

## 臨床試験での安全性評価の問題点

べっぴひろくに  
別府 宏園

医薬品・治療研究会 (TIP) 代表、  
神経内科医、都立府中療育センター副  
院長

### 日本の臨床試験の問題点 最初から疑わしきを除外

少し時間が迫っているようなので簡単に要約いたします。副作用と一般に言っているもの以外にその薬を使用している最中に起こることというのが、いろいろあるわけですし、与えられた本人にとってあまり好ましくない不都合な出来事であれば、一応は薬と関係があるかもしれないと考えてみる必要があります。昨日からの議論にも出てきましたが、ひとまずは有害事象として捉え、その上で因果関係があるかないかということのを別の角度から考えることが重要なのだらうと思います。ところが、初めから疑わしきは棄却するという姿勢で、有害事象を処理してしまうところに日本の臨床試験の問題がある。

日本の試験論文で見ますと、検査値異常を全く副作用からは除外して別の項目で扱っています。ソリブジンでもそうでしたが、検査値異常というのは副作用とは別のもので取り扱っている。異常があれば、それが何%あったと書くけれども、目につきにくい小さな文字で書かれている。説明書の中でも分離して記載してある。

しかし、実際は副作用にはいろいろあるわけですし、当然その中に白血球が減ったとか、GOT・GPTが高くなったという肝機能検査が悪くなったのも副作用の結果として考えなければならぬのに、分離されて書かれているために目を引きにくい。

その他、行動異常とか患者の気分が変化したとかいうのも、副作用としては見逃される可能性がある。あるいは故意に省かれていることもある。

### 副作用の頻度は臨床試験だけでは無理

副作用といったときに、われわれ医師が気にするのは、その頻度と臨床的な危険度です。多少危険性があってもそれが臨床的に非常に有効であれば薬としての価値は残る可能性があるわけですから。あくまで危険性というのはその利益との比較で見なければなりません。代替治療がどうかかあるいは経済的に見てどうなのかということも危険性の判断と常にカップルして考えなければいけません。

また副作用があったとしても、継続していれば消えて行く物もあるし残る物もある。中止してあまり変わらないということもある。メーカーが持って来るパンフレットには副作用は10%以下でしたとか全部ひっくるめてあったり、5%でしたとか言いますが、現実にはそれぞれ分けて考慮しなければいけません。

ご存じの通り、比較的多い副作用を見つけるのは、少数の患者さんを対象にしたもので分かるわけです。例えば副作用の頻度が100分の1程度であれば300人の患者さんを診るだけでほぼ確実に掴むことができます。しかし1000分の1になると3千人、1万分の1になると3万人、5万分の1になると15万人という人を捕まえないといけないとなる。一般に臨床試験のデータの中に薬の有効性と安全性の両方が示されていると思っ込んで入る人が多い。有効性という面からみれば、100人に1人しか効かない薬はまず考えに入れられないわけですから、有効性はかなり少人数の臨床試験でも証明できる。

一般的に臨床試験というのはもっぱら有効性のほうに目が向いて設計されている。有効性を検証するには非常に優れているけれども、副作用に関してはもともと検出しにくい仕組みに

なっているということを理解しておかなくてはならない。臨床試験論文の中で安全性が高いと書いてあってもそれは非常に危ういものだと思います。それを補うものとして市販後調査が大事だということになります。

### 大切な有害事象の考え方

有害事象、直接関連があるかどうかは分からない、しかし臨床試験をしている過程でその人にとってはまずいことが起きた、それは薬との関係があるかもしれないし、ないかもしれない、というときに有害事象として捉えるわけです。例えば内視鏡検査をして粘膜の変化はどうかを診ようとして、たまたまズブッとやって食道を破ってしまったとする。これは有害事象でしょうか。因果関係はなからうと思うのが普通でしょう。しかし、昨日のお話 (D. 日本における薬剤疫学の課題、津谷発言) だと、これも因果関係があるかも知れないわけです。例えば薬剤の作用で粘液の分泌が非常に少なくなっているかもしれない。それで滑りが悪くなってという可能性があるわけです。およそ投薬過程で起きたものは全て薬に、関係があるかも知れないと考えてみるべきなのですが、そういうものをだいたい棄却して書いているのが日本の臨床試験です。

また、治験に付随して行われる食待ち (空腹) とか採血、あるいは同じ部屋にずっと閉じ込められるということも、いろいろな影響を被験者に及ぼすわけです。そういう間接的な影響も、すべて説明できるとってはいいけない。説明できるとしてしまうと、例えば抗癌剤の場合、これはガンがあるんだから白血球が減ってもおかしくない、というふうな判断にすぐ結びついてしまうわけです。

### 次のステップに生かされない安全性評価

安全性評価に関しては、前段階の試験、動物実験の結果とかフェイズ1、フェイズ2の結果が次のステップに反映されていない。ソリブジンがいい例です。フルオロウラシル系の抗癌剤を

使っているものは当然省かなくてはいけないというデータが出ていても、それを分かっていて、あるいは見落としている。こういう患者さんを除外しなくてはいいけませんよと書くべき、臨床試験を始める前に規定しておくべきなのに、ソリブジンの場合にはそういうことは遂に行われなかった。

使ってはならない薬を使っている患者さんが、既に臨床試験の段階から入っていた、という点で、すでに致命的な間違いを犯しているのです。

それは一つには、非常にスケジュールが密に立ててあって、第1相から第2相に行くのに1~2ヵ月も間を置かないで、結果がどうなるかもわからない段階で次に移るわけです。そうすると、そもそも前段階で安全性やらを瀬踏みしなければならぬのに、そういう情報が少しも次に組み込まれないということが起こるわけです。

対象患者の選定。これも非常に問題があります。本来の薬の目的から言えば臨床試験に入ってはいいけない患者さん、例えば薬の効果を良くするために軽い病態の患者さんを多く入れる。非常に重篤な病気に対する治療薬として開発しているときに、実際に臨床試験に選ばれている患者さんは非常に軽い人ばかりが全体を占めていて、このために、治療成績が上がっているということもある。

### 「因果関係不明」の本当の意味

それから検査値異常のことも申し上げましたが、副作用と有害事象の区別が非常に曖昧である。

例えばこれ (OHP省略) はソリブジンの件ですが、2番目と5番目にはUFT使用とかフルツロン内服中とかちゃんと担当医のコメントとして書いてある。しかし、実際には、「因果関係不明」で処理している。因果関係不明というのは、何か他に理由があればそこに逃げ込みたいという治験を行う側の逃げ口上として使われている。

関連はなさそうだが、多分関係ないだろう、という表現はあまりあって欲しくないのです。そういうものはだいたい関連がある、と考えなく

てはならないのですが、多くの臨床試験は関連性が否定できないものを関連性なしとしたり、かなり関連性があるものを因果関係不明としている。

#### 甘いリスク／ベネフィット比の判断

先程申し上げたように、リスクとベネフィットの割合によって一般的に使えるかどうか、有用であるかどうかの判断が問われているのですが、その判断が非常に曖昧です。例えば死亡例は最初から有用性評価から外してしまえば関連はないことになる。

あるいは有用性判断を個別毎に行うことにも問題があります。10例の患者さんのたった1例で非常に重い副作用が起こったとします。しかし、それはその薬で1例でも起これば困ることなのですが、そういう場合も10分の1という数がきいてくる。個別に薬によって良くなったか悪くなったか、不都合なことが起こったかどうかどうか、起こっていなければそれは有用性ありとする。そういう評価方法をとっていった最終的には、有用性ありということで縦に串刺ししていきますと、有用性ありという判断の中に、非常に重篤な副作用があっても全体の中に埋没してしまうのです。そういうことで、有用性判断を個別にすることには非常に大きな問題があります。

もちろん、両方から見なければならぬこともあると思います。全体をまとめて比率で見なくてはならない評価もあるとは思いますが、現実には、個別評価に対して全く考慮していない臨床試験が余りにも多いと思います。

概括安全度も同じです。全体を見てどう考えるかという場合、例えばこのあとのセッションで抗痲呆剤が出てきますが、医師によっては個別に診るしか方法がないじゃないか、という人もいるわけですが、その場合に改善とか改善でないとか、だれが見てもきちんとわかる評価基準なしに、非常に曖昧なまま個々のドクターの主観に判定が預けられている。

ランダム割付とか二重盲検とか検査者の主観

やバイアスを省くのに必要な方法はとられていても、それですべて信頼性が保証されているわけではない。概括的な評価方法が持っている危険性を常に意識しながら使う必要があるだろうと思います。ましてもっと厳密で再現性のある評価方法があるのに概括的な評価だけに頼るのは問題があると思います。

## 質疑／討論

**谷田：**今までの話で、現在の臨床試験の問題点とか、毒性試験ではデータをよく見ればいろんなことが見えてくるということでした。

**津谷：**インテンション・トゥ・トリートの質問で先程は良い例がすぐに出なかったのですが、例えば前立腺肥大の薬の臨床試験で女性の患者が来ることがある。しかし、女性に前立腺はないですから、ではそのデータを取り扱うのかどうか。臨床試験は現場の実践ですからいろいろな問題が起きる。インテンション・トゥ・トリートのどのレベルでやるかということになります。

**浜：**承認された薬の中で、これまでの薬よりも効果が強いから承認されることはよくありますが、逆に最近はそれがなかなかむづかしいので、副作用を少なくしたと、それで高い値段がついているのですが、よくよく臨床試験のデータを見ますと、確かに副作用は少ないのですが効果のところを見ますと有意の差を持って効果も少なかった。それが高い値段がついている。こんなことは詐欺じゃないかと思いますが、そういうものもあります。

**谷田：**津谷さんの話で思い出したのですが、臨床試験は女性が意外と含まれないというのは大きな問題ではないか、小児があまりなされていないということ指摘した人がいます。これが有効だと言われる中にもいろんな問題点があるのだということを理解しておかなくてはならないと思います。

**革島：**(医師、外科)無作為化臨床試験に関しましては、わたくしは決してこれは良い臨床試験だとは思いません。例えば安全性に関して言いますと、途中で安全性に問題があるとわかってもできるだけ伏せておくことになっており、よほど重大な副作用でしかも薬が原因であるとわかれば担当医にも知らされるがそうでなければ知らされることはありません。

せん。別府氏も指摘されたように、安全性の評価というのはないがしろにされている。安全性ではなく有効性の仮説だけを検定しているわけです。それを第3相で義務づけているというのは、いわば医者を養成するのに数学や理数系が得意な者を入学させてあとは野放しでそのまま医者になるに任せるようなものです。むしろ安全性のほうを最初にじっくり検討する、有用性のほうは後で時間をかけて調べればよい。つまり入学試験ではいろんな能力の人を入れて6年間で落として行く。薬も仮免許のような形で使わず安全性を調べてから有効性を検討する、そのほうが道徳的というか倫理的だと思います。

もう一点、浜氏の動物実験を行うということに関しての話に痛みを感じられない、動物を一方的に人間の幸福のために無制限に実験に使って苦痛を与えることが良いことなのか倫理的根拠をお聞かせ願いたい。

**谷田：**これは津谷、別府、浜の順にお答えください。

**津谷：**確かに日本の臨床試験は市販前に1000人くらい、アメリカでも2000人くらいです。さっき別府さんから話がありましたように、頻度は分母が大きくなると分かりませぬし、臨床試験というものは基本的に有効性をみるものだと思います。臨床試験中の、安全性情報、今、臨床試験はICHGCPの日本版、世界で行っているのはアドバース・イベントでモニタリングする、因果関係は問わずにモニタリングするというのを今年(1997年)4月から全ての臨床試験に作って、重症度に従って治験医師はいつまでに届けなければいけない、という話がありますから改善されてきたと思います。

第2点の動物愛護の問題ですが、これはかなり国民性によって感覚がかなり違う問題だと思います。日本は人間が大事で動物はあまり

大事ではない。ヒューマン・スクリーニングというコンセプトなのです。例えばエイズの薬でいろんな動物実験をして、最後に残ったものからフェイズ1を始める。ただしそれが失敗する可能性がある。もっと先の段階で可能性があるものをごく少量ヒトに使ってスクリーニングしてみる。それによって全体の開発の期間なりコストを下げてという考えなのですが、ヒトを使ってスクリーニングすることって、日本ではなかなか認められない。やはりヒトはヒトでなくては分からない、というのが世界的なコンセプトだと思います。そのバランスは国によって少し違うと思います。

**浜：** 治験中の安全性について。これは有効性だけでなく安全性をきちんと見るといっていい、ということのほうにむしろ問題だと思います。副作用が起こればその患者さんを除外するか除外した者を有効性の評価に入れていない、そのほうがむしろ問題ではないか。実際に危険性があれば当然それは外すべきであって、それがブラシーボだろうが試験物だろうがそれでおかしなことが起きているな、と思えばそれは常識的な頭で考えて中止するようにしなくてはいけない。それがあるからといって、ランダムイズド・コントロール・トライアルが無意味だということにはならないと思います。

動物愛護に関しては、動物だけではなくて植物も愛護しなくてはならないかなと思いますが、痛みを感じていないという人は多分いないと思います。そういう中で人間の勝手に、人間ほど残虐な動物はいないと思いますが、いろんな物を食べているわけで、これは動物として生きて行くために食べないわけにいかない。それと似てないこともないと思いますが、そういうことまで議論するなら、人間としての存在そのものも、他の動物の存在も全部考え直さなくてはならないと思います。現在のところは人間の安全性を優先して動いて

いますし、それは仕方ないというのは語弊があるかもしれませんが、別の場で議論することだと思います。

**別府：** 試験中に何か悪いことが起きたときにどうするか、いくつかの組み合わせがあると思います。それに対するきちんとした対応を読むプロトコールが少ないのです。だからそれぞれの医師がそれぞれの判断においてやっている。そして事が起こった後で、「ああ、こんなことが起きてしまった。あのとき言われていれば…」となるわけで、やはりプロトコールの不完全さにまず帰すると思います。

動物の問題に関しても、無駄に殺す動物が多過ぎるんだと思います。つまり、開発しなくてもいいような薬まで動物実験をやっているから、せめて意味のある死に方をしてもらうのが供養だろうと思います。

**谷田：** もう時間も過ぎて次のセッションの時間に入ってしまったのですが、もうお一人くらい質問を受けます。

**林：** (医師) 安全性の問題の総合評価、安全度という評価は、重篤なものを見逃すからだと書いておられましたが、僕はそれに留まらないのではないかと思います。別府さんがここで指摘された安全性の厳密性を欠くという問題を覆い被せるような形で、例えば臨床検査などは安全度については臨床試験対象の70%くらいしかしていない、そういう問題が多々ある。第3相においても。そういうことを全部、安全度というオブラートの中にくるみ込んで誤魔化していけるシステムだと思うのですが、その点いかがでしょう。

**別府：** 全くその通りです。何も言うことはありません。

**谷田：** 時間が迫っておりますので、このセッションは終わりにしたいと思います。昨日の話の中にソリブジンのこともありました。アメリカでは患者さんは自分が抗癌剤を飲んでいることを知っていますから、日本の

ようなことは起こらないということで開発を進めたのですが、日本でこのような有害事象が起きたからということで、残念ながらアメリカの食品衛生庁(FDA)での認可を蹴られてしまいました。ソリブジンの誘導体ユースビルと同じような誘導体がエイズの治療薬として今後どんどん世の中に出てきます。ですからソリブジンの問題は決して終わった問題ではない。つい最近、アメリカでサリドマイドが認可されるという報道がありましたし(編集部注：1998年に認可)、サリドマイドは問題が指摘されてからでも長年、今でも、ハンセン病に使われている薬です。

今回のセッションのタイトルが「よい薬、悪い薬の見分け方」ですが、薬というのはそういうところ、つまりある面では悪い作用をしてしまったけれども別の治療には使えるという面がある。一番大事なのは情報公開と、それをよく読めば意外といろいろなことが分かってくる。それを読む、理解するための基礎をこのセッションで、参加者の方々が得られたことを願いますし、わたくしたちも今後もっともっと努力しなければならないと思っています。