臨床試験での不完全例の取り扱い

佐久間 昭^{*1} 楠 正^{*2}

(受付: 1984年4月19日)

Handling of Cases with Incomplete Data in Clinical Trials

Akira SAKUMA^{*1} and Tadashi KUSUNOKI^{*2}

^{*1} Department of Clinical Pharmacology, M. R. I., Tokyo Medical and Dental University

^{*2} Stat. Res. Sec., Development Department Takeda Chemical Ind., Ltd.

In randomized clinical trials it is rather rare that all the subjects who entered the trial complete the trial schedule given in the protocol, and more or less "incomplete cases", called exclusions, withdrawals and dropouts, often occur. How to deal with the data of these "incomplete cases" may critically influence the conclusion to be reached. The problem is particularly important in the clinical evaluation of new drugs. There is, however, no generally agreed upon, definite guideline for the handling of these "incomplete cases". Since they occur in various forms it seems that uniform handling of the problem is not appropriate and handling corresponding to the nature of each will be necessary. In this article we tentatively define terms which classify the "incomplete cases" and discuss practical and valid ways to deal with the problem depending upon the purpose and character of the study.

Key words : clinical trial, ineligibility, withdrawal, dropout

臨床試験では、試験治療を開始した対象のすべてが、計画通りに試験を終了することは稀であり、

脱落例、中止例などが多少とも生じる。これら試験計画の規定から外れた対象をどのように扱うかは薬効評価の現実問題であり、とくに比較試験ではその結論を左右することもある重要な問題である。これにどう対処するか、明確な指針として合意されたものはなく、一律に処理できる問題とも思わ

^{*1} 東京医科歯科大学・難研・臨床薬理部門

〒101 東京都千代田区神田駿河台2丁目3-10

^{*2} 武田薬品工業株式会社・企画開発本部・試験解析室

れない。ここでは、試験計画の規定から外れた対象を一括して「不完全例」と呼び、まず不完全例の内容を分類する用語を暫定的に整理する。次いで、試験の目的、性格に応じて不完全例のデータをどのように扱うのが実際のであり、妥当であるかを考察する。無作為割りつけによる比較試験を主に考えるが、必ずしもそれに限定しないで検討する。

不完全例の分類

不完全例を本質的に試験開始前の問題にかかわる不適格と、試験開始後の試験治療にかかわる違反、中止、脱落などに大別して整理してみる。本節の主題は、試験実施中に起きる計画とのずれを事実関係の上でどう整理するかであり、解析段階での採否の問題には触れず、これについては次節で検討する。「除外」という用語は、対象を試験に組み入れるときのことなのか、解析段階でのデータの扱いのことなのか明確ではなく、誤解を避けるため、ここでは使わない。なお、「不完全」の内容によっては、治療上、試験データの収集面、さらには試験をその後も適切に継続していく上で、一般的に注意すべき事項を伴うので、それらについてはここで触れる。

不適格 ineligibility：患者を試験に組み入れ、対象として登録するとき entry, registration の問題である。計画の対象選択基準に合致せず、除外基準に抵触する対象は、「不適格、資格なし」と呼び「適格」と区別する。その具体的な内容としては、診断の誤り、遅れ、診断根拠のデータ不備など、診断に直結するものと、年齢、病歴、合併症、前治療など、対象となる患者の背景特性に関係したものがあ。試験中に逐次対象として登録したもの inclusion, しなかったもの exclusion の内容の記録 patient log を作って検討していけば、その後の不適格例の発生防止や、最終的なデータ採否の検討に役立つであろう¹⁾²⁾。本来、選択基準や除外基準は、試験が安全に遂行され、医学的に妥当な結論がえられるように設定されるものであるから、試験計画において選択基準に多少の不明

確さがあっても、この基本に立ち、より具体的な上記の patient log を参照しつつ、その後の対象の組み入れを行えば、実際上の問題はほとんど解決されるであろう。

治療違反 deviation, violation：対象が選択ないし登録され、試験治療が開始されても、最後まで計画通りの治療がなされとは限らない。試験治療そのもの、薬その他の併用療法に関する試験計画との不一致を、ここでは「治療違反」とする。これは、担当医が合併症や悪化に対処するため、意識的に試験薬の用法、用量を変更したり、禁止されている併用療法を追加したりということのほか、無意識的な投薬ミスでも生じる。

治療不遵守 therapeutic incompliance：対象患者が規定通りに服薬しなかったり、投薬を受付けない場合である。単純な服薬忘れのほか、副作用の懸念、悪化や完治の判断など、試験と関係した理由によっても発生するので、事後的な調査を行い、できるだけ理由を明らかにすることが大切である。患者の単純な思いすごしであることもあるし、思いすごしではなくて臨床的にみても明らかな副作用であるということもある。観察期、試行期を設け、非協力的な患者を対象としない考えもあるが、このことは試験目的と対象集団規定、あるいはえられる結論の性格との関係で、計画段階で決めるべきであろう³⁾⁴⁾。

中止 withdrawal：担当医の医学的判断により、対象患者の試験治療を意識的に、所定の試験期間を完了しないで中止する場合である。理由としては、副作用、悪化、治癒、併発症発生による治療変更、試験治療開始後の不適格の判明などがあり、すべてが計画違反になる訳ではない。用法、用量の変更、併用治療の実施などの治療違反を経て中止に至る場合もある。中止の時点では、検査、観察、記録、評価などを可能な限り実施し、中止の内容とともに正確に記録すべきであるが、副作用、悪化、治癒などの理由で中止となった場合、これらは試験治療の重要な結末 outcome, event であるから、とくに記録が大切である。

不適格例のうち、診断の変更に關しては、患者

Tab. 1 完全・不完全例の分類とその実例

	不完全の問題内容	得られるデータの問題			実例(*) (総計1745例) における 例数と(%)
	開始 ⇒ 途中 ⇒ 終了	投薬・服薬 の条件	測定の方法・時点	欠測・歯抜け	
問題なし	[適格] ⇒ 問題なし ⇒ 完了	満足	満足	なし	1321 (75.7)
担当医	[不適格] ⇒ 問題なし ⇒ 完了 別疾患, 診断基準不適, 観察時期症状・検査データ不適, 観察時期薬剤, 合併症, 年齢	満足	満足	なし	77 (4.4)
担当医	[治療違反] ⇒ 完了 試験薬用法用量違反, 休薬, 併用薬・併用療法違反	不備	満足	なし	27 (1.5)
	[中止] 副作用, 無効, 悪化, 治癒, 合併症併発, 担当医交代, 途中で調査期間満了	満足	そこまで 満足	以後全欠測	194 (11.1)
	→ [治療違反に続き中止]	不備	そこまで 満足	以後全欠測	
対象患者	[治療不遵守] ⇒ 完了 服薬時期・回数・量の違反, 服薬の不規則・忘れ	不備	満足	なし	7 (0.4)
	[脱落・追跡不能] 来院せず(転居, 転医, 出張, 改善, 悪化, 副作用などによる), 事故(*)	満足	そこまで 満足	以後全欠測	99 (5.7)
	→ [治療不遵守に続き脱落]	不備	そこまで 満足	以後全欠測	
担当医	[観測不備] ⇒ 完了 測定・評価の時期・方法の違反, 測定・評価忘れ, 調査表紛失	満足	不備	なし・歯抜け 全欠測	6 (0.3)
対象患者	[観測不備] ⇒ 完了 来院不規則, 測定・評価に非協力	不備	不備	歯抜け	14 (0.8)
	→ [その後 脱落]	不備	不備	以後全欠測	

*実例: 「医学のあゆみ」127巻に掲載された比較臨床試験6報から問題の内容例と、
6報の症例数の合計をとった。

合計1745 (100)

- (1) 麻酔前投薬(術前夜睡眠), 1回投薬
- (2) 胃潰瘍, 試験治療6週
- (3) 本態性高血圧症, 観察4週, 試験治療12週(単独療法, 併用療法)
- (4) 慢性関節リウマチ, 観察4週, 試験治療24週
- (5) 原発開放隅角緑内障, 高眼圧症, 観察4週, 試験治療8週
- (6) TIA, 試験治療12か月以上

*事故: 麻酔前投薬(術前夜睡眠)の試験における「投与〇分(時間)後の排便」など。

の治療を優先させ、試験治療を中止、変更するなどの適切な処置が必要となる。軽度の検査値異常の発見、年齢制限における若干のずれなどでは、特に中止、変更の必要を認めず、試験治療を続行して完了させるべきであろう。

予期しない重篤な副作用の発見、基本的な計画の変更などによる試験全体の中止に伴う中止を「試験管理上の中止」administrative withdrawalということがある。

脱落 dropout, 追跡不能 loss to follow-up: 対象患者の転居、多忙、性格など、試験治療と直接関係しない理由により、検査、観察、評価の実施が打ち切られた場合である⁴⁵⁾。多くは来院しない状態となり、試験と無関係の死亡もこれに含まれる。試験と関係した理由の治療不遵守があり、来院しなくなる場合もあるが、ここではそれも「脱落」とする³⁾。

診療予定日に患者が来院しない場合には、早急に問い合わせを行い、服薬状況、副作用の有無、症状の経過などについて調査しておくことが望ましい²⁾。脱落と追跡不能とはほぼ同義に使われることもあるが、追跡調査してもその後の情報が全くえられないものを「追跡不能」とするのが自然と思われる。中止、脱落をあわせて「未完了」premature exitと呼ぶこともある^{34),5)}。

観測不備: 検査、観察、評価の方法や時期、記載方法に関する違反をここでは「観測不備」とする。担当医の管理ミス、休日、学会と来院予定日との重なり、対象患者の来院日のずれなどによって生ずる。来院日のずれは治療不遵守と同時に発生する場合が多く、その理由には、試験治療と無関係の単なる都合である場合と、副作用、悪化、治癒の判断など、試験治療と関係する場合とが考えられる。

実例: 以上、不適格、治療違反、治療不遵守、中止、脱落、観測不備を分類したが、実例にこの分類をあてはめ、それぞれの内容を整理するとTab. 1 のようになる。当然一例で二重三重の問題が発生する場合もある。不完全例ではTab. 1 に示す通り、データのえられた条件に問題があっ

たり、欠測が出たりする。これに伴い「判定不能」という場合も発生しよう。欠測の程度には一部の「歯抜け」から、ある時点以後のすべてが欠測、対照観察期の欠測、さらに試験開始直後からのすべてが欠測と、さまざまなタイプがある。

実例でみると、解析に含めないものを「除外」、不完全例でも解析に含めるものを「脱落」、あるいは「脱落・中止」とする用語が多く使われ、事実関係の分類と解析への採否との混同がみられるが、「解析対象」、「採用例」という用語で両者を区別している例もある。ここでは、発生した事実関係の分類を上記のように定め、次節で述べる解析段階での採用・非採用と区別する。

解析段階での取扱い

不完全例を解析段階でどう扱うかは、その発生が試験の初期であるか後期であるか、理由が試験治療に直結しているかないか、担当医側にあるのか対象側にあるのか、内容的に全般的か部分的かなど、前節記載の観点から検討される。データ採否の方針は、試験の性格、試験治療の長さ、試験の規模、二重盲検法であるか否か、などによって変わってくるであろう。

試験計画の設定ならびに解釈を厳格にするか柔軟にするかは、試験の目的、性格に依存しており、解析段階において不完全例のデータを採用するか否かの判断も、この点に密接に関連している。臨床試験には、研究的、実験室的、説明的 explanatory な目的・性格をもつものと、治療的、野外的、実践的 pragmatic なものとがあり、それによってデータの妥当な採否は変わってくる¹⁶⁾。採否の基準は、試験が終了してから事後的に不完全例の様子をみて決めるのではなく、事前に定められるべきものであるから、不完全例の多寡は、その試験の質的レベルを示唆するであろう。

研究的試験と治療的試験: 研究的な試験の場合、臨床試験の場をできるだけ等質な土俵に求めるが、治療的試験ではありのままの診療現場を求める。この両極端の立場において、想定している母集団、求めている結論は明らかに異なっている。

Tab. 2 不完全例のデータの取り扱い（研究的⇒治療的）

種 類	有効性	安全性	有用性	例
不 適 格	×⇒○	○	×⇒○	診断の誤り
治療違反	×⇒○	×⇒○	×⇒○	併用薬の追加
治療不遵守	×⇒○	×⇒○	×⇒○	途中から服薬不規則
中 止	×⇒○	○	○	初期に副作用
脱 落	○	○	○	後期事故死
観測不備	×⇒○	○	×⇒○	観測時期かなりのずれ

注) ×：非採用，○：採用を示す。

研究的⇒治療的：研究的試験と治療的試験でのデータ取扱いの違いを示す。

比較試験でも、研究的な場合には、ある意味での薬理学的な対比の検索に力点があり、規定通り服薬され、規定通り収集されたデータを前提として、目的とする薬効の強さなり特徴を研究する。一方、治療的な比較試験では、とりあげた試験治療のうち、どれがより選択されるべきかの意志決定に力点があり、試験に組み入れられた対象のすべてについて、試験中に起きるいろいろな事実を含めた比較を指向する¹⁾。通常の臨床試験は、おそらくこれらの中間的な性格をもつであろう。つまり、不完全となった時点、理由を吟味し、対象を一個体全体として取り扱うよりも、データの種類に応じて、たとえば有効性については非採用であるが安全性については採用する取り扱いがより現実的と考えられ、Tab. 2 のようになるであろう。

不適格：二重盲検下の比較試験では、診断の誤りと、その対象にどの試験治療が割りつけられたかとは無関係と考えられ、研究的試験では別疾患の対象にかかわるデータは採用できないという考えが成立つ²⁾。一方、治療的試験では、診療現場で試験治療が開始されれば、不適格例であっても、登録された対象すべてを解析に採用することも考えられる。

診断の誤りと年齢の僅かなずれとを一律に扱うことは不合理で、研究的試験でもこうした問題では非採用とはしない扱いが考えられる。

腎障害のある対象には副作用の恐れがあり、あらかじめ除外規定に明記されているにもかかわらず、

ず、この種の対象が試験に組み込まれた際には、副作用の発現の有無とも関係して特別な扱いが必要になろう。

中止、脱落の理由と関係したデータ：副作用の発現、悪化、治癒など、中止、脱落に至った理由が明らかに試験と関係していれば、データ解析に採用しなければならない³⁾。試験初期の副作用中止では、研究的試験の立場からは有効性評価は無理としても、安全性は勿論、有用性の評価も採用可能であろう。治療的試験の立場からは、判定不能、あるいは無効の評価を与えて有効性の解析に採用してもおかしくない。

対象患者の判断による副作用の発現による脱落でも、事後的に追跡して医学的に妥当であれば、中止に準じた取り扱いになろう。患者が試験治療中に転倒して骨折し、通院不能で脱落した場合、単純な脱落ではなく、使用薬剤がもつ中枢抑制作用、あるいは骨代謝への影響による副作用である可能性がある。交通事故による打ち切りでもなどでも、同様の考察が必要となる。生命表解析では、死亡原因が試験と明らかに無関係であると認められない限り、打ち切りとしない立場もある。

治療違反、治療不遵守および観測不備：研究的試験では、一般に治療違反や治療不遵守に関係したデータはできるだけ採用しないし、検査、観察、評価の時点の常識的な範囲を越えるずれも問題にする。これに対し、治療的試験では、データの採用範囲はかなり広くなるであろう。たとえば有効

率を考えると、研究的試験では、適格な対象、しかも計画通りに試験治療を完了した対象だけを厳密に選んで分子、分母とする取り扱いが設定される。一方、治療的な試験であっても、治療違反、治療不遵守などの発生頻度は日常診療に比べれば少ないであろうが、これらは診療現場ではありうることであり、違反とはせず、試験治療がなされた対象は、その長短には無関係に採用し、たとえば有効としての分子になりうる可能性はなくても分母に加えるべきであるという主張が多い¹¹⁻⁸⁾。また、2通り以上の分母を設定し、反応率を出しておくことを推奨する考えもある⁹⁾。こうした観点からは、比較という枠を離れ、特定の薬についての有効率などを議論するに当たっては、十分に注意することが必要になる。逆に言えば、有効率などの一人歩きには、かなり危険を伴うと考えるべきであろう。

欠測：中間時点の欠測については、対象患者の状態が安定していれば、前後のデータから類推して埋めてもよいが、そのことを報文に明記しなければならないし、常識的範囲に留めるべきであろう¹⁾。中止、脱落、追跡不能となった後の欠測、あるいは判定不能のデータについては、治療的試験では有効率などの分母に含めるのが適切であるが、研究的試験では採用しない解析が考えられる。生命表分析では、これらを中止・脱落に伴う打ち切りデータとすることができ³⁹⁾、欠測を解析の枠外におくとしても、そのことを図表に明示するといった配慮が必要な場合もある。

比較試験での問題：比較試験では、採否の決定が試験結果に偏りをもたらす可能性に注意しなければならない。とくに封筒法など非盲検下の比較試験では、どの対象にどの治療が割りつけられているかがわかっているため、比較にあたって偏りが生じないように扱うには、特別な工夫を要するであろう。一般的に言えば、各種の不完全例を除くことで試験結果のノイズが減少し、感度のよい結論が期待できるが、このことは例数の減少という逆方向の影響を生むことになる¹⁰⁾。非採用例が多ければ、はじめの無作為化の意義が失われると

の議論もある¹³³⁴⁾。また、対象の均質化により、結論をあてはめるべき母集団の性格はより明確にされるが、治療的にみれば適用範囲が狭められることにもなる¹⁰⁾。

考 察

臨床試験における除外例、中止例、脱落例などの用語は、計画通りに試験を完了しなかった「不完全例」の分類として用いられるが、一面解析でのデータの「採否」の問題と絡み、用語と内容が混乱した現状がみられる。そこで我々は、まず、試験実施中に計画とのずれが発生する内容を用語の上で整理し、次いでデータ解析で不完全例をどう扱うべきかを検討することとした。

ここに定めた分類では、一般に試験計画とは「どんな対象にどんな試験治療を行ってどの時点で何を評価するのか」を骨格とするものであるとの考えを基本におき、不適格、治療違反、治療不遵守、中止、脱落、観測不備を区分した。「除外」は、対象を組み入れる時点と解析段階の両方の問題であり、まぎらわしいので意識的に使用を控えたが、「解析除外」という使い方があってもよい。臨床試験には、試験治療の開始を起点とし、時点を決めて観測、評価を行うように計画されたものが多いが、そうではなく、治癒、死亡、再発などの治療の結末 outcome, event がいつ発生するかを追う形の計画もある。ここでは、暗に前者のタイプの試験を主に考えたが、ここでいう中止例には治療の結末を迎えたものが多く、「不完全例」というよりは「完全例」とすべきであろう。

次に、不完全例が過去どのように受け止められてきたかを振り返りつつ、不完全例の取り扱いについて考察する。従来の臨床試験は、研究的な内容の妥当性に強い関心をもつ半面、実施可能性への関心は、担当医についても対象患者についても十分ではなかったふしがある。その一つの現れとして、やや無理な試験計画のもとに試験が実施され、不完全例がかなり多数発生しても、それに伴う条件不備のデータや欠測を除外することは、試験の質を向上させる当然の手段として受け止めら

れてきた。

比較試験についてみると、ある程度研究的なところに目標をしばって実施し、結論の適用範囲はやや狭いとしても、不完全例のデータは採用せず、一例一例のデータは完備した状態で試験結果を評価すべきであるという考えが一般的であった。また、無作為割りつけに頼り、不完全例が比較に関してバランスよく発生していれば、それらを除外しても問題ないとしてきたふしがある。しかし、対象の数は有限であり、この原理が常に働いてくれるとは限らず、解析に採用しない不完全例が増えれば無作為化はくずれ、比較の結果に偏りが介入する恐れが常にある。

一方、数カ月から一年以上に及ぶ長期臨床試験が次第に多くなり、不完全例の内容として、薬剤の有効性、安全性、有用性との関連を否定できない場合の発生が増加し、完全例にしばった評価の不合理が認識されてきた。これは長期試験や比較試験に限った問題ではなく、治療的な考え方の重視に結びついてきたが、試験の目的・性格に応じて、ときに研究的であり、ときに治療的であるという使い分けがあってもよいように思われる。

新薬の開発について考えると、開発初期の臨床試験では、少数例であっても、計画通りに実施可能な対象にしばって研究的な試験とし、適格例、かつ治療不遵守や観測不備が少ないデータによって薬剤の性能を評価する方向をとり、次第に対象を広げて治療的試験による有用性評価に向かうのが自然と思われる。研究的といっても、いずれ現場で表現される薬剤の性質を推定するためのものである。現場での参考情報としては、治療的試験からえられる情報と、研究的な純粋な場で明確にされた情報の両方が必要であるが、ただ一つの試験によってあらゆる目的に答えを出すことは無理である。

おそらく、この辺の認識が不徹底であったことが、常に研究的な試験計画を用意し、結果的に生じた多数の不完全例の処理に苦慮し、「せっかくの一例」であるからとして解析に加え、事後的に計画を緩めるという習慣を定着させることになっ

たのであろう。この苦しい状況を打開する手段として、解析をより研究的、より治療的の二本立てとすることもあながち誤りとはいえないが、結果の解釈には試験の性格を振り返る必要があり、両方の結果があまりにも不一致であれば、結論は保留されるべきであろう。

今後は、対象患者の組み入れ、試験の実施、採否決定、解析について、治療的試験では広く、研究的試験では目的にあわせて限定的に、という一貫した態度が必要になろう。試験が治療的か研究の性格を明らかにし、それぞれに適した実施可能な計画書を用意すべきであろう。それぞれで計画から解析に至る一貫した態度をもてば、いずれにしても不完全例は少ないという状況の実現が可能ではないだろうか。

不完全例の取り扱い、一つの正しい答えを出せない困難な問題ではあるが、試験の性格による考え方の違いを中心にして、議論の土俵を整理することを試みた。不完全例に伴う「不明」、「判定不能」を含めて評価する必要があるとして、どのような方法がよいのかは困難な、重要な問題である。この点について、本稿では系統的な検討に至っていないが、今後整理する必要がある。

文 献

- 1) Armitage, P. : Exclusions, losses to follow-up, and withdrawals in clinical trials. Shapiro, S. H. and Louis, T. A. (Eds.) : Clinical trials. Marcel Dekker, N. Y. and Basel, pp. 99-114 (1983).
- 2) Peto, P., Pike, M. C., Armitage, P., et al. : Design and analysis of randomized clinical trials requiring prolonged observation of each patient. B. J. Cancer, 34 : 585-612 (1976).
- 3) Blum, A. L. : Clinical evaluation of success. Tygstrup, N., Lachin, J. M. and Juhl, E. (Eds.) : The randomized clinical trials and therapeutic decisions, Marcel Dekker, N. Y. and Basel, pp. 59-73 (1982).
- 4) Lachin, J. M. : Statistical elements of the randomized clinical trials. Tygstrup, N., Lachin, J. M. and Juhl, E. (Eds.) : The randomized clinical trials and therapeutic decisions, Marcel Dekker, N. Y. and Basel, pp. 77-103

- (1982).
- 5) Juhl, E. : Clinical elements of the randomized clinical trials. Tygstrup, N., Lachin, J. M. and Juhl, E. (Eds.) : The randomized clinical trials and therapeutic decisions, Marcel Dekker, N. Y. and Basel, pp. 33-41 (1982).
 - 6) Sackett, D. L. and Gent, M. : Controversy in counting and attributing events in clinical trials. N. Engl. J. Med., 301 : 1410-1412 (1979).
 - 7) Temple, R. and Pledger, G. W. : The FDA's critique of the Anturane reinfarction trial. N. Engl. J. Med., 303 : 1488-1492 (1980).
 - 8) 内藤周幸 : 臨床治験における問題点, 臨床評価, 11 : 253-263 (1983).
 - 9) WHO : WHO Handbook for Reporting Results of Cancer Treatment. (1979). (高谷 治, 市川平三郎 訳 : WHO 癌治療結果報告規準. 金原出版 (1981).)
 - 10) Blum, A. L. : Principles for selection and exclusion. Tygstrup, N., Lachin, J. M. and Juhl, E. (Eds.) : The randomized clinical trials and therapeutic decisions, Marcel Dekker, N. Y. and Basel, pp. 43-57 (1982).