

研究対象者保護法制を考える会

科学技術文明研究所

くすりネット・くすり勉強会 共催

「再生医療の医学的評価：骨髄と胎児由来の幹細胞臨床研究を例に」

「再生医療の医学的評価と倫理的論点 - 胎児脳組織移植臨床試験を中心に - 」

福島雅典氏：講義

2004年4月3日(土)

京都大学医学部附属病院探索医療センター検証部教授
(財)先端医療振興財団、臨床研究情報センター臨床試験運営部部长
福島 雅典

§ はじめに

記憶のかなたに忘れかけていた話を光石先生から紹介してもらって、恥ずかしい限りです。今はインフォームド・コンセントは当たり前前の時代になりましたが、私どもがキャンペーンを始めたのが20年前で、そのときは四面楚歌、あちらの思想を持ち込むのはアメリカかぶれで、がんの患者さんにはインフォームド・コンセントはケース・バイ・ケースと言われました。だけど、今や光石先生がおっしゃったような苦闘の歴史があったことは世の中の人誰もご存じありません。

そして当時は患者さんにがんの告知なんかとんでもないと言っていた先生が、今はインフォームド・コンセントの旗頭のような格好をして言っているのを見ると、本当に笑ってしまいますが、今とんでもないことが世の中でいっぱい起こっているということも、きょうのスライドのなかでお話しします。

私は京大で臨床試験の仕組みをつくるというミッションを帯びて仕事をしているわけですが、今まさにトランスレーショナルリサーチの拠点をつくり、さらに神戸に全国的な仕組みとして、その大規模なものをつかって、多くの臨床試験を管理しています。先ほど岡野先生がきれいにまとめられて、いろいろなレギュレーションをしていかないといけないとおっしゃいましたが、まさに的を射た議論だと思います。とても重要な点として、倫理審査レベルが低いことがあります。日本の大学ではほとんど実質的な議論なしに倫理委員会で「研究」としてこの実施が認められている実態があります。

具体的な例は、科研費の申請書類みたいなものを倫理委員会に出して、それでオーケーと言って、場合によっては研究者が倫理委員長の自宅まで直談判に行ったりするという話も聞いたりします。そういうかたちで、患者さんのためにという錦の御旗で、1%でも可能性があったらヒトに対して実験治療を行うということも聞きます。まさに栗原さんがレビューした250例中3例、1%です。こういうのが抗がん薬だったら、そんなのはカスだからすぐ捨てて、棄却するという話になります。だから、わらにもすがると言って、わらにすがったら、患者さんはひどい目に遭うわけです。ほとんどの人、99%は確率的にわらで沈む、副作用に苦しむことになるわけですから、確率の議論をよく理解しておかないと、臨床論文を読んだりして効いているのだと思い込んでしまったりします。

それから、もう一つ、こういう新しい医療というのはそう簡単ではありません。遺伝子治療を見るとわかりますが、まだ何も物になっていません。いまや、分子標的薬、あるいは特に抗体医薬なんて花盛りで、猫も杓子もそんなことを言って、脚光を浴びる新しい医療の開発に取り組んでいます。

かつてTNFがそうでした。今から20年前に、TNFは週刊誌や新聞紙上にぎわいました。もうお忘れの方も多いと思います。今や抗TNF薬は自己免疫疾患の標準治療の一つになりつつあるけれども、日本はそのときフィーバーだったのに、何も物にできませんでした。また、抗体医薬、当時夢のミサイル療法と言いましたが、これも日本は何も物にできませんでした。

つまり、新しい治療を実用化するには20年かかります。はじめて人類が遺伝子治療なるものに取り組んで、もう15年を経過しつつありますが、まだ物になっていないのが現実です。再生医療もそういう状況と考えていいでしょう。だけど、我々はこういう新しいものにチャレンジするというので、いま拠点をつくって、着実といういろいろな仕組みをつくりつつあります。

倫理審査がまだきちんとできていないのは、倫理委員の先生方が倫理審査とは何かというのがわかっていないからです。どういう手順で何を議論すればいいか、わかっていないからで場当たり的になります。そういうことではいけないので、東大医科研、京大、阪大、名大、九大、それから神戸の先端医療センター、5大学1施設が合意して共通の倫理審査ガイドライン、何をどう審査するかの基準をつくりました。

これは臨床評価誌のほうで取り計らってくれて、もう近々出版されますので、ホームページ等でご覧いただけるとは思います。今後は医師、研究者による自主的な臨床研究にはこれが適用されることになります。非常に厳しいものですが、基本的には厚労省の臨床研究の指針を補い、より具体化したものと理解していただければと思

います。来週か再来週ぐらいに出ますので、ぜひご覧ください。

胎児脳組織移植臨床試験

きょうはこの再生医療、いわゆる細胞療法というか、胎児の脳組織移植の臨床試験。先ほどの Freed と Olanow、二つの N I H のスポンサーによる正規の臨床試験を比較して、何が問題でどういうことを議論しなければいけないのか。こういうことを議論するには、どういうことを考えなければいけないのかということをお話しします。

お手元にお配りした資料はきょうのパワーポイントと、それからその二つの論文を読むのに手助けになるよう、両論文のサマリーの翻訳と、それから T A P S、臨床試験の評価のスケールです。論文を恣意的に読んでも何もわかりません。厳密に読むには一定の基準をもとにして読んでいく必要があるのです、その手ほどきに最もいい T A P S のスケールを紹介します。

きょうお手元に配りましたのは、神戸の臨床研究情報センターと京都大学で今年 2 月と 3 月に行ったトランスレーショナルリサーチおよび臨床試験の研修会でもこれを演習として用いました。きょうは、その回答ということになるかと思えます。

このような機会を与えてくださいます、本当に感謝しています。今日、『読売新聞』の行成さんから大変的を得たコメントをいただいていますのでそれを参考に話したいと思えます。先に触れた共通倫理審査指針、もうじきネットで見られますがこれこそ彼の質問に答えるかたちのものです。

ここでまさしく先ほど岡野先生も触れられていた血管再生療法について指摘しなければならない点は、一部の大学病院で「高度先進医療」としてあたかも確立した医療のごとくやっているということです。高度先進医療というカテゴリー自体に私は非常に疑問を持っています。まだこれは実験治療なのにもかかわらず、もういかにも確立した医療のように行われるのは大問題です。

そういう点に触れて、先に、2月に神戸で行ったトランスレーショナルリサーチ研修会と、3月に京大で行った臨床試験研修会で関西医大、京都府立医大の血管再生療法を行っている先生方に発表してもらいましたが、私は厳しく批判しました。きちっとしたプロトコルが提示されないのです。しかも、倫理委員会もほぼ素通りしているようです。そういうのは年内に改善したいと思えます。

臨床医学における論理

まず、こういう臨床研究について議論するのは、それほど簡単ではありません。先ほど来、皆さんはすぐに本当に効くのか効かないのかという議論をしたがりますが、はっきり言ひまして、効くか効かないかというのは、こういう症例を重ねたようなものではほとんど議論の対象外と考えなければならないということです。意味もありません。

図 1

クリニカルサイエンス（臨床科学）の世界は確率と意思決定の世界でありまして、そう単純ではありません。効くか効かないかという議論は決定論的になされていますが、クロード・ベルナルの言う決定論的なロジック（論理）は臨床では通用しません。そもそも入力、これはプロバビリティ A で表していますが、施設で入力自体が違うのです。

例えば、外科の手術でいうと、あまり公にはなっていませんが、実は外科医にも差があります。つまり、入力には差があるのです。そして患者さん、ここはまさにブラックボックスです。試験管内の実験やマウスやモルモットの実験とは違います。十人十色で条件をそろえられないのです。そして出てくる結果はプロバビリティ B でしか表わすことができず、実験系はコントロールできません。まず、それを理解しないとイケません。

パーキンソン病というのも、単一の病気ではなく、多因子疾患です。しかも、外部要因があります。そして、薬物療法への反応の問題、併存疾患の問題がありますが、これだけ考えるだけでも非常に複雑な事であることが分かります。で、そこにインプットして出てくるアウトプットを評価するわけです。いくつかの症例報告でしかない論文で効くのか効かないのか、効果だけを議論してもしょうがないのです。

しかも、移植片となった場合には、この品質管理なんてまさにアウト・オブ・コントロールです。そして、一人ひとりの患者さんで意思決定しないといけないのだから、実際はもっとややこしくなります。医療として新しい方法を確立していく過程、すなわち、すべて技術の評価は確率論の世界であって、まずここをまず踏まえないといけません。このロジックがわからないと議論できなくなりますし、科学的見地から見ると、議論はおよその外れな、効くか効かないかというほとんどナンセンスな議論になってしまいます。

医療技術の評価

医療技術の評価するには原則があります。今日は与えられたテーマが胎児脳組織移植について論文の評価から倫理的な議論ということですので、医療技術の評価について、本当は臨床試験の話をしてからでないとう理解できないですけれども、骨子だけ言います。

表1

1例ずつ症例を積み重ねる、またはパイロット試験というアプローチはありえません。しかも、プロトコルのないままやるというのはうまくいくことのない無謀な冒険的な実験であって、やってはなりません。プロトコルのない臨床試験、臨床研究というのは100%ありえません。それはナンセンス以外の何物でもありません。

通常は難治性の疾患については患者さんには何らかの効果があることが前提ですから、デザインとしては - 相ということになります。そして 相試験というのは、はっきりした効力試験ができるものに限ってやることができます。ですから、トランスレーショナルリサーチというのは基本的には - 相、そういうデザインでないといけません。

患者さんに全然ベネフィットがないような試験をたとえば従来のタイプの第 相試験を、患者さんにいいですかとたずねてやるなんていうのは、それはヘルシンキ宣言違反になりますし、それは患者さんにとって意味のないことだから、やってはけません。だから、何らかの一定の確率でベネフィットが得られるという保証がないと、実際にはやれません。

最終的にランダム化比較対照試験でのみ、有効性について議論できますが、第 相までの段階で有効性について議論してはなりません。これが新規医療技術の評価についての教科書的な原則です。

TR臨床試験の原則

以上のような原則を理解せずに、有効性という言葉を使っているものから、おかしくなります。果ては有用であるなんて書いている論文もあったので、笑ってしまうというか、馬鹿馬鹿しい限りです。言葉の用法は厳密にしなければなりません。Efficacy (効力、臨床効果) と effectiveness (有効性) は違うのです。そして usefulness (有用性) はもっと広く、経済的、社会的意味も含めた評価になります。これらは教科書的にきちんと定義されている用語ですから、間違えないようにして下さい。

だから、日本では単純に抗がん剤が効いた効かないというけれども、あれは response、response rate といいます。response は反応です。ですから、有効率では間違っているし、奏効率というのもインチキで、反応率と言うべきものなのです。本当の意味での有効性というのは、やはり survival、生存がどれだけ改善したかで議論しないと行けません。そういうサロゲート・エンドポイント、代理エンドポイントといいます。がんが小さくなった、よって効いたという論はだめなのです。本当に長く生きただろうか、そういう最終的なエンドポイントで評価しないと有効性についての議論はできません。これが原則です。あとでひどい目に遭います。研究者の側からいくと、延々と人生をかけてやってきた研究がパーということになります。

表2

臨床試験、新しい治療の開発のトランスレーショナルリサーチというものは、医療の一形態として社会的に認められる形でなければなりません。難治性疾患に対する最終的な選択肢として妥当でなければなりません。です

から、その時点でやる施設、医師はまずその疾患の治療について State-of-the Art に達していないとだめです。つまり、日常診療レベルが最高水準にあること。そして、その証拠のデータ（臨床成績データ）が公開されていることが前提となります。

近々かなりの施設の肺がんの治療の手術成績が公開されます。これは世界的に見ても初めてです。肺がんの手術の成績は、日本は平均的には世界のトップです。ただし、それぞれの各病期別、つまり進行度合い別に見て、治療成績に差があります。ある施設では90%が5年生存しているのに、ある施設では50%に満たないということがあります。それらが施設名入りで公開されますから、かなりのインパクトがあると思います。

そういう State-of-the Art に達していない施設で、肺がんの新しい遺伝子治療なんてやったら、とんでもない話は子供にも分かる話で、血管再生医療についていうならば、CABG、P T C A の治療成績が State of the Art にも達していないような施設で、もし心筋や心臓の血管再生なんていったらとんでもないことです。

それからもう一つ、これは研究ではない、事業です。だから、従来の科研費ベースで、研究費をそこに充てて研究としてやるというのはだめです。ちゃんと事業として位置づけて、国のなかで方針をもった上でやらなければいけません。今はそういう方向に文科省は方針転換しています。

表3

そして確実に試験物の安全性と臨床効果について一定のデータが得られるようなデザインで実施しないとだめです。これは岡野先生も強調された点で、当然のことです。ですから精度の高い研究のデザイン、統計学的なデザインがなかったらだめです。先ほど言ったように1例1例積み重ねていくというやり方は通用しないので、やってはだめです。ですから、統計家が最初から関与しないIR 臨床試験というのはあり得ません。客観性、普遍性、再現性が保証されていないようなやり方は駄目です。これは科学の科学たるゆえんで、科学的な基盤が不可欠。これをどうするかというところで、日本にはまだ科学的基盤がなくて、私どものミッションはこれをつくるということで、ほぼつくり上げました。

それから社会的に公式に認められていなければいけない。したがって水準の高い倫理審査が不可欠です。先ほど共通倫理審査指針について触れましたが、ようやくわれわれはこれに対しても一歩を踏み出しました。将来的にはFDAの査察に耐えられるところまで持っていくという方向で整備を進めます。

それから当然、被験者の同意が得られ、安全性が確保されなければなりません。ですから、万全の医療体制がないとだめです。リスクマネジメントの体制が整っていないようなところでやってははいけませんし、有害事象が起きたときの対応マニュアルがないとだめです。こういうことは、今回の共通倫理審査指針で記しました。

ですから、まず論文を読む場合は、少なくとも以上のような基本的知識をもった上で、評価しなければいけません。そして、論文を読むには手順があるのだということです。その手順をお手元に配っていますが、日本語に訳してありますので、論文を見るときは、それを使って評価してください。

臨床評価手順スケール（T A P S）

表4

すなわち、T A P S、臨床試験評価手順スケール、Trial Assessment Procedure Scale です。これは臨床試験の品質評価を目的とした系統的技法です。品質評価というと違和感があるかもしれないけれども、臨床試験の質、クオリティを評価するということです。つまり、ここでいう品質とはその臨床試験は科学的な論文としてまともに議論するに足る意味のあるものかどうかということです。品質の低い論文なら、まゆにつばをつけて読むか、あるいはごみ箱に捨てます。

表5

T A P S はプロトコル・試験報告書、専門誌論文の分析に使います。でも、これは例えば試験物の品質とか何かについての評価項目はないので、臨床試験に入ってしまったからの話になりますから、このT A P S だけで研

究全体を評価することには限界があります。試験全部の品質については別に評価が必要です。

TAPSでは臨床試験を8つの要素に分けて、それぞれの要素について質を独立して評価します。ですから、これは治療の効果とか安全性に関する知見に関係なく、試験の品質を格付けすることができます。

表6

TAPSの8項目の第1は研究課題です。背景と根拠、ラショナルレ、バックグラウンドそれから目的または仮説。これらはそのままプロトコルの項目に相当します。そして、外部レビュー、モニタリング、研究の管理、これがあるかどうか。あるいは倫理審査とかクオリティアシュアランスつまり、品質の保証の仕組みをちゃんと臨床試験の実施に際して持っているか。施設選択、組織。研究組織はどういうものか。研究要員のレベルはどうか。研究期間、これらもすべてプロトコルの項目そのものです。

表7

そしてデザインの特徴として、独立変数、エンドポイントを何にとっているか。デザインの構成、どのように検査とかの項目を評価しているか。そのスケジュール、研究のカレンダーはどうなっているのか。被験者数、割り付け、それから治療関連のバイアスをどう管理しているか、統計的にどのようなデザインか、これらすべてが重要な項目です。

統計的な要点としては、解析手法を最初にデザインして、最初に決めておかないといけません。あとから結果を統計的に検討したらどうだったという議論はナンセンスです。検定方法を変えれば、有意に出ることも多いのです。それから外来変数、外部要因の変数をどうやって管理するか、これはものすごく難しい問題です。

そして治療の特徴について、どのように記述されているか。治療計画はどうか、薬剤情報、治療変更基準、有害事象、これらについてきちんと記述があるかどうかを見ていくわけです。ですから、こういう項目について具体的に追試できる形で記述がなければ、プロトコルに記述がなかったということですから、そこに研究の欠陥があったということを指摘します。

表8

そして被験者の特徴について、選択規準、適格規準、除外規準、そしてサンプルの代表性。ある試験で、例えばがんの臨床試験で、対象とした患者さんが非常に限られた特殊なタイプなら、サンプルの代表性はなくなるから、その研究結果をもってほかの患者さんにも適用するような議論は全然だめなのです。

こういう基本的なことについて厚労省の当局の担当者はよくわかっていません。だから、イレッサでも適用を臨床試験での適格規準から大幅に拡大して承認し、大規模な薬害を起こすのです。こういう科学のごく簡単なことをわかっていないから、薬害がしょっちゅう起こり、反省する能力も欠くため同じ過ちを繰り返しているのです。

それから被験者の組み入れの原理がどうなっているか、規準どうなっているか。それからリクルートの手順がどうなっているか。そして、被験者のコンプライアンスがどうだったか。治療に対するコンプライアンスはどうだったか。特に内服薬みたいなものでコンプライアンスについて記述がなかったら、全然信用できないということになってしまいます。実際にふたを開けてみたら、半分の人しか薬を飲んでいませんでしたと、そのようになりやすいのです。

評価の範囲、有害事象。評価の方法、評価基準。それから評価のスケジュール、評価の実施と、このようにデータの収集についての記述がどうなっているか。こういう記述があるかどうか。あったとしたら、そのレベルはどうか。

表9

データの分析。症例報告書をどのように回収して、それをデータベース化して解析にもっていくか。研究者が

自分のところでデータをとってきて、自分でコンピュータ解析したら、それは全然客観性がないから臨床試験として信用できません。そういうのはインチキだと考えます。

あとは結論と解釈。得られた結果を実際に具体的にいろいろなことに適用できるのか。追試できるようなものなのかどうかの検討ということです。

パーキンソン病に対する論文の評価

表 10

このようにして、パーキンソン病についての二つの論文を評価します。パーキンソン病というのはどういう病気かというのを教科書から取り出しますと、要は体がだんだん動かなくなって、硬くなってきます。この病気は、病的には中枢神経の変性疾患です。脳の黒質線条体系の変性がある、ドーパミンが枯渇しています。だから、治療はドーパミン作動薬とか抗コリン薬で、このバランスを取るようにします。あと、モノアミンオキシダーゼ B の阻害薬というものも使ったりします。そして予後は、日常生活の活動が徐々に困難となって、50%に痴呆が出て、うつ状態がよく見られます。だから、この病気のことをよく理解していないと、当然のことながら Freed の論文と Olanow の論文を議論することはできません。だから、パーキンソン病という病気についてきちんと勉強した上で、議論しなければいけません。

まず、いずれの研究についても、移植品の品質について大問題があります。ドナーは選択的中絶と書いてあるだけで、選択的中絶の内容については一切書いてありません。それから移植片は胎児の脳組織ということで、サイズについても片方は固形としか書いていないのに、片方は 200 μ m と書いてあるのです。そして採取部位は中脳となっていますが、中脳をどのように同定したかも、よくわからないわけです。

表 11

それから採取した脳組織の性質としてはドーパミンニューロンを含んでいると書いてあります。そして、Freed は受精後 7 ~ 8 週の胎児、Olanow は 6 ~ 9 週の胎児で、培養ないし保存については、Freed は 6 週までに使って、Olanow は 2 日までに使ったとしています。ドナーの数は、Freed は一人の胎児から採ったものを患者の脳の片側に、もう一方には別の胎児から採ったものを入れました。だから、二つの胎児から採ったものということです。Olanow のほうは患者の片方の脳に 1 または 4 で、三群のランダムイズトライアル。よくも思い切ったことをやります。ふつうの非常に簡単なものでも、三群にすると臨床試験というのは失敗する確率が高くなるのです。このようにして、両側の被殻に胎児由来の脳組織を入れるということです。

そして、移植量は Freed 論文には書いてあるけれども、Olanow のには書いてないということで、ここらへんの品質管理についても、もしもこの試験を追試しようとしたときに、どうやっていいかわからなくなってしまいます。つまり、追試しようにも、どこをどのようにしたかわからないような論文というのは、あまり信用しないほうがいいということです。

表 12

それからエンドポイントについて。エンドポイントとは、最終的に何で評価するか、その指標のことですが、Freed のほうは GRS、Global Rating Scale で、総合的な全般的な改善度を指標として評価しています。これが彼らの研究での主要エンドポイントで、12 カ月後にこれを評価します。そして、Olanow のほうは UPDRS、Unified Parkinson's Disease Rating Scale で評価しています。術後に患者さんの経過を追ってデータを取り、統計的に検討して優劣を見ます。

そして記録者は、Freed は患者本人、Olanow では第三者。結果は、主要エンドポイントについてみると Freed は無効、Olanow も無効。両研究とも、目標としたエンドポイントに関しては無効でした。また、副次エンドポイントとして、Freed の方は UPDRS をもって、Olanow は Fluoro-dopa の PET の所見をもって評価しました。そして、Freed は Schwab & England Score をもう一つ副次エンドポイントとしています。

Freed は、60 歳以下では有意差があったと強調しています。ほかの指標はさておき、Olanow は、取り込みが有

意に上昇したと主張し、生着の証拠であるということです。いずれにせよ、この「効果」の主張は科学的に意味がない。後からのサブセット分析は？に誤りと考えてよいのです。せいぜい仮説としての意味しかない。Freedの論文で60歳以下で有効というのは仮説でしかない。このように、研究の重要なスタートとフィニッシュ、つまり、移植術の品質、脳外科手術のコントロールのとり方、そして評価方法、いいかえれば臨床試験のインプットとアウトプットについて、このように客観性と信頼性に重大な欠陥があることが明白です。

TAPSスコアリング

表 1 3

次に、TAPS：臨床試験評価スケールで両研究の質をレーティングしてみます。まず Freed の論文。研究課題、研究管理、デザインの特徴。デザインはランダム化治療後にクロスオーバーする方法。コントロール群、プラセボというか、シヤム手術にあたったほうは、あとから希望があれば移植を実施します。こういうのは、はじめからして研究デザインとしてだめです。最終的なエンドポイントが追えなくなってしまうからです。もっと低い点を付けてもいいです。

そしてこの研究における治療の特徴というのは、まずシヤム手術を行うことです。先ほど岡野先生も言われましたが、私も同感で、シヤム手術なんていうのは倫理的に許されませんし、本当にいいコントロールとは思えません。

被験者の特徴についても、被験者のバリエーションについての説明は何もないわけです。それが均質だという証拠はどこにもありません。

あと、データの収集、データの解析についても非常に低い点となります。データの解析については、特にサブセット分析して有効性を強調しているところは、もう開いた口がふさがりません。サブセット分析というのはレトロスペクティブな分析ですから、いかなる結果であれ、これは仮説を提供するものでしかない。通常サブセット分析というと、統計家はカスだ、全然信用できないと言います。サブセット分析の結果は、どんな結果であっても仮説でしかないのです。検証ではありません。

だから、総合的な品質は全体として見て50点ぐらいですが、具体的に一つずつ評価をしてみると、41点で不合格です。だまされてはいけません。『New England Journal of Medicine』であろうが、何であろうが、読む側は冷静に厳格な方法で研究論文を評価しないとイケません。どこそこの学術誌に出たから、インパクトファクターいくつだからといふ云々というのは世界の笑い物です。よく見ないとイケません。インパクトファクターで論文を評価してうのみにするというのは、極めて危険だナンセンスです。それから、インパクトファクターをもって研究者を評価して、それでうんぬんするのもまったくナンセンスです。そういうことをやっていたら、今日のように科学がビジネス化している時代にはなおさら問題で、地道なサイエンスは滅んでしまいます。

表 1 4

Olanow のほうはクロスオーバーなどというばかげたことはしていないから、まだまだけれども、コントロールとしてやはりシヤム手術をやっています。あとのところは、もう少し具体的にモニタリングをどうするかということも書いてありますから、こういう点数を付けました。

これは論文を緻密に読むということで、臨床試験、臨床研究の場合にはそういう評価をしないとイケないという目安ですから、研究者も、メディアも、市民もこういうことに慣れるのが、まずこういう議論する第一歩だろうと思います。

本質的な問題点

表 1 5

問題点を整理します。まず、疾患自体の問題。パーキンソン病は多因子疾患で、患者は不均一です。どんな病

気でも基本的には不均一です。だから、数を上げればいいというのは統計的にはそうだけれども、そう単純なものではありません。

また、パーキンソン病に関しては、病期、病態分類がまだ不完全です。ここに非常に大きな問題があります。それから予後因子も不明です。何が疾患を増悪させるかについてもまだ不明な点が多すぎます。

臨床経過に変動があります。具体的な例をいうと、私の知り合いのパーキンソン病の患者さんが体操と整体を合わせてするスクールに通い始めたら非常に調子がよくなったのです。そのようなことさえあるのです。薬物療法が行われますし、その効果も決して安定しているとはいえません。当然のことながら反応は患者さんによってまちまちです。

表 1 6

そして移植片の問題は極め付きです。まず第一に臨床試験で最も重要な点ですが、そもそも試験物（医薬品）を規格化できない、標準化できない、定量化が困難。例えば、この典型的なたぐいが漢方薬で、日本の論文で漢方薬についてランダムイズトライアルはいくつかあるけれども、それで本当に評価できるものは何もありません。科学になっていない。一言でいえばクズです。英文誌にパブリッシュされているものをもってしてもです。

米国の標準テキスト、『カレント・メディカル 診断と治療』に代替医療、補完医療についてのチャプターがあって、ハーブについての評価の記述があるけれども、アメリカで大規模にこういうものを検討し始めると、規格化、標準化していくたびに有意差がなくなって、結局は効かないということになってしまうことが指摘されています。

だけど、例外もあります。例えば針治療について言うと、日本ではふつうの医学部の先生方とか学生はばかにしているけれども、アメリカのその最新テキストには、例えば歯の痛みについては針治療がファーストチョイスの一つであって、ランダム化比較試験で実証されていることも記述されています。

話がそれましたが、もう一つ薬効原理の実証（PDP）にかかわる重要なポイントである。ドーパミンニューロン数、活性の定量化が困難です。Freed の論文ではこの点について触れてはいるけれども、具体的にはまったくあいまいで、どの程度の活性があるか、どういう刺激に対してどのように反応しているか、訳がわかりません。

用量 - 反応が不明であるということは非常にやっかいで、データ解釈を正確にできなくしている。以上のように、試験物そのものの品質の保証がない、効果・安全性の根拠付けのもと自体の情報が不十分ということです。

表 1 7

次にエンドポイントの問題があります。真のエンドポイントは、パーキンソン病に関しては 50% にできる痴呆または最終的な結果である不能または死亡ということになります。ですから、これを見ない限りは、単純に症状がちょっとよくなっているから、これは効いているんだという判断は下せません。

症状の一時的な改善や、何らかのマーカー値の改善などの代理エンドポイントは常に誤りと考えてよい。これは今までの 50 年間の抗がん剤の開発でもはっきりしています。がんが一時的に小さくなくても生存期間が必ずしも強じるとは限らないのは事実です。過去にさんざんこの誤りを繰り返してきましたので、まして、主観的な指標の改善度とか、症状がどうなったという評価は非常に危険です。いったんよくなったかに見えたけれども、早くに不能になってしまったことということでは、目も当てられません。

抗エイズ薬でもそうです。典型的な例が AZT で、これが出たときに早期に投与すればエイズの発症を遅らせるのではないかとみんなが思いました。私も思いました。ですけど、アメリカとヨーロッパで 1000 名を超す大規模臨床試験をやったらまったく逆で、ぴったり両方のデータが一致して、AZT を早期に飲んだらより多くの人々が早く死んでしまうことが判明したのです。そしてより多くの HIV キャリアで、早くエイズを発症してしまいました。ですから、代理エンドポイントは常に誤りと考えてよい。真のエンドポイントについて周到にデザインされた臨床試験で評価しない限り、結論はでないのです。

ソフトなエンドポイントはどうしても妥当性、信頼性、客観性に保証がないということです。

表 1 8

それから治療方法。シヤム手術というのはコントロールとして明らかに不適切です。脳外科的治療として、もうすでに標準的な手術として淡蒼球の切除とか、あるいは深部の脳の刺激とかがあります。かなり難しいとは思いますが、基本的にはこういう脳内に介入する手技があることはよく考えておかないといけません。ですから、試験群で移植をやってその効果を見ながら、コントロール群でこの確立した手術をやるというようなデザインなら評価に耐えられたかもしれません。

それから薬物療法を併用しなければならないというのは、決定的な難点です。抗パーキンソン病薬を飲ませておいて、オフのときにどうのこうのというのは、我々評価する側から見て、とてもそれはやっていられないという感じがします。それから、あと免疫抑制薬の意味がまだわかっていません。Freed のほうはそれらを使っていないし、Olanow は使っています。ですから、生着について比較できません。

それから外部要因のコントロールが全く不能です。先ほど少し触れましたが、睡眠、ストレス、支援の質で、パーキンソン病というのは変動があります。しかも、もっとやっかいなことに State-of-the Art (治療成績) が不明であります。ですから、この治療のレベルがどうか、もともと基準がありません。

胎児脳組織移植臨床試験とは

表 1 9

これまでのことはある意味机上の話です。ここで具体的にこの試験を眼に浮かべてみるとよいと思います。いったい何をしたのか。いったい何をしようとしているのかというのを、自分の眼でもっと生き生きと見てみないといけません。それが倫理的なポイントになります。

図 2、3

これは京都大学の解剖発生学の専門の塩田教授からお借りしたスライドですが、7週の胎児はここまで形になっています。もうちょっといくと8週で、もうちゃんと姿形がわかります。選択的中絶というけれども、死亡胎児ではないのです。それは、生きている胎児です。婦人科の医者に聞くと手足が動いているといいます。

どのように脳組織を採ったか、手技がよくわかりません。中脳というのはこのへんですけれども、脳の領域をこのような胎児から、太い注射針のようなもので吸い出すということです。そして、一つの胎児から1ほじりを入れて、もう1ほじりを別の胎児から採ってきて入れます。Olanow の論文では一人の患者さんに対して2または8胎児から、こういうことをやったのです。

表 2 0

だから、結局、フレッシュな脳がほしいということになります。この領域の研究者、誰もが言います。先日、ある会合では、産総研の金村先生がフレッシュな脳がよいだろうと言いました。そして、もうはっきり言ったほうがいい。彼は、国立大阪病院では中絶胎児の研究について、インフォームド・コンセントをマザーからとっている、そして主治医ではなくコーディネーターからインフォームド・コンセントをもらえばよい、そのためにコーディネーターの養成をする、などと狂気じみたことを言っていました。

図 4

そういうことをいみじくも非常にリアルな小説として描いたのが、帚木蓬生先生です。彼はまぎれもなく、日本の医療倫理、生命倫理の第一人者と言っていいと思いますが、彼の作品、『エンブリオ』をぜひ読んでいただきたい。胎児利用はビジネスと結びつくということです。今はサイエンスはサイエンスとして独立していない。もうれっきとしたビジネスです。ナイーブな考えは捨てて下さい。

今後の展望

表 2 1

ですから、結論と今後の方向性は明快です。これは禁止するべきだと私は思います。NIHが助成した2つの正規のランダム化臨床試験の結果が出て、いろいろな問題点が明白になりました。胎児利用は直ちに法律で禁止すべきです。

ES細胞はもうつくってしまって、京都大学の中辻憲夫先生も供給を始めていますが、厳格な管理の下に推進します。そして、それで一定の経過を見ていきます。

それから自家幹細胞の移植研究は当然、推進します。そのうち、クローン胚を利用したくなるに決まっています。ですけど、きょうの新聞に出たように、アメリカ大統領の評議会ですか、クローンからの人間の製造は禁止しました。それからもう一つ、胎児の人格を受精直後から認めるという法律が制定される方向になってきております。ここらへんが行き着くところではないかと思えます。

パーキンソン病に対しては、もっと冷静に考えて、まったく別の方法を模索するという手があります。脳の刺激をするということで、ある程度の効果があることがわかってきているのならば、いろいろな手があるはずで、もっと冷静に考えたほうがいい。

医学の本筋として、疾患の征圧は予防、早期診断、早期治療が決め手なのです。がんも積極的に予防する時代です。米国では死亡率が下がって、さらに発生も減少しつつある。日本ではがんはどんどん増加すると多くの人が思っているけれども、米国のようにそれを低下させることはさほど困難なことではありません。

例えばがんの発生率を下げようと思ったら、たばこを禁止しなければだめです。公共の場からたばこを追放することが、まず第一に行うべき治療、社会の治療です。

それからPETで、術後にフォローすることによって、早期に再発治療することができるから、治療成績はもっと上がります。いま、日本の手術、いろいろながんの治療成績というのは紛れもなく世界のトップです。ですから、これをもっと充実させることによって、治療成績は安く確実に上げることができるのです。

この表で見ると上から下にいくにしたがって、アプローチとして有望または確実です。

表 2 1

先ほど言いましたように、再生医療というものはまだ10年20年はかかると考えないといけません。

倫理的論点

図 5

最後に、倫理的な論点をまとめてみます。今日誰も主張しませんが選択の自由だと、自己決定権、幸福追求権、生存権だと、たしかにそうです。これを科学に問いかける。そして科学は市場と結びついてそれに応える。こういうサイクルの行き着くところは何か。0.ハックスリーの『すばらしい新世界』ではないかというのが、私に限らずいろいろな人が指摘しております。まず、こういうことを議論するのなら少なくとも梶木先生の『エンブリオ』は読まなければいけませんし、『すばらしい新世界』も必読書です。

幸福とは何か、人間の価値とは何かというのを足元から問い直さないと、市場で、選択の自由だ、医療特区だなんてやっていると、最終的には必ずそれはEugenics、優生学に結びついてきます。まさに、自由市場という名目下の選別に他ならず、ナチスのやったことを批判なんてできません。結局はそういうところに行き着いて、おかしいことになっていくのではないかと。何かそういう感じがします。

図 6

これがその『すばらしい新世界』ですが、講談社文庫から出ていますので、ぜひ読んでいただきたいと思います。倒錯した世界、クローンの世界です。1932年に初版が出ています。

まとめますと、非常に重要なポイントがあります。サイエンス、とくにクリニカルサイエンス（臨床科学）とは何か。先ほど述べたごとく確率と意思決定の世界で、その原点は、自分が病気になったときにどうかということとあります。すぐにとびついてはいけません。それから、有効性とは何か。サイエンスというのはどのようにしてそれを決定するのか。そういう臨床試験の原理と原則そして実際によく理解しないとはいけません。そういう研究を支える基盤が、倫理審査です。

表 2 2

最後に、まさに今、われわれがこの問題の方向について決定しないとイケない時期にあるのではないかと痛切に思うのであります。

T R 共通倫理審査指針 2004

そのような背景から、トランスレーショナルリサーチ（T R）に関する共通倫理審査指針というのをつくりました。昨年8月1日にT Rをすすめる5大学と1研究機関が懇話会をもちまして、各先端医療に取り組む責任者が集まって非常に激しい議論をしました。

表 2 4

T Rに関して、研究者の行動規範、倫理審査の水準の確保の努力、それから業務手順を定めました。そして、倫理審査委員会への提出書類と審査受け付け受領の条件というのを決めました。そして倫理審査の手順も決めました。それから監査を行うように規定しました。

表 2 5

すでに厚労省の臨床研究の倫理指針がありますが、指針では倫理審査委員会の提出書類を、はっきりと、概要書とプロトコルと説明同意文書と重篤な有害事象発生時の対応マニュアルと規定しました。そして、それらに含まれるべき内容も記述しました。

それから審査の手順は書類がそろっているか、必須項目についてきちっと記述があるかどうか。それから、概要書の系統的な審査、プロトコルの系統的審査、被験者への説明、同意文書の系統的審査、それから重篤な有害事象発生時の対応、マニュアルの系統的審査をどのようにするか、明確に記述しました。

表 2 6

審査にあたって、倫理委員が発するべき疑問、こういうことについてはどうなっているか。そういうかたちでクエスチオンの形式で列記してあります。それらについてきちんと回答できるようなものであれば、まずは本当の意味での審査の対象となり得るのです。そしてそれをやってみるべきだと倫理審査の先生方が合意されれば、承認したらいいわけです。

倫理委員の方たちは、今は申請書とやらをばらばらめくったり、いろいろ頭を悩ませたりしているけれども、まずこれだけのことをしないと倫理審査のクオリティコントロールができないということで、これを決めました。これによって、少なくとも我々の施設、それから他の施設は倫理審査をして、実際に臨床試験をやっていくこととなります。すでに『臨床評価』のホームページで見ることができますし、各施設のホームページでも見ることができます。（<http://tri-kobe.org> , <http://www.kutrc.org>）

ですから、この点については読売新聞の行成さんのコメントに対する一つの回答になろうかと思います。以上でございます。（拍手）