

The Informed Prescriber



正しい治療と薬の情報



医薬品・治療研究会

Feb. 2008 Vol.23 No.2

Critical Choices in Drugs and Therapeutic Alternatives

提携誌

Adverse Drug Reaction Bulletin (英) / Arznei-telegramm (独)

Der Arzneimittelpflichtbrief (独) / Arzneiverordnung in der Praxis (独)

Australian Prescriber (豪) / Drug & Therapeutics Bulletin (英)

Drugs Bulletin (印) / Pharma-Kritik (瑞) / La Revue Prescrire (仏)

Drug Bulletins Review (WHO)

臨床試験は常に患者に利益をもたらすか?

Clinical trials are always for the benefit of patients?

シルヴィオ・ガラティーニ*

本稿は、昨年第13回日本薬剤疫学会学術総会で行われたマリオ・ネグリ薬理学研究所所長シルヴィオ・ガラティーニ教授の特別講演である。臨床試験の中には、読者を誤解させるようなさまざまなバイアスが含まれており、その問題点を具体的事例をあげながら解き明かし、新薬承認の基礎となる臨床試験が本当に患者の利益になっているか?という厳しい疑問を投げかけている論説である。昨年のTIP誌12月号に紹介した「インデペンデントな医療品研究の推進」は、この講演後の質疑応答で明らかにされたイタリアでの新たな試みだが、この講演録を読むとなぜそのような試みが必要となったかがよく分かる。長い講演だったので、2~3月号の2回に分けて掲載する。

H.B.

おはようございます。日本薬剤疫学会にお招き頂きましたことは、私にとって大変光栄であり、この会議に出席できますことを大変嬉しく存じております。まず最初のスライドで、つい最近新装なったマリオ・ネグリ薬理学研究所の建物をご紹介致します。今日、私は臨床試験に関してお話ししたいと思います。ご承知のとおり、臨床試験は、新薬の効能と危険性を確かめるために非常に重要なプロセスです。臨床試験がなければ、薬効に関するエビデンスは存在せず、単なる効き目の印象でしかありません。したがって、私たちは臨床試験に頼らざるを得ません。

大きな商業利益がからみ、しかも、その80%以上が製薬会社自身によって実施されたり、その財源で賄われるため、試験は商業的な利益によって操作され、バイアスを生じ、試験結果にも影響を及ぼすかもしれません。したがって、薬の重要な側面と弱い側面をきちんと見分けるためには、臨床試験の批判的分析を行うことが非常に重要であり、そうすることによってはじめて薬の作用をよりよく評価することが出来るのです。ここでは、私たちが考慮に入れなければならないこれらの重要な問題について論じてみようと思います。

ご承知のとおり、通常、臨床実験には3つの相があります。第1相試験では忍容性が検討され、ヒトにとって適切な用量を決定します。そして、第2相試験では、その薬に何らかの効果があるかどうかを確認しようと試みます。すべてがうまく行くと、次はランダム化比較試験と呼ばれる第3相試験に進みます。そこでは、その薬の利益と危険性が研究され、既存薬との関係でこれにどのような位置付けを与えるべきかが検討されます。

残念ながら、これらの臨床試験には大

プラセボの使いすぎ

まず第一の問題はプラセボの使いすぎという点です。ご承知のとおり、臨床実験の倫理的側面を規定したヘルシンキ宣言では、プラセボの使用は他に使える薬が全くない場合に限って許されると述べています。他に利用可能な薬があれば、プラセボを使用すべきではありません。しかしながら、臨床試験では非常に頻繁にプラセボが用いられており、他に利用可能な薬があるときですら使用されています。これはもちろん、その薬の効果に関してより楽観的な見通しを得るためです。何故なら、既存薬(同効薬)より勝ることを証明するよりは、プラセボよりも勝ることを示すほうが容易だからです。

プラセボ使用の問題点を示すために、ある研究をご紹介します(図1、表1)。詳細に立ち入る必要はありませんが、研究は次のようにデザインされました。対象は糖尿病患者であり、患者は2群に割り付けられました。いずれの群も、従来の糖尿病薬であるメトホルミンに加えて、ロジグリタゾンかピオグリタゾンのいずれかが投与されました。次に、片方の群にはエキセナチドが、他の群にはプラセボが投与され、その効果が比較されたのです。ご覧のように、エキセナチドを投与された群ではプラセボ群

イソトレチノイン	23	シクロスポリン	15	フルオキシセチン	16
エキセナチド	13	ジクロフェナク	15	フルダラビン	22,23
エトボシド	24	シクロホスファミド	24	プロピルチオウラシル	24
エボエチン _a	22	スルフォニル尿素	16	ベニシラミン	23
オセルタミビル	21	セボフルラン	24	ミトキサントロン	23
オランザピン	15	タクロリムス	15,22	メトホルミン	13
クエチアピン	15	デキサメタゾン	19	メルファラン	22
クラリスロマイシン	24	トリメトプリム	23	リスベリドン	15
克蘭ベリー・ジュース	22	ナプロキセン	15	リツキシマブ	23
ゲフィチニブ	17	ハロペリドール	15	ロジグリタゾン	13
コルヒチン	24	ピオグリタゾン	13,23		
サラゾスルファピリジン	23	プスルファン	22		
シクロオキシゲナーゼ2	15	プラセボ	13		

目次

臨床試験は常に患者に利益をもたらすか?	13
未熟児への初期ステロイド剤治療: 神経系の後遺症	19
2008年1月号「オセルタミビル(タミフル)と異常行動」正誤表	21
CAPSULE	22

*マリオ・ネグリ薬理学研究所(伊)

図1

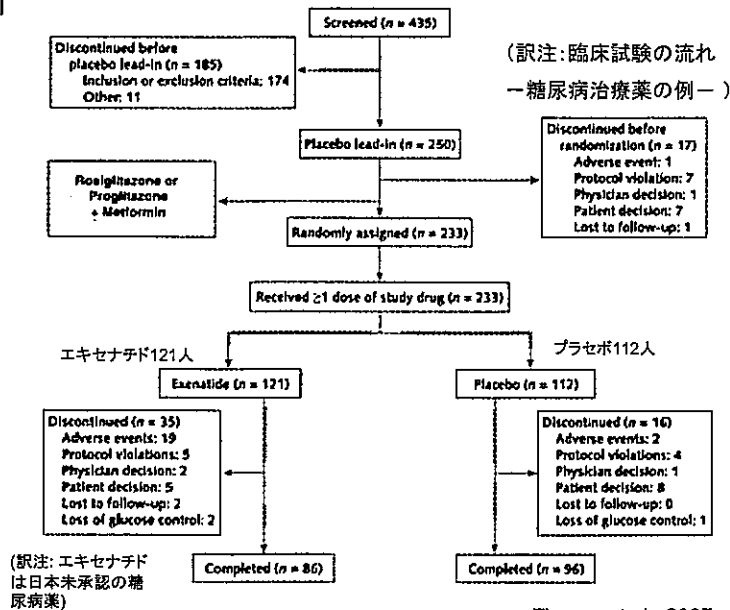


表1: 効果の推定: ベースラインから16週目までの平均変化量 (糖尿病薬の治験から)

評価項目	エキセナチド (n=121)	プラセボ (n=112)	差 (95%CI)
HbA1c 値			
ベースラインレベル, %	7.89	7.91	
ベースラインからの変化 (16W)	-0.89	0.09	-0.98 (-1.21 ~ -0.74)
空腹時血糖			
ベースラインレベル, mmol/L	9.13	8.84	
ベースラインからの変化 (16W)	-1.59	0.10	-1.69 (-2.22 ~ -1.17)
自己測定血糖 (1日平均), mmol/L			
ベースラインレベル, mmol/L	9.46	9.18	
ベースラインからの変化 (16W)	-1.85	-0.14	-1.71 (-2.09 ~ -1.33)
食後血糖 (自己測定, 平均)			
ベースラインレベル, mmol/L	1.74	1.99	
ベースラインからの変化 (16W)	-1.58	-0.31	-1.27 (-1.64 ~ -0.91)
体重			
ベースライン, kg	97.53	96.75	
ベースラインからの変化 (16W)	-1.75	-0.24	-1.51 (-2.15 ~ -0.88)

に比べて、グリコシル化ヘモグロビンや空腹時血糖値などの値が良いことが分かります。プラセボ群よりもよい成績です。でもこれは良い効果を示すための作為的なやり方です。実験で求められていることは、プラセボより勝ることではなく、既存の他の糖尿病薬よりもすぐれていることを示すことなのです。このように、たとえばスルフォニル尿素剤とエキセナチドを比べた上で、エキセナチドが勝っていたというのであれば、それは有用な薬だということができるでしょうし、よりバランスのとれた研究だと言えるでしょう。しかし、他に有効な薬があるときに、プラセボとの比較を行うというのはフェアでないし、その薬を実力以上に良くみせかけるためということになります。

臨床試験のデザイン

重要なことは、臨床試験のデザインです。薬の研究を行うにあたって、私たちがとりうる方法としては、3種類の主要なデザインがあります。すなわち、ある薬が他の薬よりも優っていることを証明するための優越性試験、同等であることを示す同等性試験、および、劣っていないことを示す非劣性試験です。優越性の証明は、私たちが真に求めているものであり、きわめて明確ですが、同等性や非劣性は定性的な概念であり、定量的でないため、非常にあいまいな用語です。したがって、私たちは何をもって同等であるというのかを明確にする必要があります。同等であるということを確認するためには、ここに示すようにデルタ値で示す2つの線を引く必要があり、何をもって等しいと見なすがこれで示されるこ

とになるのです(図2)。これら二つの線の間隔が幅広くとられれば(Δ値を大きくとれば)、すべてが同等であるということになりますが、非常に狭くとりますと(訳注: Δ値を0に収めさせると)、優越性試験になります。したがって、同等性が示されたといっても、その効果は対照薬のそれに近いのかもしれないし、また大きく異なることもありうるわけで、Δ線をどのようにとるかで、異なる結果が得られることとなります。事実、表2でもお分かりのように、最後のコラムをみるとΔ値が16%で同等という試験もあれば、Δ値が63%で同等という試験もあります。ですから、ある試験においてΔ値63%で2つの薬が同等であるという結果であったとしても、それは同等性を求めているのではなく、差が見いだせなかったということなのです。これらの同等性試験の多くは、実のところ、対照薬と大きな差がないことを示すためにデザインされたものであり、科学的な理由よりはむしろ商業的な目的で行われたものなのです。

このようなことは、抗がん剤の分野を含めて、多くの分野で起こっています。表3に示したのは、欧州で承認された18種、21適応の抗がん剤です。うち12の適応に関しては、第2相試験だけで承認されています。この場合、第3相試験が行われていないのですから、その薬が有効かどうかは不確かです。他の9つの適応については第3相試験が行われていますが、優越性試験のデザインで行われているのは、その中の3適応だけでした。これは、21の適応のうち、既存治療法と比べて新薬がいかなる位置づけにあるかを明確に分かるように示したものは3つしかなかったということです。

約400件の臨床試験を分析してみたところ、その64%は既存薬との違いが50%未満であれば同等であると見なしており、84%は既存薬との違いが25%未満なら同等とみなすという約束で同等性を証明していました²⁾。このことは、同等性があると信じる理由はどこにもないということを示しています。なぜなら、先ほども私が述べたように、これらの8線はいかようにでも広くとることができるので何でも同等と見なせるからです。

ところで、これらのタイプの研究には倫理的問題があります。何の利益も期待できない研究に患者を従事させることになるからです。したがって、私の意見では、これらの同等性試験や非劣性試験

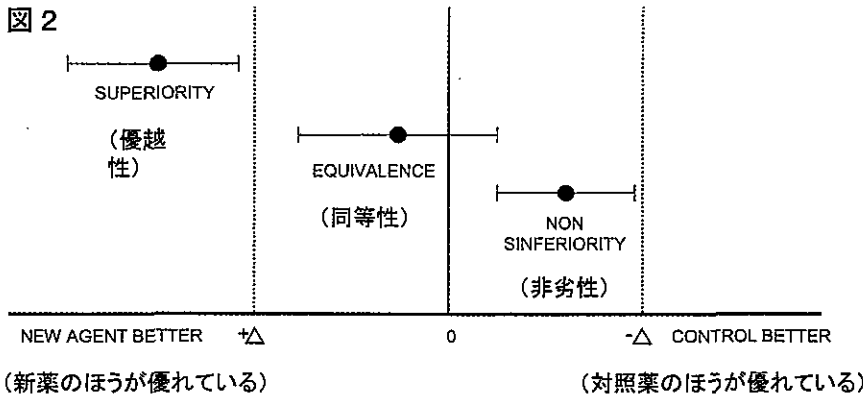


表2: SSRI、三環系抗うつ剤、プラセボの比較
同等性を示す試験で、最大限許容できる効果の差Δはどのように設定されたか

臨床試験	比較した薬/プラセボ	対象患者数	Δ
Feighner et al., 1989a	Fluvoxamine, Imipramine, PLO	86	40%
Feighner et al., 1989b	Fluoxetine, Imipramine, PLO	178	30%
March et al., 1990	Fluvoxamine, imipramine, PLO	54	63%
Dunbar et al., 1991	Paroxetine, imipramine, PLO	717	16%
Shrivastave et al., 1992	Paroxetine, imipramine, PLO	120	39%
Doogan & Langdon, 1994	Sertraline, dothiepine, PLO	308	23%

PLO = プラセボ

は禁止すべきです。これらの研究は患者の利益ではなく、商業上の利益のためだけに行われるのですから、実施すべきではないのです。

組込条件と除外条件

また、対象患者の組込基準、除外基準をみることも重要です。多くの場合、臨床試験の対象患者の選択は、当該薬が市場に出て最終的に投与されるような患者とは異なっています。妊娠可能年齢の女性や、小児、老年者はもちろん試験対象から外されますが、市場にできればこのような患者も投与対象となります。ここに約1万人を対象に行われた変形性関節症や関節リウマチの薬がありますが、その試験患者のうち65才以上は僅か200人、2.1%にすぎず、75~84才が14人、85才以上の患者は1人もいませんでした³⁾。関節炎は一般に年齢をとってからかかる病気ですから、これらの人々がこの薬を飲む可能性は大いにありうることです。つまり、この薬を飲むであろう患者が、その評価を行う試験対象として含まれていないのです。

同じ傾向は小児の患者についても見られます。小児に使用される薬の50%は本当の比較対照試験で検討されたことの

ないものです。もちろん、投与量は体重に基づいて決められますが、子供が小さな成人でないことは常に銘記すべきことです。子供は成長しつつある生き物ですから、成人と比べて異なる問題を抱えているのです。

不適切な対照

先に、私たちはプラセボが多用されていることを問題としましたが、対照薬を用いる場合、どんなタイプの薬を比較の対照とするかは、新薬の効果を確認する上でとても重要です。例えば、この研究では(図3)、臓器移植後の拒絶反応の治療に利用される2つの薬剤間で、さまざまな比較が行われました。これら薬剤の間には、多くの比較があるのがわかります。この図をみると、タクロリムスのほうがシクロスポリンよりも拒絶反応が少なく、ずっと良いようにみえます。しかし、私たちはシクロスポリンの作用は300~400ngの血中濃度で使用したときに有効であることを知っていますが、非常に注意深くみると、この研究では、シクロスポリンのレベルは300ng/ml未満であり、至適用量よりも少ない量が使われているのです。したがって、タクロリムスのほうがシクロスポリンより効果的

表3

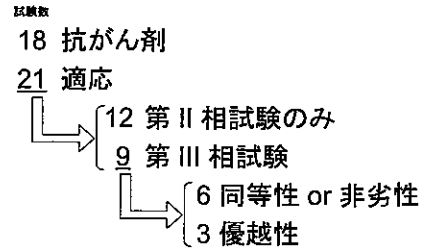
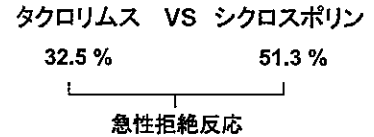


図3



急性拒絶反応の発生は、トラフレベル(最低血中濃度)が330 - 430 ng/mlに維持されたときに最低になる。

MARGREITER et al., 2002 TRIALでは、シクロスポリンのトラフレベルは < 300 ng/ml

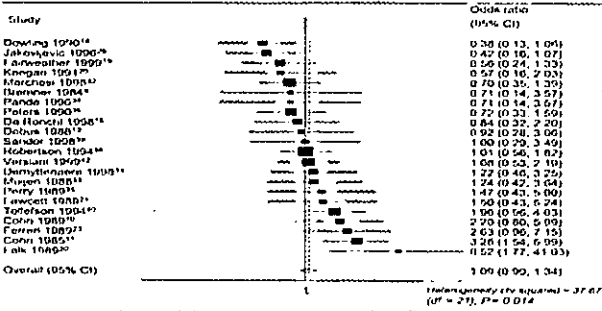
にみえたのは当然です。シクロスポリンが最大の効果を発揮するには、用量が少なすぎたのですから。

同じことは他の臨床試験でも起こっています。例えば、オランザピン、リスペリドン、クエチアピンなどの新しい抗精神病薬に関心もたれてきましたが、それは錐体外路系の副作用が少ないためです。しかしながら、その違いはしばしば比べた対照薬が非常に大用量である場合に確認されたものなのです。対照薬であるハロペリドールが12mgよりも多い用量で用いられれば、第2世代の抗精神病薬が優るのですが、12mg前後の用量のハロペリドールと比べると、その差は見られなくなります⁴⁾。事実、精神病に対する通常の治療法では、ハロペリドールの投与量は常に約12mgです。もちろん、より高用量を投与すれば、被検薬のほうが優ることになるでしょうが、これはいわば作為的に作り出された違いなのです。

同様に、これはいわゆる第2世代のシクロオキシゲナーゼ2, COXIBsやFecoxib等の非ステロイド抗炎症剤についても言えることです。これらの薬は胃腸への毒性が少ないので、従来の抗炎症剤よりも良いと考えられていますが、心毒性が強いのです。しかし、その心毒性の検出は、比較を行う対照薬に何をを選ぶかで異なるのです。もしジクロフェナクを対照に選べば、差はないように見えます⁵⁾。しかし、ナプロキセンなど別の非ステロイド抗炎症剤を使用すれば、第2世代のシクロオキシゲナーゼのほうが従来の抗炎症剤よりはるかに心臓毒性が強

図4：フルオキセチンの効果に関するランダム化試験の結果 (フルオキセチンが試験薬の場合)

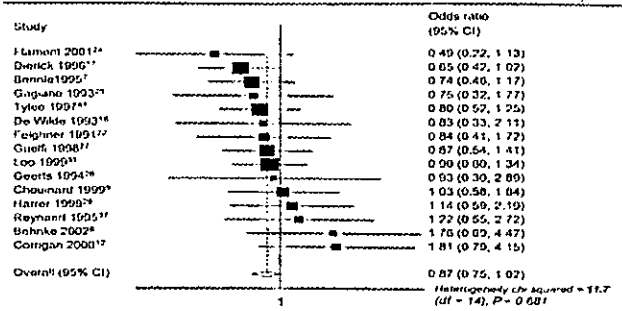
(訳注：複数の試験結果をまとめるとフルオキセチンの方が優れているという結果)



Barbui et al., 2004

図5：フルオキセチンの効果に関するランダム化試験の結果 (フルオキセチンが比較対照薬になった場合)

(訳注：複数の試験結果をまとめると試験薬のほうが優れているという結果)



Barbui et al., 2004

表4：27の適応症のもととなった、主な臨床試験での主要評価項目

	No.	%
全生存期間	2	7
無増悪期間/無病生存期間	11	41
反応率	13	48
その他 ^a	1	4

無増悪期間: TTP; 無病生存期間: PFS

^a 臨床試験としてのデータなく、非臨床研究、臨床研究の論文レビューに基づく承認

APOLONE et al., 2005

いこととなります。しかし、プラセボを使って比べてみると、心毒性がはるかに高いことがわかります。ですから、ことは何を対照とするかによって決まるのです。原則は最良の既存薬を対照薬とすべきだということであり、比較が公平に行われるように、それを至適用量と至適な治療計画で用いることなのです。このように、同じ薬であっても異なる作用があるようにみえるのは面白いことです。例えばフルオキセチンを例にとると、これはこの薬が発見された1990年頃のデータで、まだ治験段階のもので、当時のご覧のように対照薬よりもはるかに良い成績です⁶⁾(図4, 5)。しかし、このフルオキセチン自体が他の新薬の対照として使われるようになると、その効果はずっと低くなります。ですから、薬の効果は使用される場面によって変化するのは、つまり、試験薬(新薬)の段階では有効だが、対照薬として使われるようになると、効果は少なくなるのです。対照薬として使われるときには、試験薬として使われるときよりも少ない用量で用いられるからです。

サロゲート・エンドポイント

もう一つの考慮すべき重要な点は、臨

床試験で何を測定しようとしているかということであり、サロゲート(代理)エンドポイントとハード・エンドポイントを区別しなければなりません。代理エンドポイントは、通常生化学的パラメーターやハード・エンドポイントの成績を予測させるパラメーターが用いられます。しかし、患者にとって本当に重要なのはハード・エンドポイントであり、典型的にはQOLや罹患、死亡に関連するパラメーターです。私たちに関心があるのは、当該薬が与えられた生化学的パラメーターを変えるかどうかではないのですから、すべての臨床試験の最終目標はハード・エンドポイントであるべきです。私たちに興味があるのは、薬がQOLや死亡を変えるかどうかなのです。代理エンドポイントが常にハードエンドポイントを代表しているわけではありません。例えば、エンカイニドやフレカイニドなどの抗不整脈薬について言えば、これらの薬は不整脈を減らしますが、死亡のリスクを増大させます。HDL-コレステロールは心血管イベントの良い代理パラメーターであると考えられており、エストロゲンはHDL-コレステロールを増加させますが、エストロゲンを投与しても心血管イベントにはいかなる影響もみ

表5：1995年から2004年に欧州医薬品庁が審査した抗がん剤の評価：承認根拠となった48試験のまとめ

Clinical trial design (48 trials)	Type of end point (primary) (48 trials)	Difference in survival, when available (13 trials)
RCT 25	Survival 4	Range 0-3.7 months
SAT 19	Resp. rate 30	Mean 1.5 (months)
NC-RCT 4	TTP/PFS 14	Median 1.2 (months)

ランダム化試験=RCT; 単群試験=SAT; 非比較ランダム化試験=NC-RCT

無増悪期間: TTP; 無病生存期間: PFS

APOLONE et al., 2005

られません。トルセトラピブはHDL-コレステロールを増やす作用がある新薬ですが、実際は、死亡を増加させます。

したがって、私たちが薬の効果を見定めるためには、代理エンドポイントを重視すべきではないのです。例えば、糖尿病では、スルフォニル尿素を飲めばグリコシル化ヘモグロビンの減少、つまり糖尿病代理エンドポイントを減少させますが、心筋梗塞を増やします。新薬の1つ、ロジグリタゾンもグリコシル化ヘモグロビンを減らしますが、実際は心不全を増加させます。ですから、患者にとって重要なパラメーターを評価する必要があります。

抗がん剤についてもこれは真実です。これは、また、多数の抗がん剤について27の適応を検討した私たちの研究ですが⁷⁾、これらのうち全生存をエンドポイントとして評価が行われたのは僅か2例、7%にすぎなかったのです(表4)。抗がん剤にとって、生存を延長させることは重要なことですが、それが評価尺度となることは滅多にないのです。通常は、反応率や無増悪期間などの代理パラメーターに焦点が向けられ、抗がん剤の効果判定するのに最も重要な点は考慮されないことが多いのです。そしてま