解しないで、日本語で無作為抽出試験とか無作為化制御試験とか訳している人がいます。メディアでもそういうふうに使っている人がいます。図-3を用いて説明しましょう。

例えば、ある病気を持った患者が100万人いるとします。おじさんもいればおばさんもいる、男の子も女の子も。こういった背景因子がそのまま反映されるように1000人選び出す。これが無作為抽出(random sampling)です。選ばれた人はランダムサンプルになります。ここでサンプルは被験者になります。選挙に「戸口調査」というのがありますね。投票の終わった人をつ

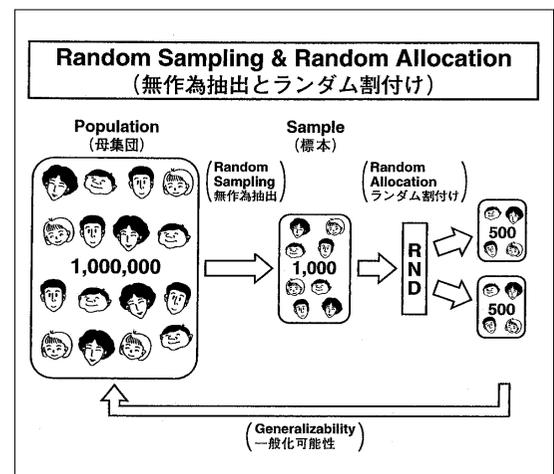


図-3 無作為抽出とランダム割付け

かまえて、誰に入れましたか、何党に入れましたかと聞く。この調査に男子学生を使うと、おばあさんはうるさいからと聞かないで若い女の子にばかり聞く、とするとバイアスが入ります。すると、これは無作為抽出ができていないことになり、予測を間違えることとなります。ただし、臨床試験においてこの無作為抽出は困難なものです。というのは、患者さんに、こちらから病院へ来るようにとは言えないからです。たまたま、その病院に来た患者を使います。そこで、ランダム・サンプルではなく、ハンディ・サンプルというような呼び方をします。

ついで、この1000人を均等な2群に割り振る作業が、ランダム割付け(random allocation, randomization、無作為割付け)です。

先ほど、性と年齢という二つの要素について話しましたが、実際には、病気の重症度だとか合併症、併用療法、さらにまだわれわれが知らない要因もあります。ある薬物は特定の遺伝子を持っている人によく効く、しかしその遺伝子はまだ分からない。こうしたunknown factor(未知の要因)についてもこのランダム割付けで均等に割り振ることができます。

ついで、一方の群に試験薬〔編集部註〕、他方の群にプラシーボ、標準薬、あるいはnon-treatment(無治療)を用います。そして、エンドポイント(評価項目)を決めて比較する。その結果を、もとの人、統計学的には母集団(population)とよびますが、そこに使うのが基本的な構造です。母集団は将来の患者(future patients)ということもできます。ここで、結果を一般化(generalize)できるかどうか問題になります。

無作為化比較試験の無作為は、図-3の右側のランダム割付けの意味です。従ってこれを無作為抽出試験とか無作為制御試験と呼ぶのは間違いです。また「割付け」というプロセスは「作為的」なものですから、以前用いられていた「無作為化」という言葉を避けて最近では「ランダム化比較試験」と呼ぶことが多くなってきました。

臨床試験の倫理

図-3は臨床試験の倫理の説明にも使えます。例えば、診療所の医師とします。そこへ患者さんが来た。その人に臨床試験に入っただき、あるプロトコルに従って試験薬を投与するわけですが、どうも効かないから2倍量を使おうとします。そのときに、それは目の前の患者さんによかれと思ってやることも知れませんが、このように使うと正しい比較ができず、バイアスが入ります。つまり、目の前の患者さんによかれと思ってやっていることも、その薬を使う将来の患者さんにとってはよかれとはならないのです。

ということで、臨床試験においての医師と目の前の患者さんとの倫理関係を「個別的倫理」(individual ethics)、将来の患者さんとの倫理関係を集合的倫理(collective ethics)と言います。

ここにジレンマが生じます。先程は、2倍量にするという有効性に関する例を示しましたが、安全性の方、この試験薬を使うと危ないから半分の量にしようとする、ここにまたバイアスが生じます。

つまり目の前の患者さんの肩ごしに、この臨床試験から得られた情報を使う1000人の患者さんがいる。あるいはベッドに寝ている患者さんの枕の向こうにもっと沢山の患者さんがいる、というのが臨床試験の基本的な構造です。つまり臨床試験は科学的な実験ですから、目の前の患者さんは実験材料なのです。

かといって、通常の医師患者関係の共感や誠意、思いやりとかがいらぬというわけではありません。これはドイツのフーフランド(Christoph Wilhelm Hufeland, 1762-1836)が書いた『医師の心得・態度』(1836)という本があります。そこには「その際、医師は患者を決して手段として見るのではなく、常に目的として見なければなりません。つまり患者を生物実験の単なる対象として、あるいは単なる医療の対象として見るのではなく、患者を人間として、自然そのものとして、最高目的として見なければなら

らないということでもあります。」(杉田絹枝、杉田勇 訳、『医の倫理』、1995)。この原本はドイツ語で、それが隣のオランダ語に訳され、江戸期に日本に伝わり、杉田成卿によって『医戒』(1849)として訳されて、広く日本で読まれたものです。

カントの思想を受けたものですが、医師にとっても目の前の患者を実験材料として使うのは、あまり気持ちのいいものではありません。カントは「人間を手段として使うな」と言っています。臨床試験の場合は、被験者(study sample)、最近では参加者(participant)とも呼びますが、その人々はやはり手段なんですね。これに従えば、臨床試験ができなくなるので、それを「カント的制約」と呼びます。それを解釈するものがインフォームド・コンセントということになるのです。

つまり、臨床試験の場合は、通常の医師患者関係とは少し違う、距離感を持ったものだということ覚えておいてください。

二重盲検法とは

「二重盲検法」という言葉が日本では一人歩きし、臨床試験というと「二重盲検法」と思ってしまう人もいます。それは間違いです。臨床試験の中にはコントロールを持つものと持たないものが、まずあります。持ったものは「比較試験」と呼びます。その中で、コントロール群とランダム割付けをしてあるものがある。それがランダム化比較試験です。その中の一部が二重盲検ランダム化比較試験ということなのです。

バイアスを避けるには

先ほど臨床試験の最大のゴールはバイアスを避けることだと言いました。バイアスにはいろいろありますが、メインのものを4つ説明します。

第1の選択バイアス(selection bias)というのは、片一方の群に病気の軽い患者さんが多いと、薬が効いたのかもともと病気が軽いから治ったのか分からない、因果関係の推論が正しくできない、このような背景因子の傾きがないように

するのが、ランダム割付けです。

第2に、観察バイアス(observation bias)を避けるのがブラインディング(blinding, 盲検法)です。ここでプラセボが関係します。医師がどちらが何かを知っていると、ああ、これは知り合いに頼まれた患者さんだからとか、わたしはポニーテールの女性が好きだから実薬を使おう、プラセボを使うのをよそう。こうなると割付けの時点でバイアスが入ってしまいます。また割付けた後も、この患者さんはプラセボあるいは無治療群だから他の治療法もやっておいたほうがいだろう、さらに実薬群だからもう少し詳しく観察したほうがいだろうとか、環境要因が変わってきます。そういったことを避けるのが、ブラインディング。最近ではマスキング(masking)とも言います。それで患者さんが知らない、医者も知らない、これを二重盲検法と言います。

このブラインディングができない場合があります。物理療法だとか、手術だとか。医師は当然何をやっているか知っているわけです。しかし、知ることによるバイアスを減らすことが出来ないかというところがありません。今、わたしは鍼の臨床試験に関わっているのですが、その場合は、鍼を行う人と評価者(assessor)とを分けています。評価者はその患者がどういう治療を受けたかを知らずに評価するのです。患者、鍼灸師、評価者とそれぞれにブラインディングがあります。さらに、統計学的な解析をするときも統計家がどちらか分かると解析手法の選択などでバイアスが入りますから統計家に対してもブラインディングすることもあります。

1948年に世界で最初の臨床試験が行われて来年でちょうど50年でロンドンで、シンポジウムがあります。ストレプトマイシンの肺結核に対する臨床試験ですが、エンドポイントをレントゲン所見としています。プラセボを使ったわけではありません。つまり、レントゲンから患者さんの名前を消している。ストレプトマイシン群と無治療群、どちらがどちらか分からないようにしている。このようにブラインディングと

いうのは、観察にいろいろなバイアスが入るのを避けるためのものです。

第3に、インテション・トゥ・トリート・アナリシス(Intention to treat analysis, ITT 解析)データが集まると解析する段階になります。そのときのバイアスを避けるものです。基本的に、割付けられた患者さん全部を対象としましょう、というものです。例えば来なくなった患者さんは解析できませんからこれをはずすとしましょう。でも、実はこの患者さんは亡くなったのかもしれないし、病気で亡くなったのかもしれないし、副作用で亡くなったということもあります。そういうバイアスを避けるために、すべての患者さんを解析の対象にしましょうというものです。ITT 解析では有効性は真の値からいくらか薄められたものとなります。

実際には、フェイズ1(第1相)、フェイズ2(第2相)のように探究的な段階ではプロトコルに合致したのだけを解析するPPB(パー・プロトコル・ベース)解析を行う場合もあります。基本的にはITT 解析で行った方が、将来の患者さんへ一般化しやすいものです。

第4に、試験の結果のいい論文だけが出版されますと出版バイアス(publication bias)が起こります。つまり総体として情報が偏ります。これを避ける一番いい方法は臨床試験を登録制にして、例えば3年経ったらアンケートを出して、あれはどうなりましたか、と聞くことです。日本では「治験」は厚生省に登録されていますが、その情報にアクセスできないのが問題です。

いずれにしても、いろんな段階でバイアスを避けるようにするわけです。

臨床試験のリソース

臨床試験に何が必要か。第1に患者さんです。患者さんがいないと絶対に臨床試験はできないのです。第2に医師。実際には医師、薬剤師、鍼灸師のこともあれば、心理学者のこともあります。第3に病院、つまり臨床試験を行う施設です。なんでも3つにすると覚えやすいと思いますが、この3つの中に他をはずして製薬会社や行政を入

れることもあります。しかし、消費者側(コンシューマー・サイド)に立てば、先の3つが臨床試験に必要なリソースということになります。

このリソースを使って、「あるリサーチ・クエスチョンにバイアスなく答える」のが「よい臨床試験」ということです。リサーチクエスチョンとは冒頭にお話しした「それ本当に効くのですか?」という疑問のことです。つまり患者さん、医師、施設、これらのリソースを無駄遣いせずに、バイアスのない情報を効率よく生産するのがよい臨床試験なのです。

編集部：医薬ビジランスセンターとしては、国際的に(ICHのガイドラインで)使用されているInvestigational product = 「試験物質」「試験製品」に相当する用語としては、「試験薬」や「治療薬」ではなく「試験物」という用語を用いている。

臨床試験でこれから有効性と安全性を評価しようとしている物質は、まだ「薬」となる資格を持っていない、「薬の候補」の段階だからである。

谷田：あとでディスカッションはあるのですが、言葉の問題とかテクニックなこととかわからないこととか、是非とも聞きたいこととか、一つ二つお受けしたいと思います。

名郷：（医師）見分け方ということで確認なのですが、第1にランダム割付け、第2に二重盲検法、第3にITT解析となっていますが、これは Evidence-Based Medicine を始めたサケット氏らのガイドによれば、第1が解析時の無作為化の維持、インテンション・トゥ・トリートとフォローアップ率が高いかどうかというのが第1のポイントになっていて、二重盲検法は第2のガイドになっていると思います。ITT解析の一番大きな効果というのは、解析時の無作為化をずっと守っている、ということ。つまり割付けた時に無作為化ということと解析時にも無作為化が守られていることがセットになっていて、はじめて無作為割付け臨床試験と認識するのがいいと思うのです。

第1、第2は、ランダム割付けと、インテンション・トゥ・トリート、追跡率が高いということ、そのつぎのポイントが二重盲検としたほうが誤解がないのではないかと思うのですが。

津谷：おっしゃる通りだと思いますが、実はインテンション・トゥ・トリートの定義は非常に混乱しているのです。微妙にずれているので、インテンション・トゥ・トリートという用語は止めましょうという議論が起きています。臨床試験は現場の科学なので、いろんなことが起きています。例えば、前立腺肥大の臨床試験に女性の患者のデータが入ることがあります。データの誤りでしょう。また、割付けられた薬を1回も服用しない患者さんのデータはどうするか、などいろいろあります。そこで、ITT解析を少しゆるめた full analysis set (FAS) 解析という考えが出てきま

した。割り付け時の患者のままを解析するという時にはオール・ランダムイズド・サブジェクトという言い方がより分かりやすいかもしれません。（その後1998年、ICH統一ガイドラインで用語が統一された）

近藤：（医師、放射線科）医薬品に関して欧米（日本以外）で行われた臨床試験での結果を日本で適用することの是非と、その理屈づけをどう考えていらっしゃるか教えてください。

津谷：海外の結果を日本人に外挿できるか、という話ですね。実は昨日お話しした国際ナショナル・カンファランス・オン・ハーモナイゼーション（ICH：International conference on Harmonization）で、もう5年ほど議論していますが、なかなかガイドラインがまとまらない分野です。基本的には、内的要因と外的要因に分かれます。内的要因は遺伝的要素により代謝活性が民族によって違うであろう、などということ。外的要因は、食事や生活習慣、物の考え方とかが違うということ。これをどう解決したらよいか。ブリッジング・スタディというコンセプトがあります。例えばある国臨床試験で作られたデータセットと別の国のデータセットとでブリッジしたらいいのではないか。しかし、このブリッジング・スタディは現時点ではまだ世界に1つもないのです。いろんな医薬品のグループについて、情報が蓄積されていけば、経験的な判断でもって、エビデンスではないですが、きっとこれは日本にもいいだろう、そういう議論が今進んでいる最中です。（これも1998年にガイドラインが出された）

谷田：これで臨床試験とは何を意味するかを少なくとも理解されたと思います。詳しいディスカッションは後ほど行います。次は、動物試験の見方、効果と毒性のバランスについて。浜さんお願いいたします。

動物試験の見方

[効果と毒性のバランス] ———— 浜 六郎



医薬ビジランスセンター JIP 代表、医薬品・治療研究会 (TIP) 副代表、内科医

大阪大学医学部卒業後、同附属病院、大阪府衛生部を経て、1977年から阪南中央病院内科勤務。1997年3月退職し、医と薬の監視機関として、医薬ビジランスセンター JIP を設立。大阪大学医学部公衆衛生学非常勤講師。主な研究分野は、医薬品の安全かつ適切な使用に関する研究、医薬品の評価（系統的レビューを含む）医療経済的研究、薬害の発生要因に関する総合的研究など。著書『クスリへの告発状』（共著）エール出版社、『薬害はなぜなくなるのか』日本評論社、『抗生物質治療ガイドライン』（翻訳監修）医薬ビジランスセンター発行など。

動物実験でも情報に偏りあり

ある物質が薬として使えるかどうかを見分けるためには、最初からヒトで行うわけにはいきません。だから初めは動物で検討する。その検討をどのようにするのかという問題です。しばしば動物実験で良い結果が出たらそれが強調される、特に製薬企業のパンフレットでは動物実験でこんなによい結果が出たということが、あたたかも人間でも良い結果が出たかのように宣伝されます。

毒性については逆に、いくつかやった実験の中で、毒性が目立たないものが選ばれるわけです。一度出た結果が都合が悪いと、別の動物でやってメーカーに都合のよい結果ができればそれをまた公表するというようにして、製薬企業に都合のよい毒性は少ないのだという情報が重点的に流されることになります。

動物実験も適切な危険／益比の評価が重要

私たちは、そのような偏った情報にいつもさらされているわけで、公表された文献情報の中から、本当の情報をどう選ぶのか。効果の面と毒性・危険の面の情報をどういうふうに見分けるかが非常に大事なことです。狙った効果、薬理作用がどの程度にあるか。毒性はどの程度に出るかを検討します。この物質が使えるかどうかは、狙った効果と毒性のバランスで決まってくるはず。英語で言えば、risk-benefit ratio、日本

語では危険／便益比（危険／益比）と訳されています。人でやる前に、まず動物実験をして、効果が大きくて危険性が少ないのが薬の候補となる資格が与えられて、人での臨床試験が実施されることになるわけですね。この段階で危険／益比が大きいものはそもそも、人で試験する「候補」となる資格もない。しかし、薬害で問題になったものを振り返ってみると、この段階を通過して、「候補」となったこと自体が問題だったとしか言いようのないものが結構ありますので、その見分け方を知っておくことは非常に大切なことです。

危険の徴候は毒性試験だけではない ——薬理試験でも危険の徴候あり——

動物実験を見る場合に、人での危険を予知できるものは、ふつうは毒性試験と考えられていますが、実はそれだけではありません。薬理作用も単に効果を見るだけではなく、ヒトにとって危険な兆候と読みとれるものもあります。

たとえば、胃潰瘍に使う薬を狙っているとしても、神経に対してどういう作用があるか、心臓の機能、たとえば心拍は増加するのか、心筋の収縮力はどうか、血管は収縮するのか拡張するのか、血圧は上がるのか下がるのか、尿は出るようになるのか、少なくなるか、解熱作用はあるかないか、気管支の筋肉は弛緩するのか緊張するのか、眠気はどうかとか、いろんな生理