

# 第2章

## 治療開発のための研究1：臨床試験



Introduction to Clinical Research



## 本講義の内容

- **治療開発の概要**
  - 「相 phase」の概念
- **研究デザイン**
  - 基本は比較
  - エンドポイント
  - 研究デザイン 3つの基本要素
- **品質管理・品質保証**
  - データマネジメント、モニタリング、監査
- **臨床試験に必要な組織**
  - 3つの基本要素



Introduction to Clinical Research

2

本講義では、治療開発のための研究として、臨床試験について説明していきます。

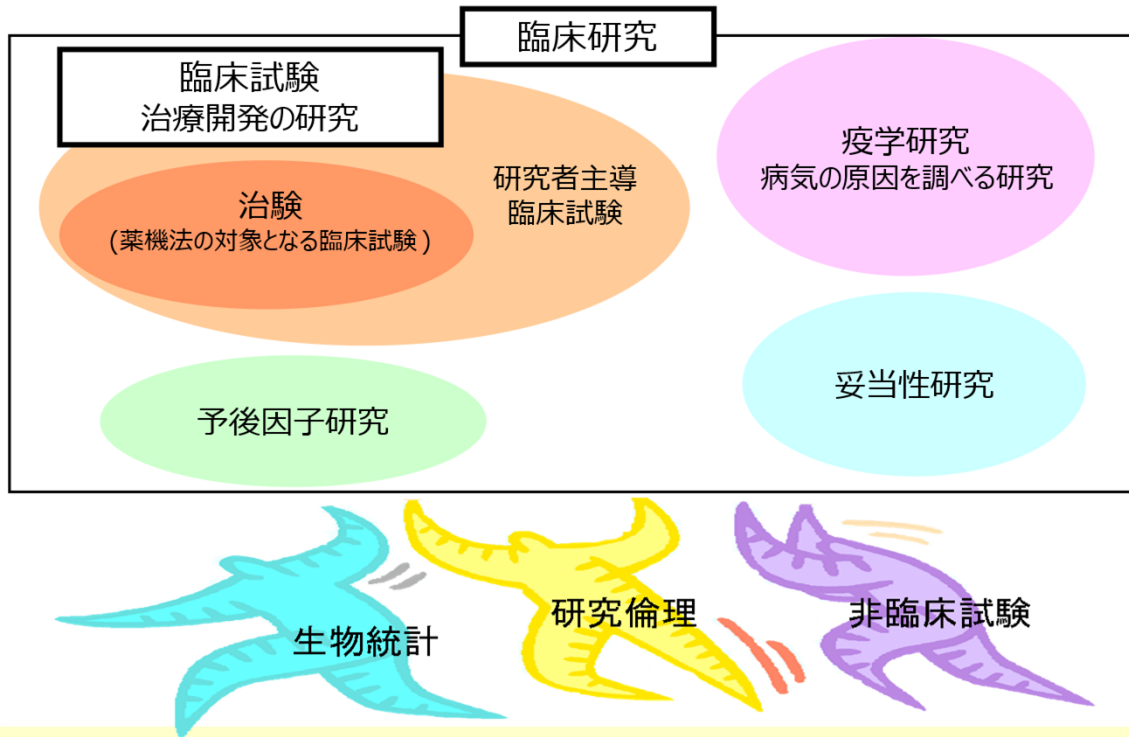
内容として、4つのポイントが挙げられます。

- ・治療開発の段階、「相/phase」という概念
- ・正しい結果を導くための研究デザインの重要性
- ・データの品質管理と品質保証を行い、結果の正しさを保証
- ・研究を行うために必要となる機能や組織

臨床試験の方法論は治療の開発だけでなく、診断や予防、ケアなど技術開発すべてに通じるものです。本講義では、治療法の開発を例として取り上げますが、自分の分野にあてはめて応用してください。

# 臨床研究とそれを支えるもの

**臨床試験**とは、あらかじめ綿密に立てられた計画に沿って、人に対して、多くの場合は患者に対して何らかの介入を行い、臨床的に未解明なことを調べる研究方法



臨床試験とは、あらかじめ綿密に立てられた計画に沿って、人に対して、多くの場合は患者に対して何らかの介入を行い、臨床的に未解明なことを調べる研究のことをいいます。

臨床試験のうち、国から新薬としての承認を得るためのデータを集めることを目的に行う臨床試験を、特別に「治験」と呼んでいます。

1984年に米国のポコック (Stuart J. Pocock) が書いた臨床試験の教科書「Clinical Trials」には、臨床試験は下記のように定義されています。

「患者を用いて行われ、かつ、ある特定の医学的条件に合致する将来の患者に対して最適な治療法を明らかにすべく企図された『計画的実験』」

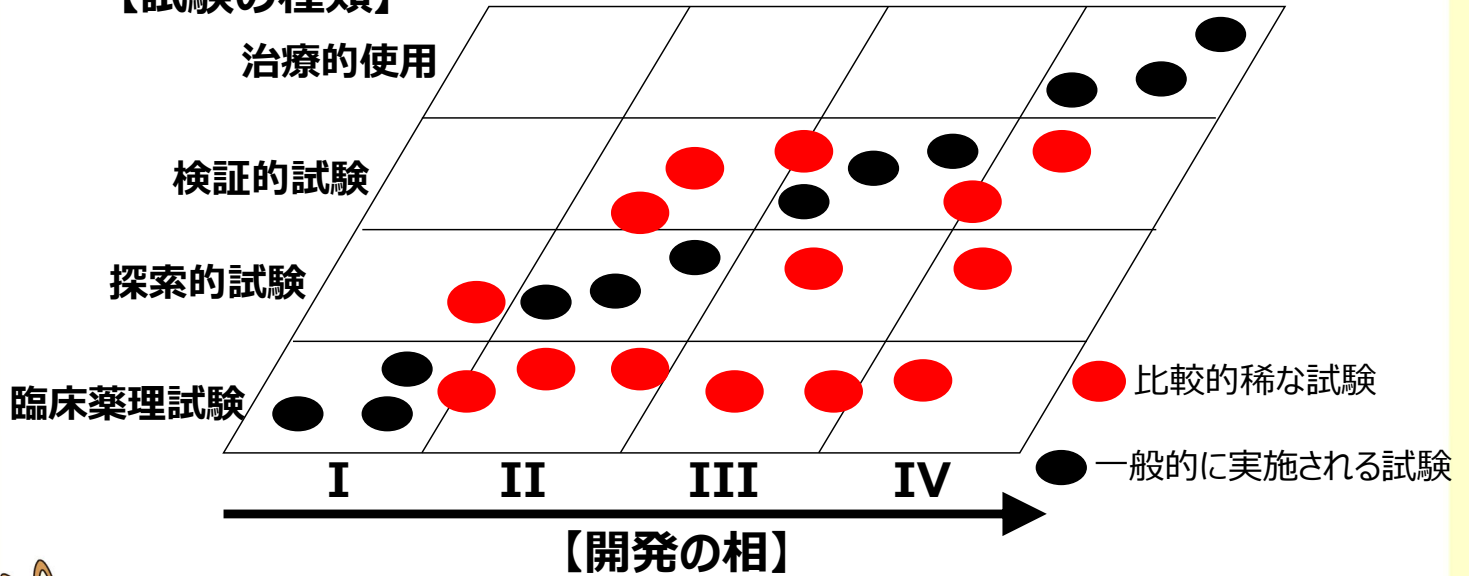
ここであえて実験といっているのは、臨床試験で行われる医療は日常臨床で行われる医療とは異なり、未解明なこと、すなわち‘実験的要素’を含むからです。実験的なレベルは臨床試験の内容により異なりますが、いずれの場合も安全に、かつ正しい結果を得られるよう綿密な計画を立てて実施し、段階を追って治療開発を進める必要があります。

では、まずは治療開発の概要として、「相/phase」の概念について説明していきます。

# 治療開発の概要 -開発の相 (phase) と試験の種類 (type) -

開発の相は治療開発の段階を表すが、相の概念は説明上のものであり、**開発の相と試験の種類は必ずしも1:1対応しない**

## 【試験の種類】



ICH-E8「臨床試験の一般指針」(<https://www.pmda.go.jp/files/000156372.pdf>) (アクセス日: 2023/2/16)  
2022/12/23通知日の改正版にはこの図は掲載されていないが、末尾の補遺に旧指針と同様の趣旨の説明や表が掲載されている



Introduction to Clinical Research

新しい有用な治療法の開発は、1つの臨床試験を行うだけでは完成しません。有望と考える新しい治療法がすでに存在している治療法に比べて安全で有効であるかどうかは、計画的に段階を追って臨床試験を行うことにより科学的根拠を集め、証明します。

この治療開発の段階のことを便宜的に開発の「相 ( phase ) 」と呼びます。医薬品の臨床開発は 4 つの相 ( 第 1 相-第 4 相 ) で成り立つといわれていますが、相の概念は説明上のものであり、要件ではないこと、相を明確に分けることができない場合があること、複数の相にまたがる試験があり得ることを理解することが重要です。

また、治療開発の段階を、臨床薬理試験や探索的試験、検証的試験など、目的に応じた分類「試験の種類 ( type ) 」で分ける場合もあります。

図では横軸には開発の相、縦軸に試験の種類をとっています。ここでは右側に位置するほど開発の相は後期となり、上側に位置するほど試験の種類は検証的なものとなります。

抗がん剤では、第 I 相(phase I)試験では臨床薬理試験を実施し、第 II 相(phase II)試験では探索的試験、第 III 相(phase III)では検証的試験を行うことが典型的です。このように、開発の相と試験の種類が1対1対応する領域も多いのですが、早期のphaseで実施されることの多い臨床薬理試験を後期のphaseにおいても実施する場合があります。重要なことは、実際に治療として使用されるまでには、より実験的な段階から臨床に近い場合へと順を追って試験をすること、その中で様々な点を確認する必要があるということです。

## 各phaseの概要（がんの治療開発を例に）

	相	一般的な対象者数と試験の目的
早期	Phase I	<ul style="list-style-type: none"> <li>・ 15-50人</li> <li>・ 安全な用量を見つける・薬剤の投与方法を決める</li> <li>・ 薬剤が人体にどのように作用するかを観察する</li> </ul>
	Phase II	<ul style="list-style-type: none"> <li>・ 100人未満</li> <li>・ 薬剤や治療法が特定のがんに有効性を持つかどうかを判定する</li> <li>・ 薬剤や治療法が人体にどのように作用するかを見る</li> </ul>
	Phase III	<ul style="list-style-type: none"> <li>・ 100人～数千人</li> <li>・ 新しい薬剤や治療法（新しい使い方）を現在の標準治療と比較する</li> </ul>
後期	Phase IV	<ul style="list-style-type: none"> <li>・ 数百人～数千人</li> <li>・ 新しい治療法の長期の安全性と有効性を評価する</li> </ul>



<https://www.cancer.gov/about-cancer/treatment/clinical-trials/what-are-trials/phases> より作成（アクセス日：2023/2/16）

Introduction to Clinical Research

がんの治療開発を例として、より詳しく各phaseの目的や特徴を説明します。

表では、上側に位置するほど早期開発段階、下側に位置するほど後期開発段階を意味します。

一般的には、phase I では安全性試験や、用量・用法を検討する試験、体内動態（吸収・分布・代謝・排泄）を評価する臨床薬理試験が行われ、phase II では有効性や安全性の関係を確認する探索的試験、phase III では標準治療と比較する検証的試験が実施されます。続く市販後に行うphase IVでは、治療的使用を評価する試験あるいは長期的な副作用の評価を目的とした調査が実施されます。

なお、通常phase I は健康なボランティアの人を対象に行いますが、抗がん剤は副作用として毒性がある可能性が高いため、健康な人を対象にして臨床試験を行うことができず、phase I から患者さんを対象として行うことが一般的です。

# 臨床試験はスクリーニング

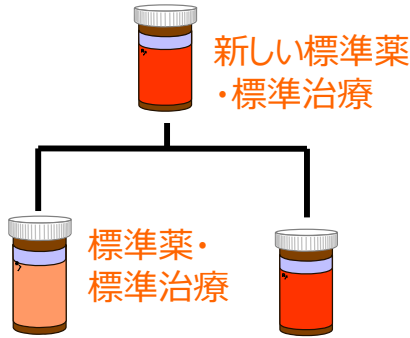
— 抗がん剤開発を例に —

Phase I → Phase IIIで標準治療よりも優れる開発成功確率は医薬品全体で

わずか **13%** 程度しかない

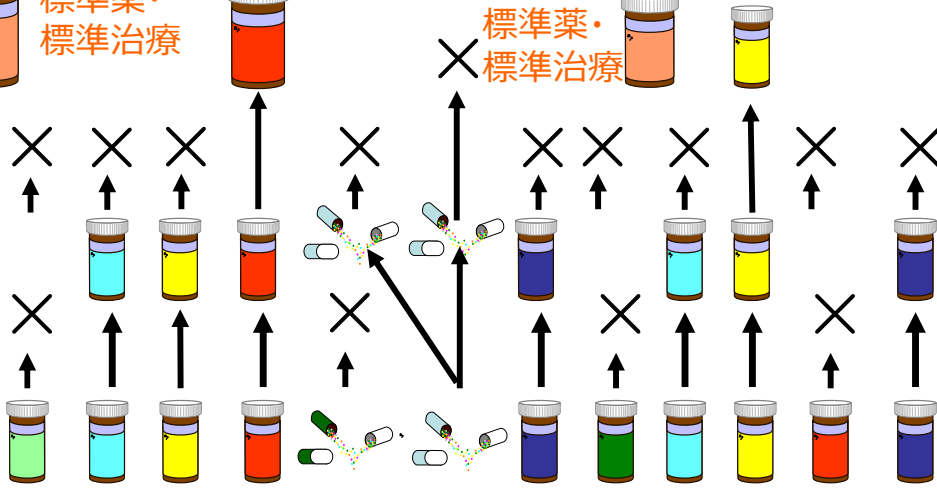
## Phase III

・総合評価による決勝戦



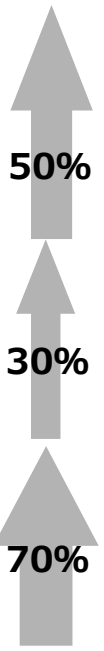
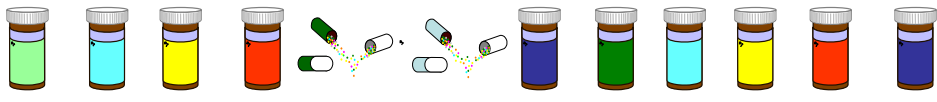
## Phase II

・有効性で screening  
・がん種別の展開



## Phase I

・毒性で足切り  
・Phase IIでの推奨用量決定



臨床試験創薬の成功確率のデータは、医薬産業政策研究所、「創薬の成功確率分析-臨床試験に焦点を当てて-」、製薬研ニュースNo.66より  
[https://www.jpma.or.jp/opir/news/066/bvd7r10000003ft-att/66\\_all.pdf](https://www.jpma.or.jp/opir/news/066/bvd7r10000003ft-att/66_all.pdf) (アクセス日: 2023/2/16)

Introduction to Clinical Research

非臨床試験を経てphase I の段階にきた薬の候補がすべてphase II、IIIへと進み、治療開発が順調にいけば、よい薬がどんどん世の中に出てくるわけですが、残念ながらそれ程うまくはいきません。

段階を追って進めていくうちに、予想していたほど効果がなかったり、毒性が強いことがわかれば、開発は中止になります。

また、効果があっても毒性が許容範囲であっても、最後は標準治療と比較する臨床試験を行って標準治療よりも有用であることが証明できなければ、新しい治療法の開発に成功したとはいえません。

すなわち、臨床試験は新しい薬を世に出すためのスクリーニングのプロセスといえます。

図は、抗がん剤を例として、開発のイメージを示しています。

抗がん剤の候補は

phase I からphase II へ進む段階で、30%が開発中止となり

phase II からphase III へ進む段階でさらに、3つに1つのみが生き残り

phase III の比較臨床試験では、2つに1つのみが標準治療よりも優れる

といった確率で、新薬開発は行われています。

医薬産業政策研究所の調査によると抗がん剤の候補がphase I まで進んだ場合のその後の開発成功確率は10%程度であり、医薬品全体での確率でも13%と低いものでした。なお、最も成功確率が低い領域は、精神系の治療薬でした。

## 一般薬開発と抗がん剤開発の比較

- Phase Iで一般薬は**健康人**、抗がん剤は**患者**を対象
- 承認申請は、一般薬は**phase III**の後、抗がん剤は**phase II**の後もあり得る

	一般薬		抗がん剤	
	対象	評価項目など	対象	評価項目など
Phase I	健康なボランティア	<ul style="list-style-type: none"> <li>用量探索と毒性評価</li> </ul>	患者	<ul style="list-style-type: none"> <li>用量<b>決定</b>と毒性評価</li> <li>探索的な有効性評価</li> </ul>
Phase II	患者	<ul style="list-style-type: none"> <li>用量決定と毒性評価</li> <li>代替りのエンドポイントによる有効性評価</li> </ul>	患者	<ul style="list-style-type: none"> <li>毒性評価</li> <li>代替りのエンドポイントによる有効性評価 <b>承認申請</b></li> </ul>
Phase III	患者	<ul style="list-style-type: none"> <li>用量決定と毒性評価</li> <li>代替りのエンドポイントによる有効性評価 <b>承認申請</b></li> </ul>	患者	<ul style="list-style-type: none"> <li>毒性評価</li> <li>真のエンドポイントによる有効性評価 <b>承認申請</b></li> </ul>
Phase IV	患者	<ul style="list-style-type: none"> <li>毒性評価</li> </ul>	患者	<ul style="list-style-type: none"> <li>毒性評価</li> </ul>



この表では、phaseごとの特徴について、一般薬と抗がん剤を比較して示しています。

Phase I の対象が、一般薬では健康ボランティア、抗がん剤では患者さんであるといった違いは先に述べましたが、両者の間の開発の大きな違いはもう1つあります。それは、一般薬はphase IIIまで治験として行った後、新薬としての承認申請を行うのが一般的ですが、抗がん剤の中にはphase IIまでの治験のデータをもって申請することができるものがある、という点です。

どのような場合にphase IIまでのデータで申請できるかという点、発生数が少なく、phase IIIを行うことが難しいがん種の場合です。

一方、比較的発生数の多い「非小細胞肺癌」、「胃癌」、「大腸癌」、「乳癌」の承認申請の際には、phase IIIまでの治験データが必要なことが、2006年に改訂された「抗悪性腫瘍薬の臨床評価に関するガイドライン」に定められています。その後、2021年に改訂された「抗悪性腫瘍薬の臨床評価に関するガイドライン」では具体的な癌種の記載はなくなり、「患者数が多いがん種を対象とした臨床試験では、生存期間の延長等に基づき、確実な有効性を示す必要がある。ただし、遺伝子異常に基づいた希少なサブタイプを対象とする等、科学的根拠に基づいて効果が期待される患者数が著しく少ない場合には、全生存期間に基づく検証的試験の実施が困難であったり相当の期間を要すると判断される場合も想定され、上記の限りではない。」と記載されました。

なお、一般薬の開発においても、発生数が少ない希少疾病の場合は、図に示したとおりのステップでは行われません。

## 「治療開発の概要」のまとめ

- 開発の「相 (phase)」と「試験の種類(type)」は本来別もの  
✓ だが、抗がん剤の開発では1対1対応するため、類型化されている
- 臨床試験は全体として無効なものを落としていく「スクリーニング」のプロセスである
- 一般薬と発生数の少ない一部のがんでは、開発のプロセスに違いがある



ここまでの話をまとめると、このようになります。

例として採り上げてきた抗がん剤は、phase II までのデータをもって新薬として薬機法により承認されて世に出てくることが多くあるため、世に出た後にphase III、すなわち、標準治療との比較試験を行わなければなりません。

しかしながら、製薬企業は抗がん剤に関し、承認取得後の臨床試験には積極的ではありません。

その理由としては、

- ・抗がん剤のマーケットはあまり大きくないため、他領域に比べ開発の優先順位が低い
- ・その時点で標準治療である他社製品との比較試験を行うためには、他社との調整が必要だが、困難であるなどが挙げられます。

そこで、最善の治療を開発したい研究者は、自ら臨床試験を企画し、実施することになります。

また、がんの治療は抗がん剤単剤で行うことは少なく、「集学的治療」と呼ばれる2剤以上の併用や、手術や放射線と併用して行うことが多いため、臨床試験の複雑さが、製薬企業が主体となって実施できる範囲を超えているともいえます。

がん領域の治療開発にはこのような事情があり、標準治療の確立には研究者主導の臨床試験が必須といえます。



# 研究デザイン

## 臨床試験方法論のエッセンス

- **正しく判断**……被験者数は**多いほどよい**
  - ✓ よい治療は正しく「よい」と判断
  - ✓ よくない治療は正しく「よくない」と判断
- **効率的に治療開発**…被験者数は**少ないほどよい**
  - ✓ すべての新しい治療（薬）に phase III はできない
  - ✓ 有望なもののみを選んで先に進める
  - ✓ 有望でないものを落とす

**「正しい判断」と「効率」のトレードオフ**  
**因果関係の推論 < 治療の技術的評価**



ここから、研究デザインの話に移ります。

臨床試験の方法論のエッセンスは「正しい判断」と「効率」のバランスを取ることです。

なぜならば、試験結果を正しく判断するためには、試験に参加する患者さん（被験者）の数は多いほうが統計学的な精度は高くなりよいといえます。

一方、効率を考えると、よい薬をなるべく早く世に出すためには、患者さんの数は少ない方が、早く試験結果が出るためよいでしょう。

なお、患者さんの数が少ない方がいいのは、効率面のみではなく、倫理的な理由もあります。

例えば、比較試験の場合、どちらかが統計学的に有意に優れるということを証明するために必要な数を大幅に超えて患者さんに試験に参加いただくということは、本当は試験の途中で、どちらがいいかわかっているのに、わかった後も優れていない方の治療を患者さんに受けさせることになってしまい、それは、非倫理的といえるからです。

このような理由から、臨床試験方法論のエッセンスは、いろいろな工夫をして「正しい判断」と「効率」のバランスを取ることといえます。

# 臨床試験：基本は比較

## Phase III

標準薬・標準的治療法と  
総合的な有用性を  
直接同時比較

標準薬・  
標準治療

## Phase II

過去のデータから導かれる  
有効性の水準や目標と比較  
(奏効割合・有効割合)

## Phase I

過去のデータから導かれる  
安全性の許容水準と比較  
(許容される毒性の割合)



いずれのphaseにおいても比較相手は標準薬・標準治療



臨床試験の基本は比較といえます。比較するのはphase IIIのランダム化比較試験だけではありません。Phase IIIでは標準治療と新しい治療を1つの試験の中で直接同時に比較しますが、phase IIやphase Iでは、過去のデータから導かれる水準と比較します。よって、いずれのphaseにおいても、標準薬・標準治療を比較の相手として考えて試験デザインを考える必要があります。

## 何を比較するのか？

### 患者さんの利益 (benefit/beneficial outcome) を比較する

– 今までの標準薬/標準治療と比べて・・

- よく効く
- 副作用が少ない
- 安い
- **有効性の比較** (狭義の benefit)
- **安全性の比較** (risk/harm)
- **経済性の比較**

– ただし医療経済の仕組みに左右される



では、治療の「よさ」は、「何を比較対照」として「どんな点」を比較することによって評価するのでしょうか。

具体的にいうと、新しい治療の「よさ」は、「従来の標準薬・標準治療」を比較対照として「患者さんの利益 (benefit)」を比較して、評価しなければなりません。

従来のものより「よく効く」というのは当然、患者さんの利益です。そこでまず有効性を比べることになります。そして、副作用が少ない、治療が楽だ、といったことも利益になりますので、安全性を比較します。

また、同じくらい効いて、同じくらい副作用が少ないのであれば、値段が安い方が患者さんにとって利益があるといえますから、経済性についても考える必要があります。

ただし、これは医療保険制度の違いにより各国間で差があったり、日本のような皆保険制度ではない国の場合には、加入する保険の種類により個人毎に異なりますから、比較する指標とはしがたいです。

よって、一般的に患者さんの利益を測る指標としては有効性と安全性の2つが基本となるといえます。

## 何をもって比較するのか？

- **エンドポイント (endpoint、評価指標、評価項目)**

- 原義は「終点/ゴール」「目的」・・・臨床研究では・・・

「患者さんの benefit を測る“ものさし”」

“ criterion by which patient benefit is measured”

- Richard Simon

Cancer Principle & Practice of Oncology : DeVita / Lippincott - Raven



有効性や安全性を比較する際に用いる具体的指標“ものさし”のことを、臨床試験の用語ではエンドポイントといいます。

日本語では「評価項目」といった言葉を用いることもあります。エンドポイントを「目的」といった意味で用いる場合もありますが、臨床試験の分野では「ものさし」という意味で用いることが誤解がないと考えられます。

# エンドポイントのタイプ

患者さんの ベネフィットを 直接反映 するか？	真のエンドポイント	<ul style="list-style-type: none"> <li>患者さんの benefit を直接反映しているエンドポイント</li> </ul>
	代替のエンドポイント	<ul style="list-style-type: none"> <li>真のエンドポイントの代わりに使われるエンドポイント               <ul style="list-style-type: none"> <li>より簡便、早く測定できる、少ない患者でわかる等の利点あり</li> <li>真のエンドポイントと評価の結果が一致</li> </ul> </li> </ul>
客観的か？	ハードなエンドポイント	<ul style="list-style-type: none"> <li>誰が見ても同じ・何回見ても同じ</li> </ul>
	ソフトなエンドポイント	<ul style="list-style-type: none"> <li>真のエンドポイントの代わりに使われるエンドポイント               <ul style="list-style-type: none"> <li>見る者によって違う・時々によって違う</li> <li>他の影響を受けやすい</li> </ul> </li> </ul>
その臨床研究 の中で 最も重要か？	Primaryエンドポイント	<ul style="list-style-type: none"> <li>その研究で最も重要なエンドポイント。</li> <li>通常は一つに絞って設定する</li> </ul>
	Secondary エンドポイント	<ul style="list-style-type: none"> <li>Primaryエンドポイント以外のエンドポイント。</li> <li>通常は複数項目設定される</li> </ul>



エンドポイントにはいろいろなタイプがありますが、分類方法として、

- ・「True (真の)」と「Surrogate (サロゲート、代替の)」
- ・「ハード」と「ソフト」
- ・「Primary」と「Secondary」

の3つが重要な考え方です。

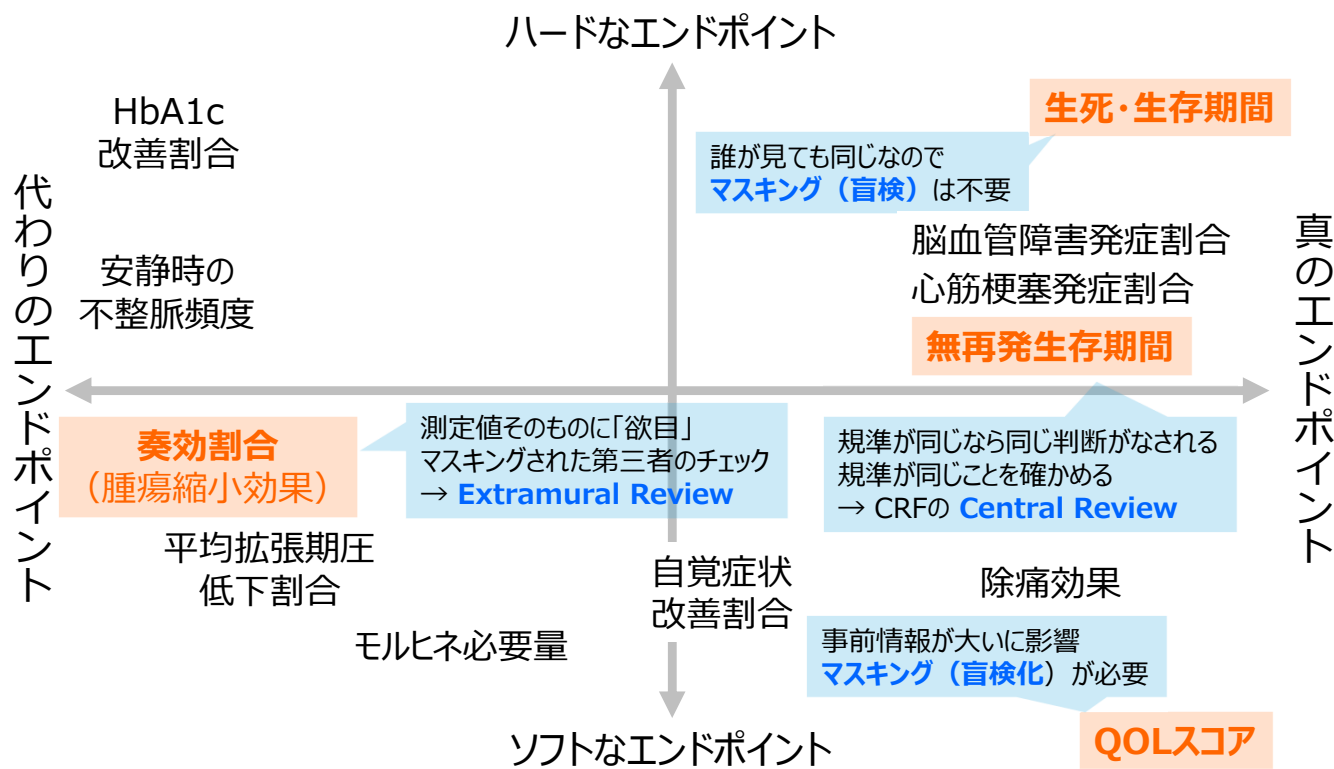
真のエンドポイントは患者さんのベネフィットを直接反映しているようなもので、代替のエンドポイントは、真のエンドポイントの代わりに用いられるものです。代替という以上、真のエンドポイントを用いるよりも簡便である、早くわかる、少ない患者でわかるといったメリットが必要ですが、真のエンドポイントによる評価とできるだけ評価が一致していることが大前提です。

ハードなエンドポイントとは、測定の仕方によらず、同じ結果が得られるエンドポイントをいいます。誰がどのように測定しても同じハードなエンドポイントなら、データの信頼性・再現性が高いといえます。

一方、ソフトなエンドポイントは、評価者や患者さんの意識の影響を受けたりするような信頼性・再現性の高くないものをいいます。信頼性を上げるには、評価の際に群をマスキングをしたり、測定方法を厳密に定めたりするなどデザイン上の工夫が必要となります。

Primaryエンドポイントとは、その研究で最も重要なエンドポイントであり、通常は一つに絞って設定されます。このprimaryエンドポイントを用いてその研究の目的が達成されたかを判断します。Secondaryエンドポイントとは、primaryエンドポイント以外の全てのエンドポイントです。通常は複数項目設定されます。例えば、がんを対象とした第III相試験であれば、primaryエンドポイントを全生存期間、secondaryエンドポイントは無増悪生存期間、奏効割合、有害事象、QOLなどを設定します。

# エンドポイントの2つの軸



**エンドポイントのタイプが品質管理の方法、結果の信頼性に関係**



先ほど挙げたエンドポイントに関する2つの観点(軸)から、よく用いられる典型的なエンドポイントを2次元に配置してみます。ここでは縦軸でハードさ、ソフトさを示し、上側に位置するほどハードなエンドポイントとなり、横軸で真のエンドポイントか代わりのエンドポイントかを示し、右側に位置するほど真のエンドポイントに近いものを表します。

臨床試験でがんの治療効果を調べる場合には、ベネフィットとしての生存期間は真のエンドポイントですし、何月何日に亡くなったというのは誰が確認しても同じですから、ハードなエンドポイントといえます。それに対して、腫瘍縮小効果はそれ自体が患者さんのベネフィットではなく、生存期間の延長につながって初めてベネフィットがあるため、サロゲートエンドポイントといえます。また、評価を画像で行うならば、評価者によって異なる可能性があるため、ソフトなエンドポイントといえます。

また、Quality of Life (QOL) は患者さんの主観的なベネフィットそのものですが、測定するのが難しく、QOL 質問票などで定義されたQOLスコアはQOL自体ではなく、QOLのサロゲートであり、質問票という性格からソフトなエンドポイントということになるでしょう。

除痛効果の指標としてモルヒネ必要量を使うならば、それもサロゲートエンドポイントとなるでしょう。

また、前のスライドで説明したがんの例を思い出してください。第III相試験のような検証的な試験では右上にある生存期間のようにハードかつ真のエンドポイントがprimaryエンドポイントに設定されます。一方で、第II相試験のように探索的な段階の試験では、ソフトではあるものの早く簡便に少ない患者でわかる事を優先して奏効割合がprimaryエンドポイントに設定されることが多くあります。このように、エンドポイントのタイプは、開発段階(相)ごとのエンドポイントの設定にも関係しています。また、ソフトなエンドポイントや代わりのエンドポイントはその欠点を補うための工夫(マスキングや中央判定など)が必要であり、エンドポイントのタイプが品質管理の方法や結果の信頼性に関係していることがわかります。

# 相ごとの方法論のエッセンス

	Phase I	Phase II	Phase III		
目的	Phase II に進むかどうかを決める 安全性のスクリーニング Phase II でのレジメン (用量/用法)を決める	Phase III に進むかどうかを決める 有効性のスクリーニング 毒性プロファイルの充実 治療変更規準等の最適化	標準治療を決める 次は日常診療 総合的な Benefit/Risk 評価の検証的試験 標準治療とのランダム化比較		
主たる Endpoint (ものさし)	・毒性 (MTD・DLT) ・探索的な有効性 (効きそうな薬かどうか)	・有効性の短期的な 代替エンドポイント ・毒性	・有効性の長期的な 真のエンドポイント ・毒性		
対象					
- 疾患要因 (がん種/stage)	広い (がん種を問わない)	>	狭い (特定)	=	狭い (特定)
- 宿主要因 (年齢/臓器機能)	狭い (正常臓器機能)	<	やや広い	<	もっとも広い
事前の安全性 情報の蓄積と 患者さんのリスク	ほとんどなし (リスク大)	<	少ない (リスク中)	<	多い (リスク小)
参加施設	単施設 (~少数の専門施設)	<	中規模 (専門病院主体)	<	大規模 (一般病院主体)



この表では、ここまでのまとめとして、相ごとに典型的な試験デザインを要約しています。

表の下2段にまとめたように、phase I の段階ではまだ薬の候補は海のものとも山のものともわかりませんので、事前の安全性情報の蓄積はほとんどなく、参加する患者さんのリスクは大きいといえます。よって、phase I はphase I 実施体制が整備された1つ、あるいは少数の専門施設で行う必要があります。

一方、phase II、phase III へと進むにつれて、安全性情報が蓄積され、リスクに対する対処の方法がわかってきますから、参加する患者さんのリスクはphase I にくらべて小さいといえます。よって、phase II は専門病院を中心とした複数施設、phase III は、将来実際に利用される現場に近い一般病院も多く参加する大規模な体制で行います。

## 研究デザイン：経験上しばしば不明確なもの・・・

- **明らかにしたい仮説は何か？**
  - **Clinical Question** は明確か？・・・「疑問文」で書いてみる
    - 胸部食道癌に対し、FP術後化療は、再発抑制 効果があるか？
- **意味がある仮説か？**
  - 研究の結果、**Clinical Decision Making** が変わるか？
    - 術後化療が勝てば（positive results）術後化療が標準治療
    - 勝たなければ（negative results）手術単独が標準治療
- **標準治療はなにか？**
  - **Standard therapy** or commonly-used therapy はなに？
  - 評価したい治療のどこが「実験的」？
- **どういう結果が出れば「よい治療」とするか？**
  - **Decision Criteria** の明確化
  - 標準治療に勝たなければならないのか：優越性・・・どれくらい？
  - 標準治療に負けなければよいのか：非劣性・・・どれくらい？



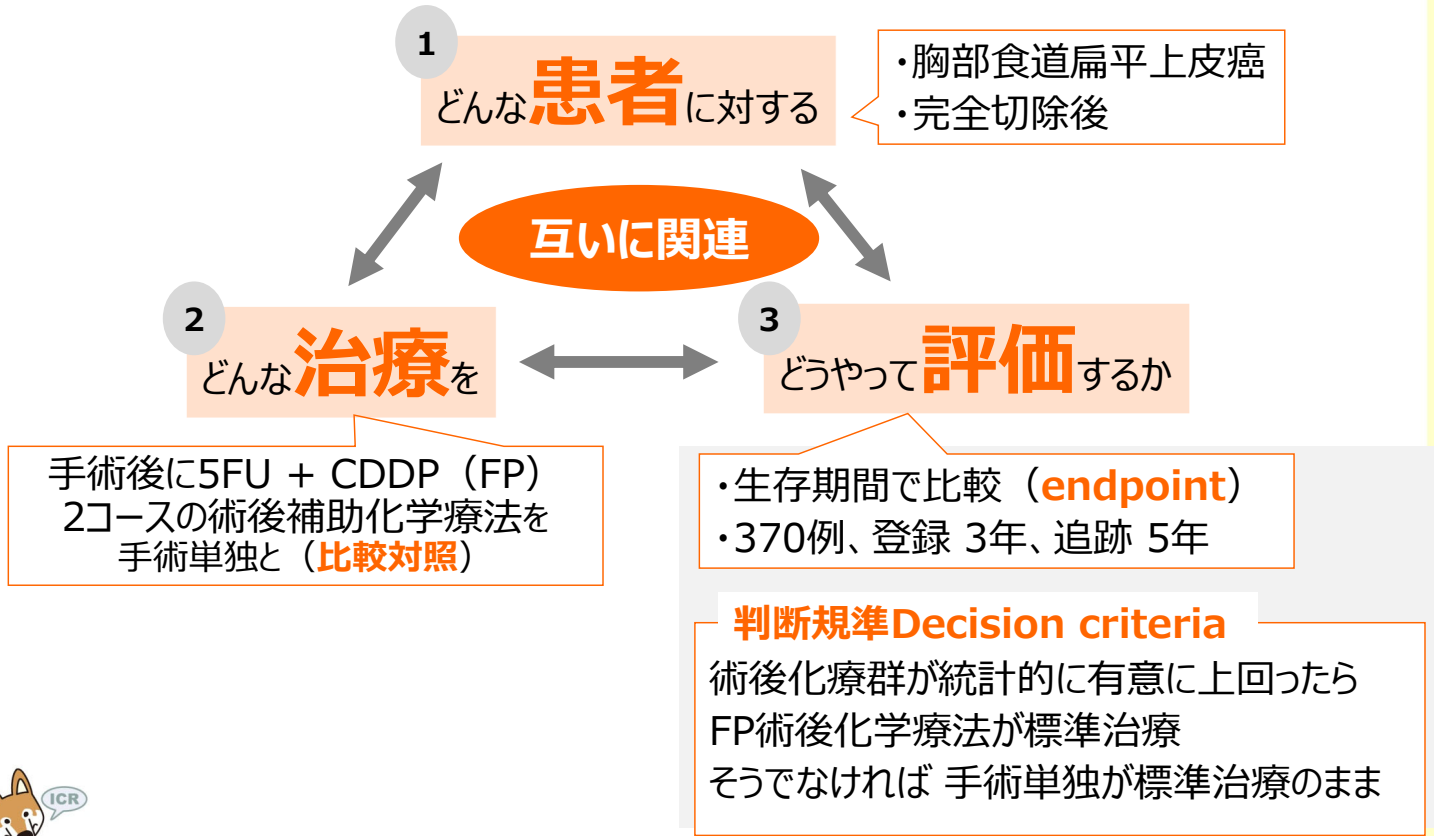
研究計画を立てる際には、すべてをできるだけ具体的に考え、得られる結果が意味のあるものとなるようにしなければなりません。

漠然と考えていてもいい研究計画を立てることはできません。

これから、具体的に研究計画を考えるために役に立つ考え方を説明していきましょう。



# 研究デザイン：3つの基本要素



臨床試験の研究デザインを具体的に考える上で考慮すべき基本要素は、「患者」、「治療」、「評価」の3つです。

どんな「患者さん」に対してどんな「治療」をしてどうやって「評価」するかについて、それぞれの要素がどのように関連しているかを捉えることが重要です。

このことは先に紹介した臨床試験の教科書 ( Pocock, Clinical Trials ) でも述べられています。

あるphase IIIを例として説明します。

この例では、

- 「患者」として胸部食道扁平上皮癌で手術が終わった人、
- 「治療」として5FUとシスプラチン(CDDP)2剤の化学療法を手術後に2コース行うという新しい治療(以下、術後化療)を標準治療である手術単独と
- 「評価」として、治療の成績を「生存期間」が長いかわりか短いかわりかというものさし(=エンドポイント)で「比較」します。

評価のためには、ものさしである「エンドポイント」とともに、それを用いた判断規準が必要です。この例の場合は、「術後化療群の生存期間が手術単独群よりも統計的に有意に長ければ術後化療を新しい標準治療とし、そうでなければ手術単独が標準治療のままとする」というのが判断規準となり、これが調べられるような臨床試験デザインを用いて試験を実施することが必要となります。

対象患者、比較したい治療、評価項目(ものさし)と判断規準が明確であれば、それを実証するための臨床試験デザインを決定することができます。

# Phase IIIの例：代表的なJCOG試験

## JCOG9511 (肺がん)

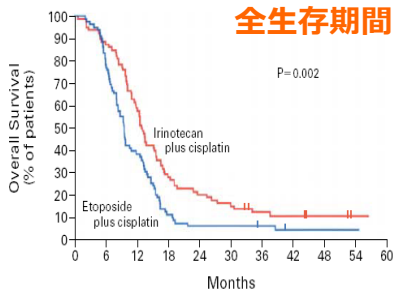
進展型小細胞肺癌  
70才以下

ランダム割付

イトホシド + シスプラチン  
(標準治療)

イリノテカン + シスプラチン  
(新治療)

全生存期間



Noda et al. N Engl J Med 2002

進展型小細胞肺癌に対しては  
延命効果が優るイリノテカン+シスプラチン  
が新しい標準治療である

## JCOG9204 (食道がん)

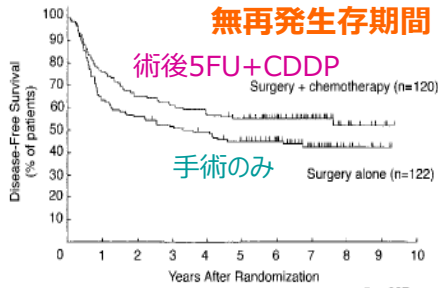
胸部食道扁平上皮癌  
肉眼的根治切除後

ランダム割付

手術のみ  
(標準治療)

術後  
5FU+CDDP  
(新治療)

無再発生存期間



Ando et al. J Clin Oncol 2003

治癒切除後進行食道癌に対して  
5FU+CDDP術後化学療法  
は再発抑制効果がある

## JCOG9502 (胃がん)

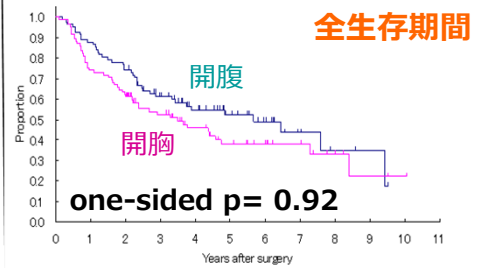
食道浸潤 (3cm以下) を有する  
胃体部癌・噴門癌

ランダム割付

開腹横隔膜  
切開法  
(標準術式)

左開胸  
開腹法  
(拡大術式)

全生存期間



Sasako et al. Lancet Oncology 2006

食道浸潤を有する体部噴門癌には  
合併症が多く生存期間が劣る  
開胸術は推奨されない



これらは、がんの多施設共同研究グループJCOG(日本臨床腫瘍研究グループ)で実施したphase IIIの例です。

それぞれにおいて、先に説明した3つの基本要素を考慮して試験を実施しました。

それぞれ3つの要素が何だったか、そして判断根拠に基づくと結果はどうであったか、考えてみてください。

## 「研究デザイン」のまとめ

1. 臨床試験の基本は「比較」
2. 何を比べて判断するのか・・・患者さんのbenefit
3. 何を使って比べるのか・・・エンドポイント
  - True endpoint vs. surrogate endpoint
  - Hard endpoint vs. soft endpoint
4. どんな患者への、どんな治療を、どうやって評価
  - 3つは互いに関連する
  - これらが明確・整合ならいい研究デザイン
  - これらが不明確・不整合ならダメな研究デザイン



研究デザインについての簡単なまとめです。

「患者」「治療」「評価」は互いに関連します。対象患者によって治療法が決まり、患者と治療法によって評価項目が決まります。

また、逆に患者と評価項目によって治療法が決まることもありますし、治療法によって患者が決まることもあります。

よい研究デザインは、これら3つの要素が明確に定められており、かつ、相互に整合性がとれている(必然性がある)ことです。

# 品質管理と品質保証

—元は工業分野から来ている—



生産ライン

Introduction to Clinical Research

ここからは研究の質を保証するための仕組みである、品質管理 (quality control、QC) と品質保証 (quality assurance、QA) について説明します。

品質管理、品質保証は、元々物作りの工業分野から出てきた概念ですが、臨床試験・臨床研究の分野にもその考え方が導入されています。

品質管理とは、例えば薬品工場で、生産ラインの各工程において検査をして問題があればフィードバックをして改善し、出荷する薬品の品質を一定以上に保つことをいいます。

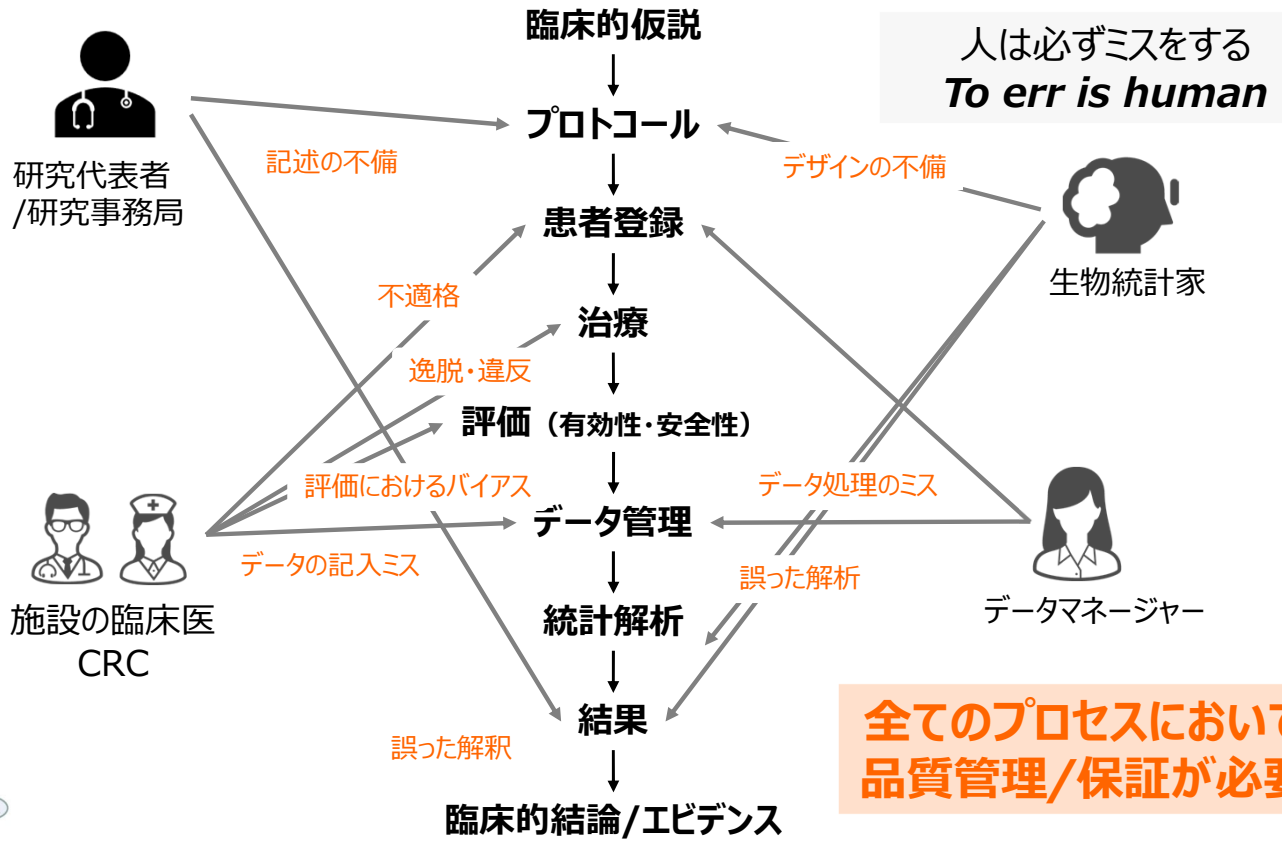
臨床試験の品質管理も同じ考えで、臨床試験の各プロセスにおいて、検査とフィードバックを繰り返して、データの質を高めることをいいます。

一方、品質保証とは、薬品工場の例では、出荷前の最終完成品である薬品の中から抜き取り検査をして、定められた品質通りの製品ができあがっていることを保証することをいいます。

臨床試験の品質保証としては、臨床試験の途中あるいは最終段階において、工場における抜き取り検査と同様に臨床試験データを抜き取って確認する「監査」が主なものとして挙げられます。

正確には、監査だけでなく品質管理活動を含めた品質を保証するシステム全体のことを品質保証と呼びます。

# すべてのプロセスで「エラー」は生じる



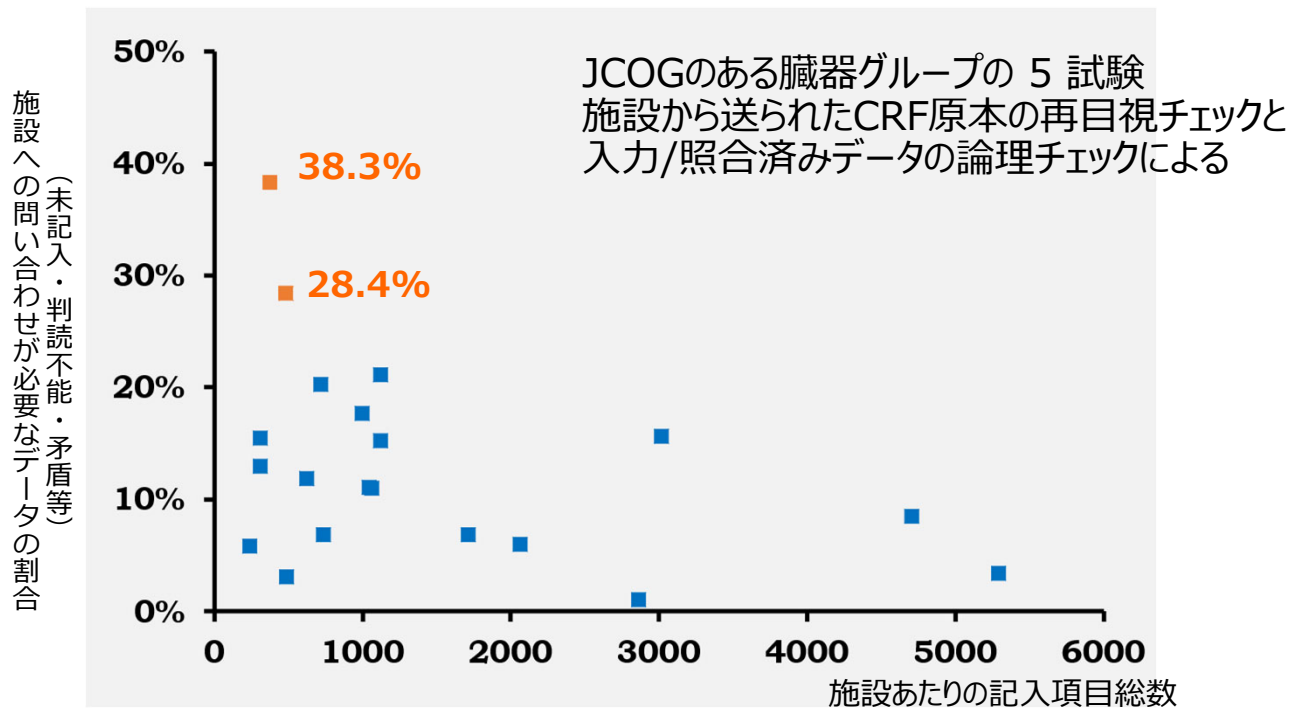
臨床試験は様々なプロセスの直列的回路といえます。1つのプロセスに関わる人は1人の場合もあれば、複数の人、多職種が関わる場合もあります。

人は必ずミスをするものですから、このすべてのプロセスのどこにでもミスが生じる可能性があります。ミスは0(ゼロ)にはできません。

しかも、直列回路ですから、それぞれのプロセスのミスの累積により、全体の結論は本当の結果からどんどんと離れて行ってしまいます。

よって、それぞれのプロセスでのミスを最小化することが必要です。すべてのプロセスにおいて許容できる範囲のミスで進んで初めて、正しい結論に結びつくといえますので、すべてのプロセスにおいて、品質管理活動(QC活動)、そして品質保証が必要となります。

## データにはどれだけエラーがあるか？



**クリーニング**して初めてエラーがあることがわかる



データのエラーと品質管理について、過去のJCOGの例を用いてお示したいと思います。

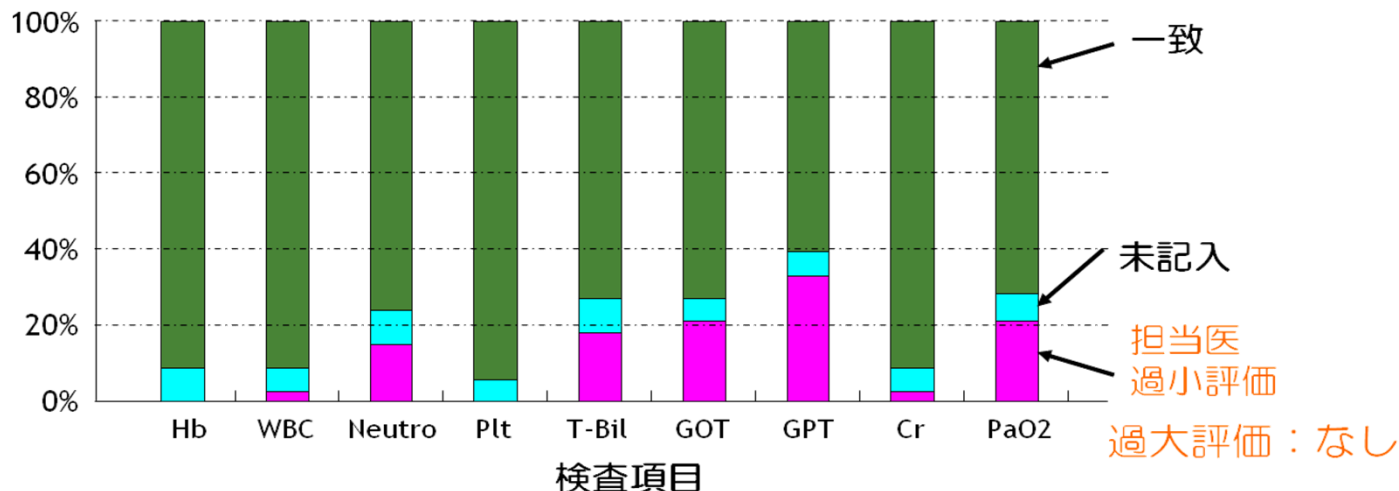
図は、あるJCOGの5つの試験を題材に、施設から送られてきた生のデータと、品質管理としてデータクリーニングを行い最終的にきれいになったデータとを照合した結果を表しています。横軸は施設あたりの記入項目数、縦軸はデータクリーニングのために施設への問い合わせが必要であったデータの割合を示しています。

施設あたりの記入項目数が多い施設のほうが問い合わせの必要なデータの割合が低い傾向が見て取れます。これらの施設は臨床試験に慣れている施設といえるでしょう。図のオレンジの点のように、記入項目数の多くない施設、すなわち臨床試験に慣れていないような施設の中には、データの3割、4割にエラーがある場合もあります。どの施設にもエラーがあり、施設から送られてきたデータをそのまま入力するだけではかなりのエラーが存在する可能性のあることがわかります。データを解析する前に「データ・クリーニング」というQC活動をすることの重要性がみてとれます。

## 毒性評価のバイアス：担当医の評価

### 担当医判定グレード vs. 検査値から直接グレード

(あるJCOGスタディにおける毒性評価：JCOG毒性規準による 最悪Gグレード)



**毒性は過小評価されやすい**  
**ふつうはどれくらい過小評価かはわからない**



毒性評価のエラーについても、JCOG試験のデータで説明します。

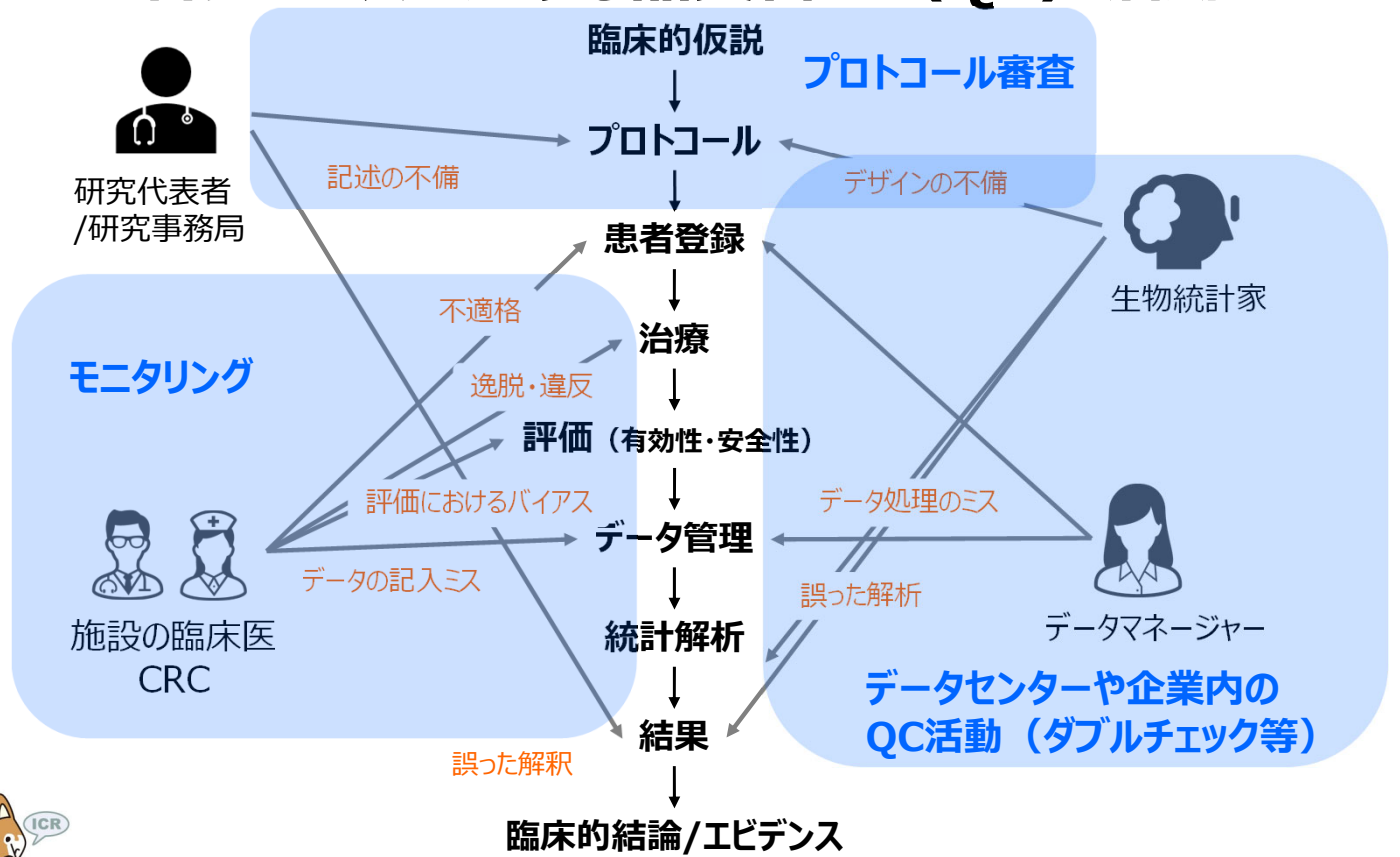
図は、施設から送られてきた検査値から直接評価した毒性のグレードと、施設の担当医が判定した毒性のグレードを突き合わせたものです。2つのグレードのうち、測定された検査値をもとに直接評価したグレードが正しいデータであると考えられます。なお、毒性グレードは0が毒性なしで、2、3、4と進むにしたがって強い毒性を示しています。

図中で緑色で示す部分は両者が一致した割合、ピンクで示す部分は担当医の方が誤って毒性を過小評価していた割合を示しています。担当医が毒性を過大評価していたものは皆無であり、誤りはすべて過小評価でした。

この例からもわかるように、一般的に毒性は過小評価されやすいといわれており、他の臨床試験グループからも同じような報告があります。

毒性は過小評価されやすく、有効性は過大評価されやすいことに注意が必要です。

# 各プロセスに対する品質管理 (QC) 活動



では、臨床試験の各プロセスに対するQC活動の例を紹介します。

プロセスの初めの部分では、臨床的仮説から研究デザインを組み立ててプロトコル作成をしますが、作成されたプロトコルが科学性、倫理性とも確保されているかを審査する「プロトコル審査」もQC活動の1つといえます。

また、データセンター内で行うQC活動には、データマネージャーによるデータチェックや、生物統計家によるプロトコル作成への関与や統計解析が挙げられます。

そして、医療機関で行われていることに対する定期的なQC活動として、モニタリングが挙げられます。

モニタリングは、一般的に以下の2つに分類することができます。

- 中央モニタリング: central (in-house) monitoring
  - ー セントラル(データセンター・企業・医薬品開発業務受託機関(CRO))に収集された症例報告書 (Case Report Form、CRF) を用いて行う
  - ー データマネージャーが行う
  - ー すべての臨床試験で必須
- 施設訪問モニタリング: site visit (on-site) monitoring
  - ー 医療現場を訪れて行う
  - ー モニターが行う
  - ー 治験では通常必須(中央モニタリング+施設訪問モニタリング)



## モニタリングとは？

### ● 医療現場で行われていることに対する定期的なQC

- IRB承認が適切に得られているか？
- インフォームドコンセントが適切に得られているか？
- 適切なプロトコール（最新版）が使用されているか？
- 適格患者のみが登録されているか？
- プロトコールが守られているか？
- データが正しく記入・提出されているか？
- カルテへの記録が正しくなされているか？
- 安全性情報が適切に共有されているか？
- 有害事象報告が適切になされているか？
- 患者さんへの情報提供が適切になされているか？



Introduction to Clinical Research

### SDVのみがモニタリングではない

モニタリングではどのようなことをチェックすべきか具体例を挙げてみます。

施設訪問モニタリングでは、データセンターに提出されるCRFに正しくデータが記載されているかどうかを、医療機関にあるカルテなどを直接閲覧して照合すること(Source Document Verification、SDV)が行われます。しかし、SDVをすることがモニタリングなのではなく、試験が適切に実施されているかどうかを様々な視点からチェックすることが本来の意味でのQC活動としてのモニタリングです。

日頃、治験のみに携わっている人はモニタリングと聞くと施設訪問モニタリングを想像するかもしれませんが、中央モニタリングでも臨床試験の質の確保に十分なモニタリングが行えます。研究者主導多施設共同研究グループでは国際的にも中央モニタリングのみ実施することが一般的です。形態ではなく、QC活動として何が必要かを考え、効率的に実践することが重要です。

## データモニタリング以外のモニタリング

- 不適格
  - やむを得ないか？ 防ぎ得たか？
  - 結論に影響を及ぼさない程度か？
- プロトコル逸脱/違反
  - 原因は？：施設？ プロトコル？
  - 重篤な有害事象に結びついていないか？
- 重篤な有害事象/治療関連死
  - 原因は？：施設？ プロトコル？
  - 防げるか？：防げないなら試験中止
- 結果の信頼性
  - データはきちんと収集されているか？
  - データエラーは結論を歪めない程度か？

施設/担当医に原因



施設に注意/警告

プロトコルに原因



プロトコルを改訂

**フィードバック  
により質を向上**

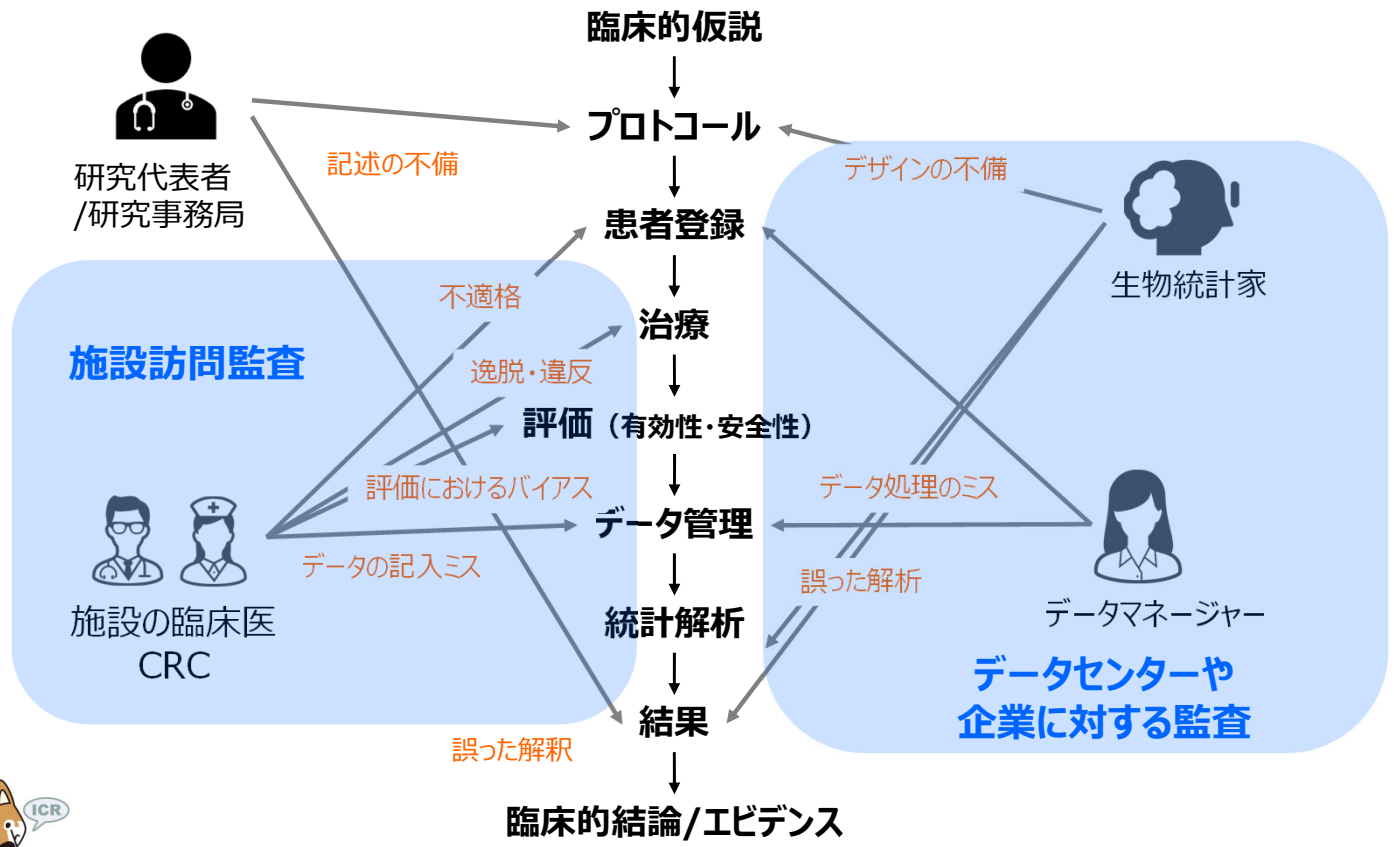


もっと詳しく見てみましょう。

医療現場に対して行うモニタリングはデータのモニタリングだけではありません。むしろ、質の高い臨床試験実施のために必要なモニタリングは、不適格やプロトコル逸脱、重篤な副作用といったものの原因は何かを考察することです。その原因がどこにあるのかを見定めて改善する必要があるからです。

施設担当医に原因があるプロトコル逸脱とか重篤な有害事象であれば、施設で注意すれば不適格や逸脱を減らせますし、プロトコルに原因があるのであればプロトコルを改訂すべきです。どこに起因があるかを考え、適切にフィードバックをすることがモニタリングというQC活動の重要な目的です。

# 各プロセスに対する品質保証 (QA)



臨床試験のすべてのプロセスにおいてQC活動が必要なことは先に述べましたが、品質保証(QA)も全プロセスに対し必要です。

QAの主なものとしては「監査」が挙げられます。

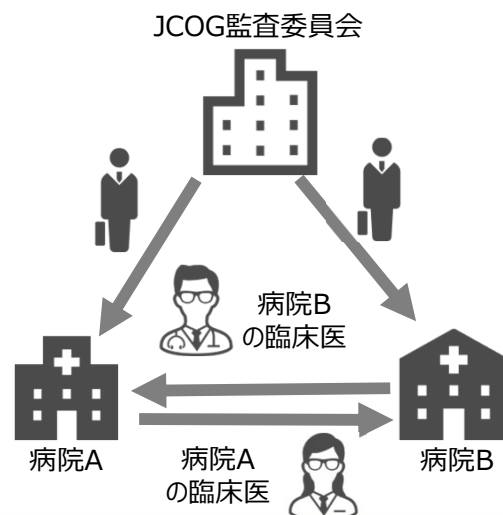
監査は医療機関に対して行うものだけではありません。以下のような種類があります。

- 施設訪問監査:site visit (on-site) audit
  - －医療現場
- 施設以外の組織への監査
  - －業務委託先  
(CRO、治験施設支援機関(SMO)、検査会社、等)
- 内部監査
  - －組織(企業、研究グループ、他)内の各部門(開発部門、データセンター、等)

# JCOGでの品質保証 Quality Assurance

## ● 施設訪問監査 2000～

- 相互訪問監査：別の医療機関の医師が監査チームの一員
- 複数の試験が対象（登録中の試験含む）
- 問題点がフィードバックされるため、品質管理でもある
- 全部のデータを見るわけではない。対象者や項目をサンプリングして確認
  - 倫理的に重要なもの：IRB承認書、同意文書、
  - Primary endpointに影響を及ぼすもの
  - 中央モニタリングでチェックしにくいもの
- 監査委員会から監査結果を関係者に提出
  - 監査実施医療期間の長
  - 研究グループ代表者
  - 研究代表者・事務局
  - データセンター
  - JCOG代表者



研究者主導臨床試験グループによる監査の例として、JCOGで実施している施設訪問監査の仕組みと内容を示します。

JCOGの監査の特徴は、専門の監査担当者に加え、監査対象の医療機関とは別の医療機関の医師が監査に加わる「相互訪問監査」であることや、複数の試験を対象にするという特徴があります。また、進捗中の試験に対して問題点がフィードバックされることがあるので、監査は品質保証のみではなく品質管理の側面も併せもつといえます。

相互訪問監査により別の施設を訪問することによって、自施設の問題点や優れている点などを確認することができるため、監査に同行した医師にとって学習の場にもなります。

## モニタリング・監査・査察の違い

研究実施側が行う	モニタリング Monitoring	監査 Audit	査察 Inspection
目的	早期発見・改善・予防	適切なQCの保証	違反摘発
対象	全施設かつ全症例 かつ全データ	一部 or 全施設・ 一部の症例・データ	問題施設のみ
時期	試験中から いつでも	原則は 試験終了後	必要時
実施者	モニターなど 試験チーム	監査担当者 規制当局	規制当局 査察官
内容	SDV フィードバック	SDV QCの問題点報告	SDV 違反の判定 罰するかどうか



典型的なモニタリングと監査と査察の特徴について、表にまとめました。

監査は、適切なQCがなされているかどうかを監査担当者が施設を訪問して確認し、品質保証ができているというお墨付きを与えるために行うものです。試験の終了時に行うことが多いですが、近年では、試験途中ですることも一般的となってきています。

査察は監査と異なり、通常違反の摘発が目的です。例えば、薬機法やGCP(医薬品の臨床試験の実施の基準に関する省令)、GLP(医薬品の安全性に関する非臨床試験の実施の基準に関する省令)に対して問題があると疑われる場合に、規制当局の査察官が施設を訪問し、罰するかどうかを決めるといったものです。

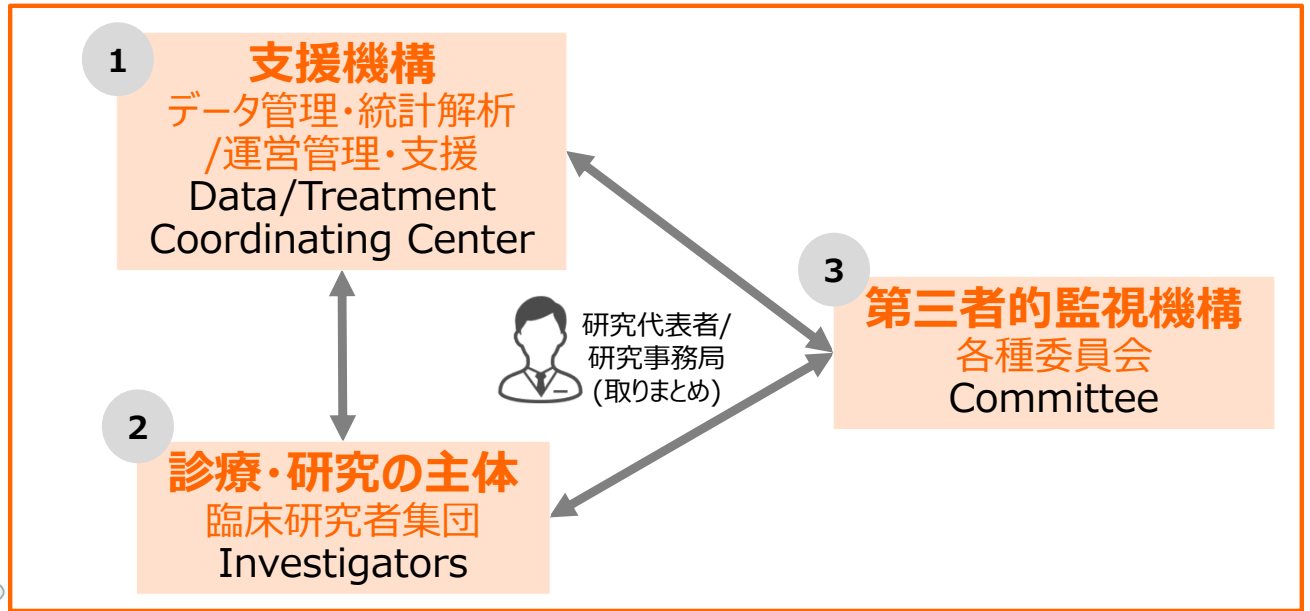
治験の場合に規制当局が行う適合性調査は、新薬承認申請時に添付された資料が厚生労働大臣の定める基準(GCP、GLP、信頼性基準)に適合しているかどうかについて、書面調査や実地調査などによりデータとの照合等の確認を行う業務です。対象、時期、内容は表の監査に相当しますが、違反が疑われ調査を行う場合もあるので、査察の意味合いをもつこともあります。

# 臨床試験に必要な組織

資金提供者  
(企業・国・医局)



## 3つの基本要素 (Essential Components)



最後に、臨床試験に必要な組織・人材について説明します。

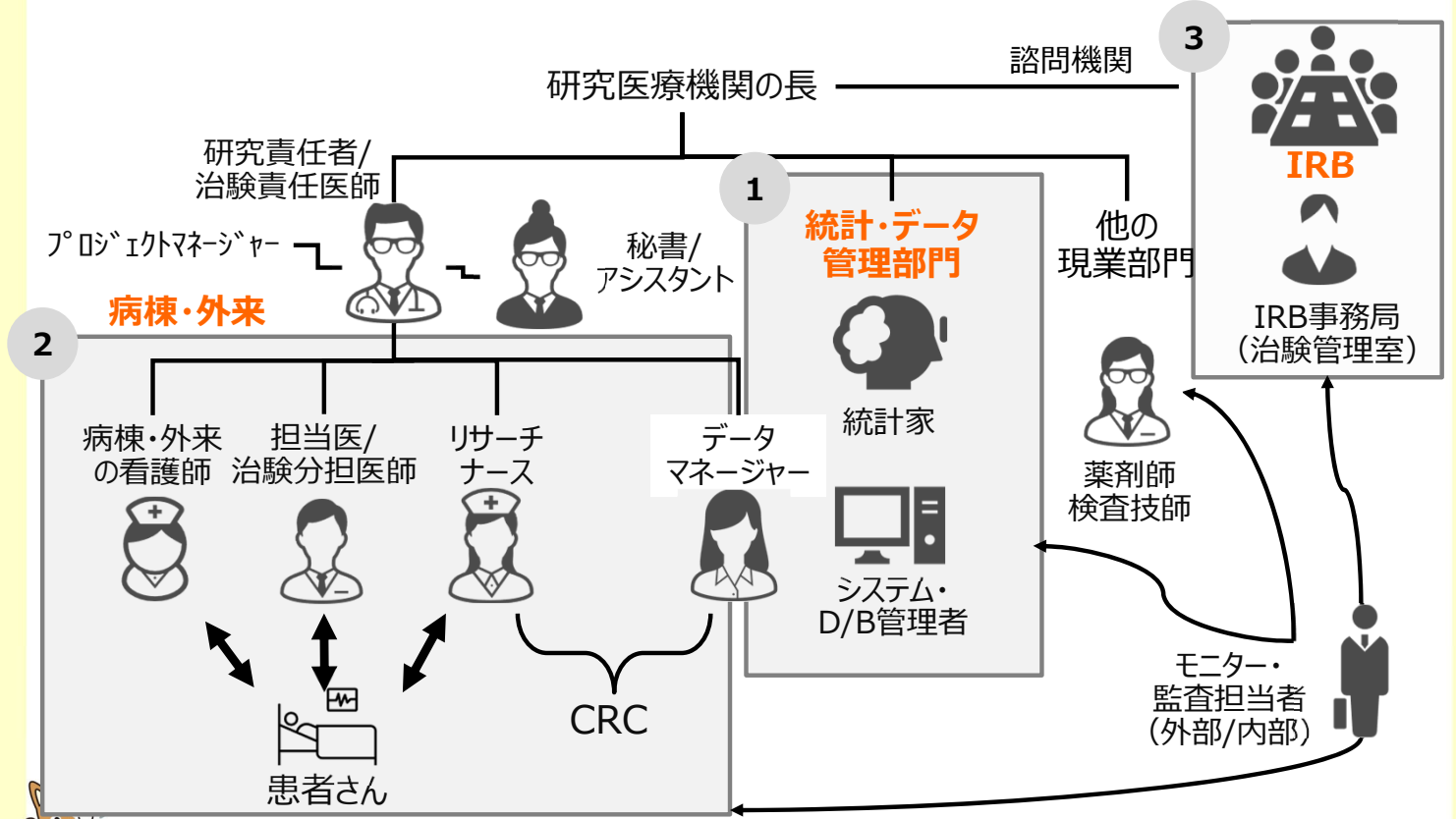
臨床試験を実施するためには、施設の臨床研究者集団の他、研究者を支援する機構と第三者的監視機構の存在が不可欠です。

その他、臨床試験を実施するためには資金が必要ですので、企業や国などの資金提供者の存在も必須といえます。

手弁当で研究を行うことは、科学性・倫理性の担保ができないことから避けるべきです。

もちろん、これら全体をまとめる研究代表者や研究事務局といったマネジメント担当がいないと効率的な運営はできません。

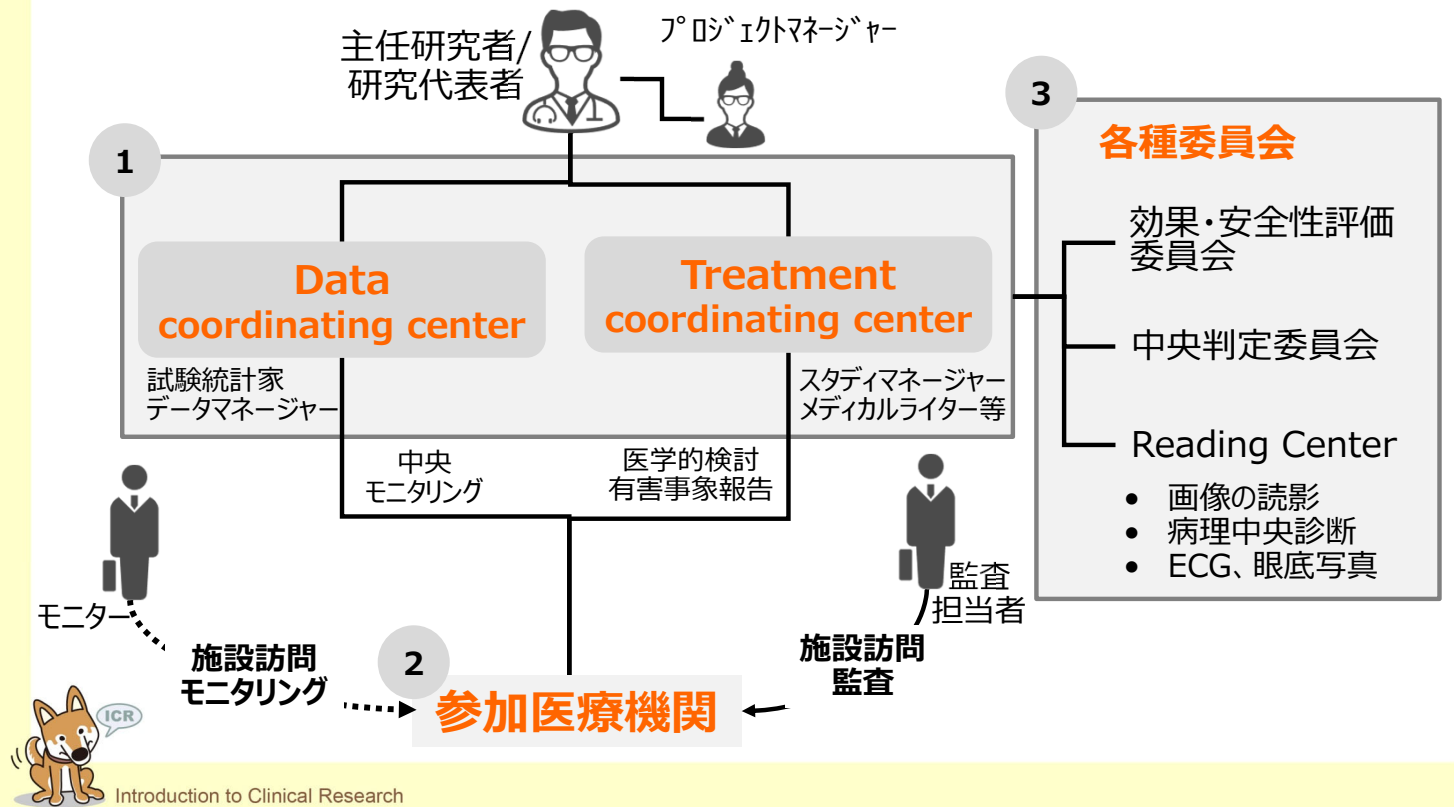
# 単施設の臨床試験に必要なチーム



臨床試験を単施設のみで行う場合にも、現場、支援機構、監視機構の3つの基本要素が必要です。監視機構としては施設の倫理審査委員会などがその役目を担うことが考えられます。

# 大規模多施設共同試験の仕組み

試験ごとに組織する場合 – 循環器・代謝疾患・がん予防



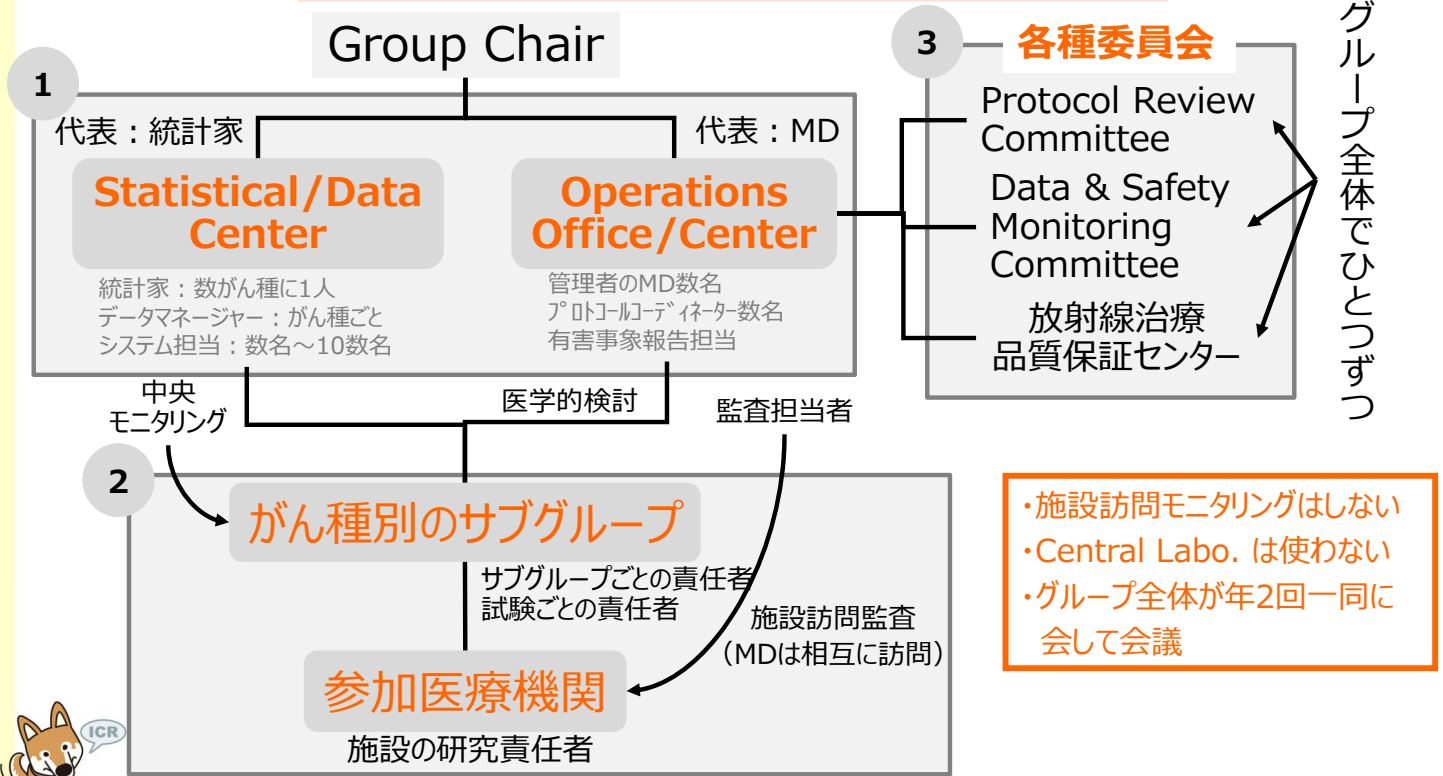
多数の施設が参加して行う大規模多施設共同試験の場合で、1つの試験に特化して組織を作る場合の例を示します。やはり3つの基本要素が必要です。



# がん治療開発の Cooperative Group の仕組み

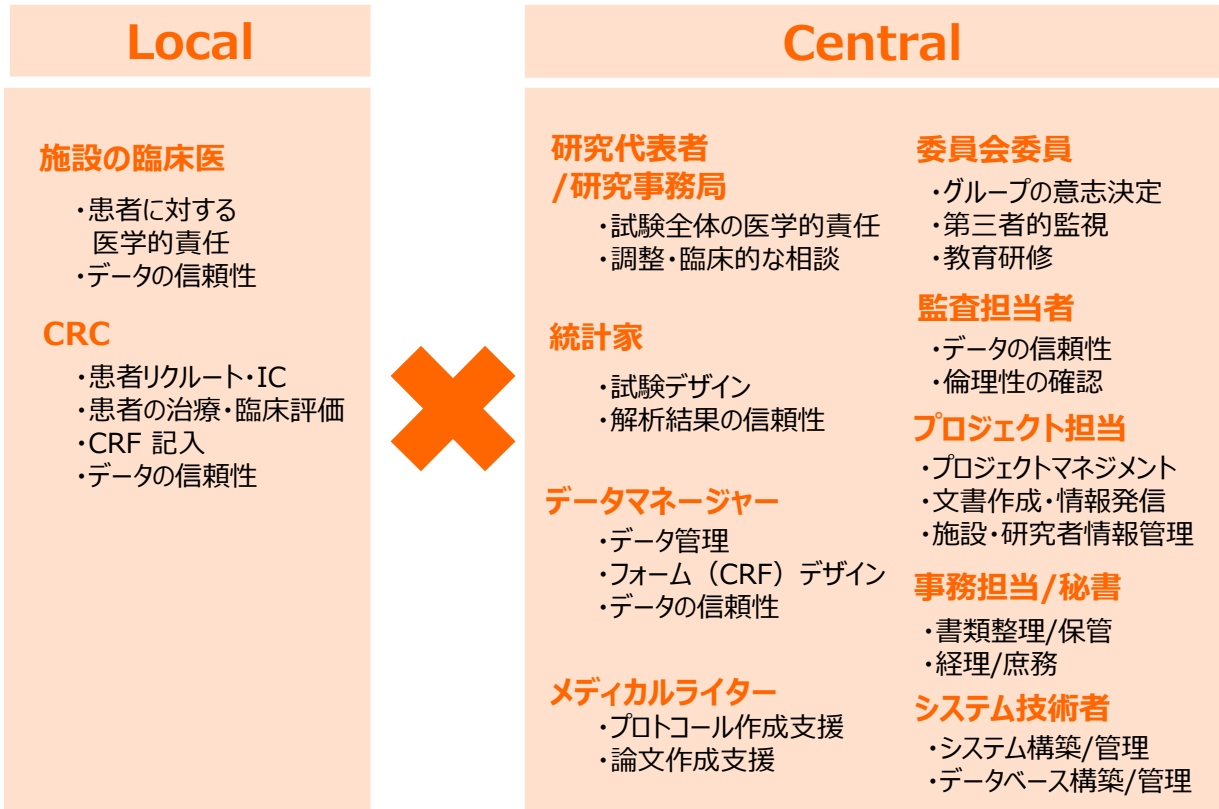
試験ごとではない恒常的な組織

原則グループ全体でひとつずつ



がんの分野では、恒常的な多施設共同研究グループ ( Cooperative Group ) を組織して臨床試験を継続的に行うことが一般的です。3つの基本要素やQC/QAなど臨床試験実施に必要な多くの概念の多くががんの Cooperative groupシステムから考案・整理されてきました。

# 臨床試験は異なる職種によるチームプレー



いずれの規模の臨床試験を行うにしても、3つの基本要素を構成して臨床試験を実施するためには、多くの異なる職種によるチームプレーが必要です。

## 「品質管理と品質保証」 「臨床試験に必要な組織」のまとめ

臨床試験の「品質」とは科学性と倫理性である

- **品質管理・データ管理しないと・・・**
  - 効かない治療を効くと結論してしまう（かもしれない）
  - 危ない治療を安全と結論してしまう（かもしれない）
- **品質管理（QC）：チェックして改善する**
  - 医療現場へのフィードバック = モニタリング
- **品質保証（QA）：チェックしてOKをだす**
  - 医療現場へのQA = 施設訪問監査
  - ただしQCとQAは明確に分けられないことも多い
- **臨床試験に必要な組織の3要素は普遍的である**
  - 研究実施主体・支援機構・第三者的監視機構



「品質管理と品質保証」と「臨床試験に必要な組織」についてのまとめです。



## 本講義のまとめ

- **治療開発の流れ**
  - スクリーニングのプロセス
  - 安全性や有効性などで順にふるいにかけていく（相/phase）
  - 最後は標準治療との比較を行う
- **研究デザイン**
  - 臨床試験の基本は比較
  - エンドポイントとは評価のためのものさしである
  - 研究デザイン 3つの基本要素（患者、治療、評価）
- **品質管理・品質保証**
  - データマネジメント、モニタリング、監査
- **臨床試験に必要な組織**
  - 3つの基本要素（研究実施主体・支援機構・第三者的監視機構）



最後に、本講義の内容をまとめました。

臨床試験は介入研究です。いい加減な結果を出すことは非倫理的であり、行ってはいけないことです。そのためには、臨床試験の手順を理解し、しっかりした研究デザインを立て、質の保証をする活動を行うことが必要です。

そのためには組織が必要です。人はエラーをするもの、何でも1人ではできないことを前提に、役に立つ結果を出すために最善の努力をする必要があります。試験結果の正しさは神のみぞ知る、です。少しでも真の結果に迫るためには、結果の正しさを過程の正しさで保証することが必要です。

臨床試験・介入研究を実施するならば、本腰入れて取り組んでください。