

なかでの奏効率、使いやすさなどを、最初に100%予測できなくてもいいと思うのです。ファーストラインでは、どちらの薬剤のほうが有効性が高いかなど、そういうことが予測できれば有用だと思っています。

西尾 まさしくおっしゃるとおりです。マーカーは100%に近いものでないと臨床試験で層別に使えません。そして、細胞傷害性抗がん剤に関してもそれに匹敵するようなマーカーは必ずあります。そう考えると非常にモチベーションが上がりますね。将来的にはおそらくコンビネーションで行わなければならない、それもまた難しい問題だと思います。

山本 何度も試験を行うことは困難ですから、すべての機会に付随研究を行い、それを有機的に次につなげていくということが必要ですね。

山田 そう、そこですね。

山本 はじめからそういうデザインをしておくというアプローチが、方向性としては望ましいですね。EGFR-TKI 同士で組み合わせるなどの試験はどうでしょうか。

山田 単剤でそれぞれ開発してから組み合わせてい

かなければなりません。追加ですが、ASCOのあとに開催された World Congress on Gastrointestinal Cancer のなかで、2つの抗体を用いた大腸がんに対する試験がありました。FOLFOX+ペバシズマブ+panitumumab で、併用での第I相試験と第II相試験がなく、いきなり第III相試験が行われました。しかし、下痢と皮疹という抗EGFR抗体の毒性が増して、グレード3以上の下痢が20%、皮膚毒性が30%以上の症例にみられたのです。そのため、毒性によって化学療法そのものの dose intensity が下がってしまいました。そして、FOLFOX+ペバシズマブはPFSが10ヵ月でしたが、2つの抗体を用いた群は9ヵ月だったので、中間解析で止まってしまいました。もう少し前段階の試験で、抗体同士の併用を行い、毒性を評価することも必要かと思います。

西尾 海外はそういうことをチャレンジングに行いますね。羨ましいぐらいです。ただし、生物学的製剤に関してはイギリスでの事件の例もありますから、より慎重に、科学的に解析していくアプローチが必要です。本日はお忙しいところ、ありがとうございました。



解説

がん治療研究におけるランダム化 第II相試験の意義*

山本精一郎**

Key Words : molecularly targeted agent, trial design, randomization, phase II trial

最近の薬剤開発事情

最近, 多くの薬剤が開発されており, その中には多くの分子標的薬が含まれる. これらは従来の細胞傷害性抗がん剤(cytotoxic drug)とは異なる毒性プロファイルをもつ. たとえば, 単剤でのactivityが必ずしも高くない, 毒性がオーバーラップしないといった特性をもつなど, 標準的な化学療法との併用が効果的と考えられるケースがある. こういった薬剤の開発には従来のcytotoxic drug開発とは異なるストラテジーが必要となってきた.

これまでのcytotoxic drugでは「用量-毒性」「用量-効果」が相関関係にあり, 結果として毒性と効果が比例していたことから, 毒性による最大耐用量を推定することがそのまま効果を最大化する用量を推定することにつながるというロジックに基づいてphase I trialが行われてきた. それに対し, 分子標的薬をはじめとするいわゆる“cytostatic drug”では, かならずしも用量と効果が比例しない. そのため, 毒性の面からの最大耐用量を推定するプロセスを通して効果の面での最適用量を推定することが難しくなった. その結果, 「用量-ベネフィット/リスク比」関係を検討して至適用量(optimal dose)を決めるという, 一般薬開発で用いられてきた考え方と同じ方法をとる必要が出てきた. つまり, 最大耐用量を探す古典的なphase I デザインでは, 至適用量が決定できなくなり, その後の開発ステージであるphase II や場合によってはphase III で用量設定

を行うという必要が生じてきたということである.

また, 最近, phase III で新規薬剤と標準治療の比較を行ったが, 有効性が証明されず失敗に終わったという例がいくつかあった. 多くのphase III を実施することは通常困難であるため, 失敗した場合には, その薬剤の開発が中止されるだけでなく, 他の薬剤の開発の遅れにもつながる. 新規薬剤がそれほど多くない時代には, phase III で検討の対象となる他の選択肢がないためにこれはそれほど大きな問題ではなかったが, ほかに多くの候補薬がある現在では, phase III の失敗は治療開発全体の遅れにつながってしまう. そのため, 医療に携わるものとしてはできるだけ成功する確率の高い薬剤を選んでphase III を行いたい.

ではどうやって有効性のスクリーニングをすればよいのであろうか. 単剤で抗腫瘍効果のみられる薬剤であれば伝統的なphase II デザインで奏効割合(response)などの腫瘍縮小効果をエンドポイントにすればよい. たとえば, CMLやGISTに対するimatinibやNSCLCに対するgefitinibなどがその例であろう. しかし, 分子標的薬の中には必ずしも単剤としてのactivityが高くなく, 標準治療との併用療法として期待されているものも多い. このような場合, phase III へ進めるべき薬剤をどのようにスクリーニングすればよいのであろうか.

このような状況下でphase II に求められる役割は, これまでの有効性によるスクリーニングとい

* Randomized phase 2 design for molecularly targeted agent.

** Seiichiro YAMAMOTO, Ph.D.: 国立がんセンターがん対策情報センター/JCOGデータセンター統計部門〔〒104-0045 東京都中央区築地5-1-1〕; CIS, National Cancer Center, Tokyo 104-0045, JAPAN

う目的に加え、いくつかの用量や候補薬剤、combinationの中から、phase IIIで成功する確率の高い薬剤や治療法を選択するということである。この場合にrandomized phase IIが有効なデザインと考えられる。なぜなら、phase IIで標準治療のconcurrent controlと比較できれば、ヒストリカルコントロール(historical control)を用いることによる選択バイアスを防ぐことができるため、有効性に関してsingle armの試験より高い確信が得られるからである¹²⁾。ランダム化は、新薬や分子標的薬との組み合わせで目安にすべきヒストリカルコントロールのデータがあまりない場合やresponseがエンドポイントとして適切でなく、time to progressionの方がエンドポイントとしてよい場合にはさらに有効である。なぜなら、time to progressionやsurvivalなどは、ヒストリカルコントロールと同じ選択規準で試験に組み入れられる対象者を選択しても、各被験者がもともと予後によって大きく影響を受ける、すなわち、治療以外の要因によってtime to eventに大きな違いが生じてしまう可能性がある。また、抗腫瘍効果は腫瘍が縮小したことを直接観察できるのに対しtime to progressionやsurvivalの延長は直接観察することができない。このようなtime to eventの欠点は、標準治療とランダム化することによって克服することができる。このように、標準治療を含めたランダム化比較をphase IIで行うことは大きな利点をもつが、2群を直接比較するとサンプルサイズが倍以上になってしまう。では、どのようにすればよいのであろうか。以下、Rubinsteinらの論文³⁾をもとに、randomized phase II designについてこれまでに提案されている例をあげ、その特徴を議論してみたい。

これまでに提案された randomized phase II design

1. Randomized phase II selection design

1985年にNCIのSimonら⁴⁾によって提案されたものが有名である。患者を2つ以上の試験治療にランダムに割り付け、もっとも高いresponseが得られた治療が次の研究に進むというデザインである(エンドポイントはほかのものでもよい)。2つの試験治療があるときに優先順位をつけるデザイ

ンであって、もっともよい群が高い確率で選択されることを保証するようにサンプルサイズを決定する。たとえば、真のresponseがレジメンAで35%、レジメンBで20%のとき、90%の確率でレジメンAを選択するためには29名が必要、といった具合である。この場合の問題点は、「どちらも同じ場合には50%の確率でAを選択!!!」することであり、その意味で α エラーが50%ともいえるかもしれない。このデザインはdoseやscheduleを選ぶためにも用いることができる。前提として、毒性の程度がほぼ同じことを仮定しているが、毒性が異なる場合にもその分効果に下駄を履かせたdecision ruleを作ることも可能である。しかし、異なる治療薬の選択の場合には本当にこれでよいのかという不安がある。というのも、2群の治療効果がまったく同じでも、50%の確率でいずれかの群を選んでしまうためである。また、観察されたresponseのみで選択するため、とくに事前にどちらがよいかの根拠がまったくない場合だけ使えるともいえる。さらに、試験の対象となる薬剤が異なる製薬企業によって開発されているものであった場合、製薬企業がよいといえばこのデザインによる優先順位づけはできるかもしれないが、大きなコストをかけて開発が行われていることを考えると、このような単純な判断基準が受け入れられるとは考え難い。

Randomized phase II selection designの別の方法として、患者を2つ以上の試験治療に割り付け、それぞれヒストリカルコントロールと比較するというものもある。群ごとにサンプルサイズ計算をするので、群の数だけphase II studyをやるのと同じサンプルサイズが必要であるが、効果と毒性の両方から判断してよい方を次の段階へ進めるといった判断が下せる。これは、効果と毒性のプロファイルがまったく違う場合に有効と考えられるが、ヒストリカルコントロールとの比較はあまり信用できないため、選んだ薬剤がphase IIIで既存治療に勝る確率が高いという確証にはつながらない。

2. Reference arm phase II design

Hersonら⁵⁾は、試験治療群と標準治療群にランダム化するが、サンプルサイズの増大を防ぐため、直接比較はしないというデザインを提案し

た。標準治療群はヒストリカルコントロールの対象者と試験の対象者が同じ集団かどうか(比較可能かどうか)をチェックするためだけに用い、新治療群はヒストリカルコントロールと比較するというものである。標準治療群の成績がヒストリカルコントロールの成績と変わらなかった場合は、試験治療群とヒストリカルコントロールを比較することに問題ないが、標準治療群が予想より低かったり高かったりした場合はどのように比較するかが問題となる。その場合にはもう一度研究を行えばよいとオリジナルの著者らは言っているが、これは現実的ではないし、もう1回やったからといって比較可能な成績が出るという保証はない。このように、ヒストリカルコントロールと同じかのチェックが失敗すると試験結果が解釈しづらくなるといえる。

3. Phase II/III design

標準治療と新治療でランダム化phase IIをやり、その結果に基づいてそのままphase IIIに進むデザインをrandomized phase II/III designと呼ぶ。いろいろなバリエーションがありえるが、共通していえることとして、同じ対象者を使えるのでサンプルサイズの節約になる一方、phase IIIに進まなかった場合にはphase IIIの準備が無駄になる、といった特徴がある。ここでは、提案されているデザインをいくつか具体的に紹介する。

Schaidら⁹⁾は、標準治療Aとそれと比較したい新治療B、さらにはphase II試験をしたい新治療C1, C2, C3...でランダム化比較試験を行い、Cに失敗したらA or Bにさらにランダム化して組み込むというデザインを提案した。このデザインの特徴は、無治療の患者をphase IIに組み込み、失敗したらmore established therapyであるA or Bへ治療を変更するというものである。しかし、無治療の患者をphase IIに入れることに対する倫理的な問題や、前治療としてのCの影響をどう取り扱うかといった問題があり、あまり現実的でないと思われる。

また、Storerら⁷⁾は、後継phase IIIと同じランダム化をphase II部分で行うというデザインを提案している。オリジナルの点としては、新治療をヒストリカルコントロールと比較し、よければphase IIIへ、よくなければphase IIとして

終了するというものである。Phase IIの段階ではランダム化した標準治療群との比較を行わないので、ヒストリカルコントロールと比較することの問題点は解消されていない。Phase IIIが終わって、ふたを開けてみたら負けていた、ということも十分ありえる。

Ellenbergら⁸⁾は、phase II部分はresponseで比較、少しでも勝っていたらphase IIIへというデザインを提案した。これは、誤ってphase IIで止めてしまう確率が十分低くなるようにして、phase IIIでの検出力低下を防ぐことを意図したデザインである。Phase IIIのサンプルサイズの1/3位、通常のsingle arm phase IIの2倍くらいのサンプルサイズを提案している。このデザインではresponseがprimary endpointである生存期間のよいサロゲートエンドポイントである場合でも、帰無仮説が正しい、すなわち両群に差がない場合にも50%の確率でphase IIIへ進んでしまうので、スクリーニング目的としてはあまり効率がよいとはいえない。

Schaid⁹⁾, Scherら¹⁰⁾は、複数の新治療群があっても行うことができるrandomized phase II/III designを提案した。Phase II部分も標準治療群と真のエンドポイントである全生存期間(overall survival; OS)で比較し、ある規準を超えていればその群のみphase IIIへ進むというもので、phase II部分は全体の半分程度のサンプルサイズという提案を行っている。検定の多重性の問題が生じるため α エラーを増大させないためには、なんらかの統計的方法を組み込む必要があるだろう。

私が所属するJCOG(Japan Clinical Oncology Group)でも、実際のいくつかの試験でphase II/III designを採用している。1種類のデザインを用いているのではなく、試験の内容によりさまざまなデザインを採用している。たとえば、phase IIからphase IIIへ進むための判断規準として、

- ・新治療群の安全性だけを確認してそれが規準を満たしていればphase IIIへ進む。
- ・標準治療群に対し有意に劣っていないことのみ証明できればphase IIIへ進む。
- ・両群別々にヒストリカルコントロールと比べて劣っていないかを検証してphase IIIへ進む。

表 1 PFSをエンドポイントとした場合のランダム化スクリーニング試験に必要な総イベント数

| | 検出したいハザード比 Δ (化療 vs. 化療+分子標的薬) | | | |
|----------------------------------|--|--------------|--------------|---------------|
| | $\Delta=1.3$ | $\Delta=1.4$ | $\Delta=1.5$ | $\Delta=1.75$ |
| $(\alpha, \beta) = (10\%, 10\%)$ | 382 | 232 | 160 | 84 |
| $(\alpha, \beta) = (10\%, 20\%)$ | 262 | 159 | 110 | 58 |
| $(\alpha, \beta) = (20\%, 20\%)$ | 165 | 100 | 69 | 36 |

(文献³⁾より改変)

といったようにphase IIIでの α エラーを消費しない仕組みを組み込んでいる。いますぐにでもphase IIIを行いたい、その前に有効性や安全性の再確認をしておく必要があるといった場合に用いることが多い。そのため、phase II部分で検証的な解析を行うことには重きを置いておらず、phase II部分で α エラーの増大が生じないように注意を払っているといえる。これらは統計的にかなりの応用問題であり、統計家が関与しても厳密には解決しつくしていない部分が残るが、サンプルサイズの節約とphase IIIを計画する時間の節約がcriticalな場合には、採用する余地があると考えられる。

Phase II/III designをまとめると、メリットとして、①phase II部分のデータをphase IIIでも使える、②phase II終了後phase IIIをすみやかに開始できる、③phase II部分のエンドポイントにはflexibilityがある、④標準治療群をconcurrent controlを使うのでphase II段階での比較の妥当性が上がる、などがあげられる。デメリットとして、①スクリーニングツールとしての性能はよくない、②通常のphase IIに比べてphase II部分のサンプルサイズが大きい、③phase IIIを実施するためのインフラと準備が必要、④phase III部分の検出力を保つためにはphase II部分の α (=controlよりたいしてよくないのに間違っphase IIIに進んでしまう確率)が大きくなる、⑤phase II部分で早く判断できるエンドポイントがなければphase IIの結果が出るまで中断しなければならない、といった点があげられる。したがって、phase II/III designはeffective screeningというより、aggressive interim monitoringを

表 2 レスポンスをエンドポイントとした場合のランダム化スクリーニング試験に必要な全体の患者数

| | 検出したいレスポンスの差 (化療 vs. 化療+分子標的薬) | | | |
|----------------------------------|-----------------------------------|-------------|-------------|-------------|
| | 20% vs. 35% | 40% vs. 55% | 20% vs. 40% | 40% vs. 60% |
| $(\alpha, \beta) = (10\%, 10\%)$ | 256 | 316 | 156 | 182 |
| $(\alpha, \beta) = (10\%, 20\%)$ | 184 | 224 | 112 | 132 |
| $(\alpha, \beta) = (20\%, 20\%)$ | 126 | 150 | 78 | 90 |

(文献³⁾より改変)

もつphase IIIと考えた方がよいであろう。

4. Randomized phase II screening design

最近提案されたものにランダム化スクリーニングデザイン(randomized phase II screening design)³⁾と呼ばれるものがある。これは、標準治療と新治療を直接に、しかし決定的ではない形(nondefinitive)で比較するものであり、標準治療 vs. 標準治療への分子標的治療のadd onなどといった比較を念頭において提案されたデザインである。これは、ヒストリカルコントロールとの比較による選択バイアスの影響を避けるために標準治療群とランダム化し、かつ、試験の規模は2群で50~100名くらいのサンプルサイズにしたいという欲求を満たすものである。そのためには、治療効果の差に対して α エラー、 β エラーを注意深く設定することによってサンプルサイズ設計を行うことが必要となる。また、決定的な結論を導くことのできるphase III試験ができなくなることがないように、OSとは異なる、より早期のエンドポイントを使うなどといった工夫が必要である。たとえばPFSやresponse、%PFSなどといったものである。ちなみに、ある時点で全員のprogressionを測定できる場合には、その時点での%PFSの方が観察間隔の影響を受けやすいPFSに比べて観察によるバイアスが少ないといえる。

PFSをエンドポイントとした場合のランダム化スクリーニング試験に必要な患者数を表1、responseをエンドポイントとした場合のランダム化スクリーニング試験に必要な患者数を表2に示す。 $\alpha > 20\%$ 、 $\beta > 20\%$ は薦められないので、それ以下で考えても、両群合わせて100名程度あ

るいはそれ以下で設計できる組み合わせがある。つまり、治療効果の差に対して、 α , β を注意深く設定することによって100以下のサンプルサイズ設計が行えるケースがある。たとえば、治療効果をハザード比で表した表1より、 $\Delta > 1.4$ であれば α , β がともに20%の場合必要イベント数は100以下となる。 α が大きすぎるとスクリーニング能力がない、つまりphase II を実施せずにphase I から一足飛びにphase III を実施することと同じになってしまう、逆に β が大きすぎると有効な薬剤の開発を誤って中止してしまうことになる。過度に楽観的な治療効果の差 Δ を考えることも β を大きくすることと同じである。これらから、スクリーニングにはHR=1.5(レスポンスでは差が20%)に対して $\alpha=\beta=0.2$ 程度が一つの目安といえるだろう。

ランダム化スクリーニング試験の他の特徴として、同じ薬剤を用いて実施されるスクリーニング試験の数が多くなるとfalse positiveの確率が高くなることがあげられる。たとえば、同じ薬剤を用いて異なる3つの対象に試験を実施したとすると、 $\alpha=\beta=0.2$ の場合には、その薬剤が3つの対象すべてに治療効果がある場合に1つも治療効果があることを見出せない確率が5%以下となる一方、3つの対象すべてに治療効果がない場合に少なくとも1つの対象で効果ありと判断してしまう確率は約50%となってしまう。

ランダム化スクリーニングデザインでの重要な点は、得られた結果がdefinitive phase III から得られたものとは違うことをしっかり意識するという点である。とくにOSをエンドポイントにしているときは、研究者がsubstantial evidence と思ってしまうことが問題である。そのため、PFS, response rate, %PFSといった、できるだけOSと異なるエンドポイントを用いることも必要であろう。このようなデザインで実施された試験結果をdefinitive phase III と考えてはならないのは、何度もランダム化スクリーニング試験を実施する状況では検定の多重性の問題と同様の状況が生じ、 $\alpha=0.05$ が判断の基準として大きすぎるためである。Phase III をやるにはかなりの証拠が必要だが、phase II はそうでもない。したがって、通常のphase II-phase III sequence

と同じような証拠を与えるためには、phase III の中間解析と同じように考えて、たとえば $\alpha < 0.005$ などといった極端な結果が得られた場合にのみsubstantial evidenceと考え、そのあとのphase III を行う必要がないといった判断の方針を採用するのがよいかもしれない。また、サンプルサイズが小さいことを考えると、有意な結果が得られたとはいっても大きな治療群間差が大きな信頼区間を伴っているはずであり、より正確な推定値、すなわちより正確な治療効果の見積もりを得るためにも、さらなる研究が必要といえる。

具体的に考えてみよう。ランダム化スクリーニングデザインが使えるシナリオとして、

(1)ある製薬企業は、多くの臓器で標準化療に対してadd onできるような分子標的薬Xを開発中であるが、薬剤Xは単剤ではinactiveである。しかし、薬剤X+標準化療をヒストリカルコントロールと比較するsingle arm phase II はあまり信頼できない。このような場合には、複数のスクリーニング試験を行うというストラテジーが提案できるだろう。それぞれの臓器で実施する一つ一つの試験に対して比較的大きな β エラーを設定しても、どれかの試験では差を検出できる可能性が高まるからである。そのほかのシナリオとして、

(2)研究者はある特定の疾患の標準化療にadd onできるような分子標的薬 Y_1, Y_2, \dots に優先順位をつけなければいけないが、single arm phase II は毒性やfeasibility程度のデータしか与えないし、薬剤の開発状況の違いや製薬企業のreluctanceによって、複数の分子標的薬のrandomized selection designはやり難い。そこで、複数のスクリーニング試験を行うというストラテジーが提案できるかもしれない。この場合も、それぞれの試験に対する比較的大きな β エラーが許容できるからである。

まとめ

本稿ではさまざまなrandomized phase II design について議論してきた。最後にphase II でどのデザインを選ぶかを簡単にまとめてみたい。

①Responseで評価が可能な場合はconventional

phase II design.

- ②複数の試験治療に優先順位をつけたい場合randomized selection phase II design.
- ③次にphase IIIに進むことがほぼ確実な場合はphase II/III design.
- ④ヒストリカルコントロールとの比較が行いにくい場合はrandomized screening design.

とくに新規分子標的薬を標準治療にadd onする場合には, definitive phase IIIを実施可能でかつ研究者が治療効果と α エラーと β エラーをバランスと分別をもって設定できるのなら, ランダム化スクリーニングデザインはよいデザインといえるだろう.

もちろん, デザインが先にあるわけではなく, 治療開発全体のストラテジーの中で適切なデザインを選ぶということが基本である. 最近randomized phase IIが多く議論されているということは, 数少ないが一発逆転狙いのcytotoxic drugの開発という時代から, 単剤での治療効果は必ずしも高くないが, 分子標的薬という数多くの候補薬剤の中から有効なものを選んで開発していくという時代へと世の中が変わっていることを意味している. これは製薬会社や臨床研究者のみならず, 臨床試験に携わる統計家にとっても大きなチャレンジといえる. ぜひ協力してよい治療を開発していきたいと思う.

文 献

- 1) European Organisation for Research and Treatment of Cancer. Phase II trials in the EORTC. The Protocol Review Committee, the Data Center, the Research and Treatment Division, and the New Drug Development Office. Eur J Cancer 1997 ; 33 : 1361.
- 2) Van Glabbeke M, Steward W, Armand JP. Non-randomised phase II trials of drug combinations : Often meaningless, sometimes misleading – Are there alternative strategies? Eur J Cancer 2000 ; 38 : 635.
- 3) Rubinstein LV, Korn EL, Freidlin B, et al. Design issues of randomized phase II trials and a proposal for phase II screening trials. J Clin Oncol 2005 ; 23 : 7199.
- 4) Simon R, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. Cancer Treat Rep 1985 ; 69 : 1375.
- 5) Herson J, Carter SK. Calibrated phase II clinical trials in oncology. Stat Med 1986 ; 5 : 441.
- 6) Schaid DJ, Ingle JN, Wieand S, et al. A design for phase II testing of anticancer agents within a phase III clinical trial. Control Clin Trials 1988 ; 9 : 107.
- 7) Storer BE. A sequential phase II/III trial for binary outcomes. Stat Med 1990 ; 9 : 229.
- 8) Ellenberg SS, Eisenberger MA. An efficient design for phase III studies of combination chemotherapies. Cancer Treat Rep 1985 ; 69 : 1147.
- 9) Schaid DJ, Wieand S, Therneau TM. Optimal two-stage screening designs for survival comparisons. Biometrika 1990 ; 77 : 507.
- 10) Scher HI, Heller G. Picking the winners in a sea of plenty. Clin Cancer Res 2002 ; 8 : 400.

* * *

特集

臨床試験

第 I 相試験*

山本 精一郎**

Key Words : phase I trial, non-cytotoxic drug, cancer study design

臨床試験のデザイン概説

臨床試験はスクリーニングである。薬剤であれば非臨床段階を経て、臨床導入から徐々に安全性や有効性、実施可能性などを調べながら、多くの候補治療がそれぞれの段階でふるい落とされる。最終的に標準治療と決勝戦を行うことによって、新しい標準治療として取って代わる治療を見つけることが臨床試験を実施するもっとも大きな目的である。臨床開発の段階は「相(phase)」としてとらえられることが多いが、これはスクリーニングである治療法の開発段階を便宜的に「呼び名」として、逐次的な複数の相に分けただけである(ICH-E8)¹⁾。慣用としてよく用いられているものの、開発の相という概念は臨床試験の分類の基礎としてふさわしくなく、試験の目的による分類のほうが望ましい。ICH-E8では、臨床試験は、その試験の臨床開発期間における実施時期および目的によって、臨床薬理試験、探索的試験、検証的試験、治療的使用のように分類可能だとしている。Piantadosiの提案する分類も、これに準ずるものである²⁾。つまり、臨床導入から標準治療になるまでには必然的にそれぞれ満たすべき・調べるべき要件があり、

それを調べるのが各試験の目的となる。たとえば、まずは安全性をみながら用量反応研究を行い、推奨用量が決まったら、その用量で探索的に効果を検討し、標準治療以上の効果が期待できそうであれば、次に標準治療とのランダム化比較試験を行う。めでたく標準治療に勝れば、新しい標準治療となる。目的が決まれば、それに応じてデザインは決まってくる。とはいえ、相という概念は慣習的に用いられ、今となつてはもともとの「段階」以上に内容を伴う用語として用いられることも多い。本稿の目的は第 I 相試験に用いられるデザインを紹介することであるが、本筋に立ち戻り、臨床導入第一段階で必要なこと、すなわち安全性を検討して、その後の開発用量を決定するという観点からデザインについて考えてみたい。

第 I 相臨床試験の標準的な臨床試験デザイン

人間に新しい治療(候補)を行う最初の段階では、それがどのような有害事象(種類と程度)をもたらすかを調べる必要がある。また、どのくらいの治療強度(薬剤の投与量など)がもっとも治療効果をもたらすかも調べる必要がある。効果や有害事象、適した治療スケジュールを調べるためには、薬理作用も調べる必要がある。これらを調べる試験を第 I 相試験(phase I trial)と

* Study design of phase I clinical trial for non-cytotoxic drug.

** Seiichiro YAMAMOTO, Ph.D.: 国立がんセンター がん対策情報センター がん情報・統計部(〒104-0045 東京都中央区築地5-1-1); Center for Cancer Control and Information Services, National Cancer Center, Tokyo 104-0045, JAPAN

呼ぶ。最初にヒトに薬剤を投与する場合だけでなく、複数薬剤の併用のレジメンを検討する場合にも同様の手順が必要となる。単剤の場合、毒性と効果が比例していると考えられるcytotoxic drugの場合には、耐えられる最大の投与量が最大効果をもたらすという考え方から、想定される毒性(用量制限毒性, Dose limiting toxicity : DLT)に対して、最大耐用量(Maximum Tolerated Dose : MTD)を調べ、推奨用量を決定する試験を第I相試験として行うことが一般的である(簡便のために、本稿ではこの考え方を「MTDに基づく用量探索」と呼ぶことにする)(図1)。新治療(候補)をヒトに初めて行うという性質上、薬理作用や有害事象などを逐次モニターし、それをすぐに試験の進捗に反映させる必要があるため、少数の施設(できれば単施設)で行うことが望ましい。

単剤のcytotoxic drugを例に、標準的な第I相臨床試験デザインを紹介する。わが国の抗悪性腫瘍薬の臨床評価方法に関するガイドライン³⁾をはじめ、海外の教科書⁴⁾に紹介されているデザインは、以下のものである。まず、原則として、マウスに対する10%致死量の用量レベル(LD10)の1/10を試験開始用量として各用量を定め、低い方から順にそれぞれの用量を一定数の患者(多くの場合3例)に投与する。もし、ある用量で初めの3例に対しDLTが現れなければ次の用量へ進み、DLTが現れればさらに3例に対し同じ用量を投与する。結果として6人中1人までの毒性発現なら次の用量レベルへ進み、6人中2人以上ならそこで試験を終了し、その用量をMTDとする方法である。増量幅に関する用量レベルの設定はModified Fibonacciの方法が用いられることが多かった。この方法は、6人中2人で終了するため、直感的には33%の毒性が出る用量をMTDとしているように思えるが、実際は33%の毒性が出るより低い用量で止まることが多い。33%発現用量より低い用量でも6人中2人が発現することは十分あるからである。慣習的にこの方法が用いられることが多いが、ストップリングール、増量幅ともにとくに科学的な根拠はなく、許容できる毒性の程度を変更できない、試験開始用量が低すぎる場合には効果のまった

くない用量を多くの患者に投与してしまう、といった問題がある。これに対し、まず用量反応曲線を想定し、それを逐次推定した上で、曲線から推定される推奨用量付近(あるいは試験実施者が決定した許容できる毒性に対応する用量付近)を投与し、これを繰り返すBayes流の方法(Continuous Reassessment Method : CRM)が提案され⁵⁾、利用されている場合もあるが、いまだそれほど一般的とはなっていない。Cytotoxic drugの第I相試験で用いられるデザインである3例コホート、CRMともにMTDに基づく用量探索という考え方に基づいている。

分子標的薬の開発に 現在用いられているデザイン

ところが、近年では分子標的薬などのさまざまな作用機序の新薬が開発の中心となってきた。単剤のcytotoxic drugが開発の中心であった時代には、MTDに基づく用量探索が第I相試験の役割であった。MTDが探索できると、MTDあるいはその下の用量を至適用量(optimal dose)として、プライマリエンドポイントを腫瘍縮小効果(たとえば奏効割合)などとした第II相試験を行う。さらに、腫瘍縮小効果としてよい成績を残した薬剤を用いて、標準治療を対照とした第III相試験を行うというのが標準的な開発ステップであった。

これに対し、分子標的薬は毒性と効果が必ずしも比例しないため、第I相試験でMTDに基づく用量探索という考え方をを用いることが最適とはいえない。しかも、これらの薬剤は必ずしも腫瘍縮小効果が高くないため、第II相試験において腫瘍縮小効果をエンドポイントとすることも常に最適とはいえない。このような場合、全生存期間や無増悪生存期間などをエンドポイントとして第II相試験を行うこともあるが、これらのエンドポイントは薬剤の効果だけでなく、試験に登録される患者によって大きく異なってしまう。つまり、用量選択が適切でなかったり、延命効果のない薬剤であっても、同じ適格規準の中で比較的予後の良い患者を登録することによって、よりよい成績を得てしまう場合がある。恣意的でなくてもこのようなことが生じてしまえ

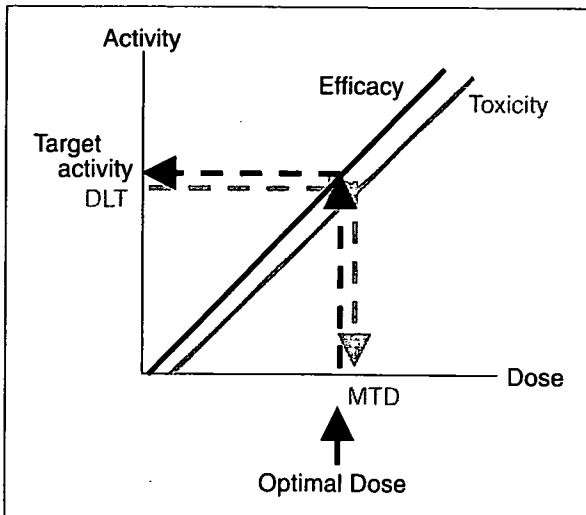


図1 用量-毒性-効果の関係(毒性と効果が比例)

ば、その後の第III相試験はnegative resultsに終わってしまう。これらの薬剤に対する第II相以降の臨床試験デザインは別稿に譲るが、臨床試験のデザインを考える上で、開発のすべての段階において、cytotoxic drugとは異なるポイントを考慮する必要がある。

第I相試験のデザインに話を戻そう。分子標的薬の開発のためには、MTDに基づいた用量探索に基礎を置いたこれまでの第I相試験のデザインと同じデザインを用いることができないのか。もしそうなら、どのようなデザインがよいのであろうか。残念ながら、これまでに実現可能性も含めて分子標的薬開発に対するベストなデザインはいまだ提案されていない。NCI sponsored trialのプロトコルをすべてレビューしている米国National Cancer InstituteのClinical Trial Evaluation Program (CTEP)の統計家に尋ねたところ、毒性と効果が必ずしも比例しないような場合でも、大多数の第I相臨床試験で、これまでと同様MTDに基づいた用量探索に基づく3例コホートを用いたデザインが用いられているとのことであった(personal communication)。

そこで、分子標的薬開発において、3例コホートに代表されるMTDに基づく用量探索法に基づくデザインを用いた場合にどのようなことが生じるのか検討してみる。MTDに基づく用量探索法によってOptimal doseが正しく得られるためには以下の条件が満たされていることが必要である⁶⁾。

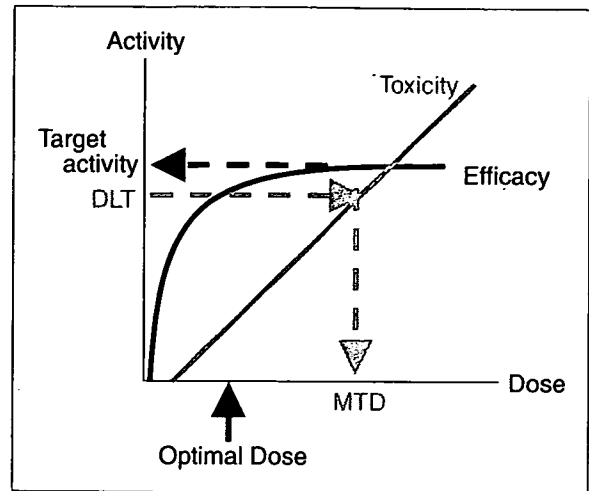


図2 用量-毒性-効果の関係 (MTDより前に効果がプラトー)

- 1) 用量が増加すれば、臨床的なベネフィット(効果)が増加する。
- 2) 用量が増加すれば、毒性が増加する。
- 3) 毒性が許容範囲で、効果が得られる用量が存在する。

最近開発が進んでいる分子標的薬では、1)に関して、用量-効果曲線にプラトーがあり、必ずしも高ければ高いほど臨床的なベネフィットが大きくない、あるいは少なくとも用量を上げても少ししか効果が増加しないという場合が考えられる。2)に関して、用量が増加しても必ずしも毒性は増加しない、あるいは毒性が発現するのは十分な効果がみられる以上の用量である場合もある。いくつか具体的に例をあげて考えてみよう。

まず、毒性は用量依存的に増加するが、効果はMTDよりも前にプラトーに達する場合である(図2)。この場合、MTDから推定した用量はsuboptimalといえる。MTDよりも前の時点から効果はプラトーに達しているため、より低い用量のほうが毒性に対する効果のバランスが高いためである。しかし、この場合でもsuboptimalな用量が得られるという意味において、MTD探索デザインはある程度有用であるといつてよいだろう。次に、毒性は用量依存的に増加するが、効果はMTDよりも後にプラトーに達する場合である(図3)。この場合、MTDまでの用量で効果が用量依存的に増加するならMTDに基づく用量

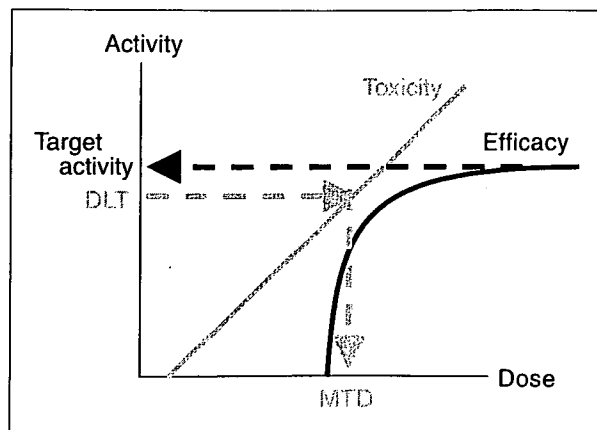


図3 用量-毒性-効果の関係 (MTDより後に効果がプラトー)

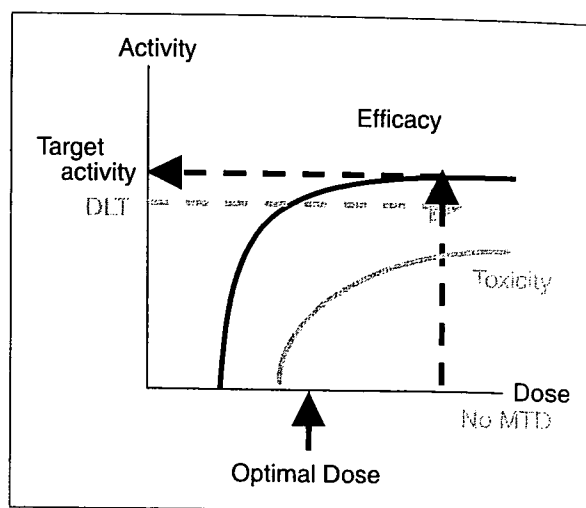


図5 用量-毒性-効果の関係 (毒性がプラトー, 効果もプラトー)

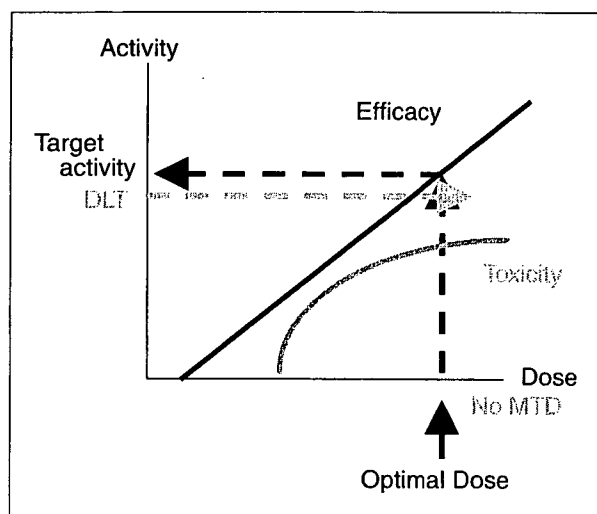


図4 用量-毒性-効果の関係 (毒性がプラトー, 効果が直線的に増加)

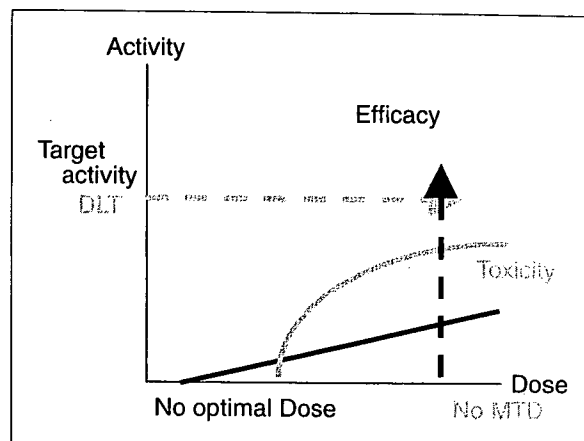


図6 用量-毒性-効果の関係 (毒性がプラトー, 効果が十分でない)

探索法によってoptimal doseを得ることができるが、効果がtarget activityに達しない場合(効果が出始めるのがMTDよりも高い用量の場合など)には上記3)の条件を満たさないことになり、optimal doseを見つけることはできない。第I相試験でも効果をモニターできれば効果についてある程度情報が得られるであろう。この場合、よくデザインされた第II相試験を行うことができれば、やはりその結果で開発を中止することになるであろう。

効果が用量依存的に増加し、毒性がプラトーに達するような場合を考えてみよう(図4~6)。この場合、毒性によるMTDが見つけられないので、それ以外の条件で用量増加を中止することになる。効果を理想的にモニターすることがで

ければ、効果がTarget activityに達した用量を見つけることができるが、そうでない場合には、用量-効果曲線がTarget activityを越えて直線的に増加する場合(図4)を除いて、optimalな用量を選択できない。たとえば、用量がtarget activityに達した後プラトーになるような場合には、効果をみないで選択するとsuboptimalな用量を選択することになる(図5)。この場合はまだよいが、効果がTarget activityに達しないような場合には、どのような用量を選んでもoptimalな用量とはいえない(図6)。この場合も、よくデザインされた第II相試験を行うことができれば、その結果で開発を中止することになるであろう。

これらの検討からわかることは、毒性と効果が必ずしも比例関係になくても、ある程度MTD

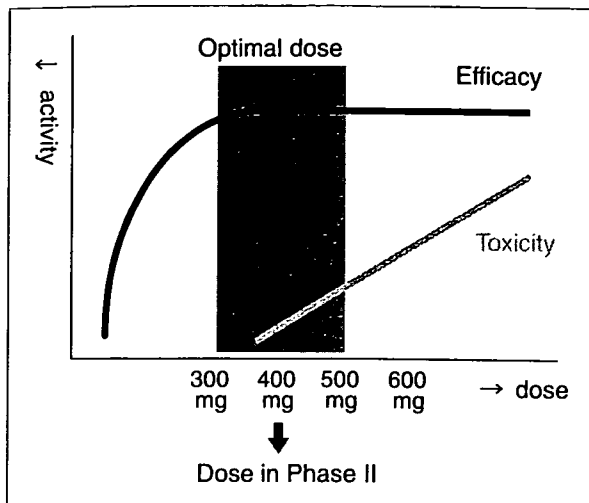


図7 Imatinib trialでの用量-毒性-効果の関係

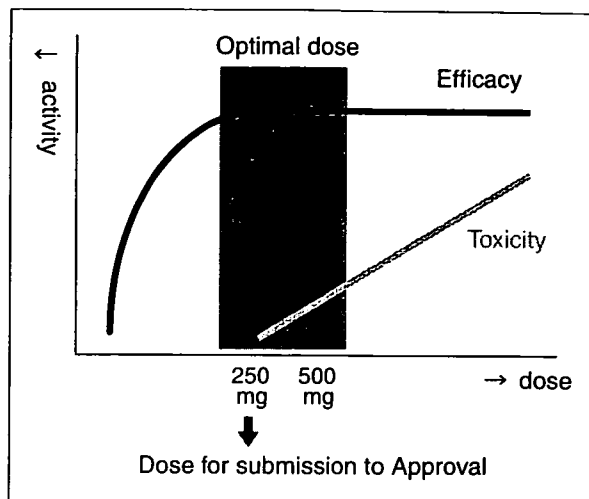


図8 Gefitinib trialでの用量-毒性-効果の関係

に基づく用量探索法が有効であることである。しかしながら、いくつかの場合においてうまく働かない場合がある。その場合でも効果をうまくモニターできればoptimalに近い用量を探索することができる。いくつか実例を示そう。図7はimatinib trialの第I相試験における用量-毒性-効果の関係を模式的に表したものである⁷⁾。1,000mgまで用量を上げて毒性はMTDの規準に達せず、効果(Hematologic response)はその用量の範囲内でプラトーに達しており、phase 2では400~600mgを中心として検討が行われた。図8はgefitinibに対する臨床試験であるIDEAL1における用量-毒性-効果の関係を模式的に表したものである⁸⁾。IDEAL1は第I相試験でなく第II相試験である。効果(response)は250mg/dayで18.4%、500mg/dayで19.0%とあまり変わらなかったのに、Grade 3/4の毒性は500mg/dayで多く発現した。結果として(もちろんそれだけが理由ではないだろうが)、日本における申請用量として250mg/dayが用いられた。

しかしながら実際の問題として、各用量に数例しか対象者がいないような第I相試験で効果をモニターするのは難しい。IDEAL1のように毒性プロファイルがある程度わかれば効果も調べられるような第II相試験へ早めに進むか、第I相試験の対象者数をもう少し増やすか、第I相試験のPKやPDなどからtarget inhibitionの情報を得ることによって効果のあたりをつけて複数の候補用量の選択を行う、などいろいろな方法が考

えられるが、単純かつ決定的な方法は現在までのところ提案されていない。

まとめに代えて

研究デザインは研究の目的が決まれば、おのずと決まると説明した。まず、第I相、第II相、第III相というデザインがあるというのではなく、治療法開発の段階に従って調べるべき目的があり、それに合わせてデザインをたてることになる。毒性と効果が比例するcytotoxic drugでは比較的単純であった各段階での研究デザインに対し、毒性と効果が必ずしも関連しないといわれている分子標的治療薬の評価はその応用問題といえる。分子標的薬に対して、ベストと考えられる第I相試験のデザインがない現在、基本的にはこれまでに十分経験を積んだMTDに基づく用量探索方法に基づくデザインをもとに、効果を見る工夫を盛り込んだ開発デザインを採用する必要がある。効果と毒性の両方が測定できるような場合には、統計的な定式化によりいくつかのデザインが考えられるが(たとえばHunsburger S. 2005⁹⁾)、それぞれ検証の難しい仮定を含んでおり、たとえ両方がわかる場合にでも、まだまだ実際の適用場面とのすり合わせが必要と思われる。実際には、薬剤と腫瘍の性質によって最適なデザインを頭を絞って考え、結果に対して慎重に検討し、思い切って判断するという時代が当分続くことが予想される。第I相、第II相、第III相それぞれで何をするとするのが決まっ

いると考えるよりも、ヒトへの導入段階から標準治療とのマッチアップまでをトータルで考えた開発戦略がより有効となるであろう。

文 献

- 1) ICH-E8 臨床試験の一般指針について. ICHガイドラインホームページ. (<http://www.nihs.go.jp/dig/ich/eindex.html>)
- 2) Piantadosi S. Clinical Trials. A Methodologic Perspective. John Wiley & Sons, Inc ; 1997.
- 3) 抗悪性腫瘍薬の臨床評価方法に関するガイドライン. (<http://www.hourei.mhlw.go.jp/hourei/doc/tsuchi/171101-b.pdf>)
- 4) Crowley J. Handbook of statistics in clinical oncology. Marcel Dekker ; 2001.
- 5) Ishizuka N, Ohashi Y. The continuous reassessment method and its applications : a Bayesian methodology for phase I cancer clinical trials. *Statistics in Medicine* 2001 ; 20 : 2661-81.
- 6) Hunsberger S, Rubinstein LV, Dancey J, et al. Dose escalation trial designs based on a molecularly targeted endpoint. *Stat Med* 2005 ; 24 : 2171-81.
- 7) Druker BJ, Talpaz M, Resta DJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 2001 ; 344 : 1031-7.
- 8) Fukuoka M, Yano S, Giaccone G, et al. A multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small cell lung cancer (The IDEAL 1 Trial). *J Clin Oncol* 2003 ; 21 : 2237-46.

* * *

V. 資 料

資料 ①

| | | | |
|-------|---|-------|--------------------|
| 研究課題名 | 「切除不能または再発胃癌に対する5-FU+I-LV+Paclitaxel併用療法の第Ⅰ／Ⅱ相試験」 | | |
| 研究分類 | PhⅠ/Ⅱ | | |
| 研究代表者 | 島田 安博 | 所属 | 国立がんセンター中央病院 消化器内科 |
| 研究担当者 | 松原 淳一 | 所属 | 国立がんセンター研究所 |
| 施設数 | 8施設 | 予定症例数 | 35例 |
| 研究期間 | 倫理審査承認後より1年6ヶ月(STEP2登録開始後より1年) | | |

H19年度データマネジメント報告

1 実施内容

- ① CRFクリーニング(CRF23枚)
- ② クエリー作成
- ③ データ入力
- ④ データ集計

2 実施件数

- ① PhaseⅠについて
19例/19例 (L1:L2:L3:L4 / 3:3:6:6(1例登録直後不採用))
PhaseⅡについて
5例/35例
- ② PhaseⅠについて
全19例
PhaseⅡについて
5例/35例
- ③ PhaseⅠについて
全19例
PhaseⅡについて
5例/35例
- ④ 17票作成
学会発表用データ集計
癌治療学会(2007/10)
論文作成用データ集計
2007/11及び2008/01

3 試験進捗

2007/6/11よりPhaseⅡの登録を開始
RDはLevel3(5-FU/I-LV/TXL=600/250/80)

-以上-

| | | | |
|-------|--|-------|--------------------|
| 研究課題名 | 臨床病期Ⅱ/Ⅲ食道癌(T4を除く)に対する50.4Gy, modified 5-FU+CDDP併用化学放射線療法 の臨床第Ⅱ相試験 | | |
| 研究分類 | PhⅡ | | |
| 研究代表者 | 加藤 健 | 所属 | 国立がんセンター中央病院 消化器内科 |
| 研究担当者 | 中島 貴子 | 所属 | 国立がんセンター中央病院 消化器内科 |
| 施設数 | 9施設 | 予定症例数 | 50例 |
| 研究期間 | 登録期間:2年、追跡期間:登録終了後3年、総研究期間:5年 | | |

H19年度データマネジメント報告

1 実施内容

- ① CRFクリーニング(CRF11枚)
- ② クエリー作成
- ③ データ入力
- ④ データ集計

2 実施件数

- ① CRFクリーニング
49例/50例
- ② クエリー作成
49例/50例
- ③ データ入力
49例/50例
- ④ 25票作成
 班会議用データ集計
 2007/05
 2007/10
 次期プロトコル検討のためのデータ集計
 2008/01
 2008/03

3 試験進捗

49症例の登録(のこり1例)

-以上-

| | | | |
|-------|--|-------|--------------------|
| 研究課題名 | 「胃癌術後補助化学療法としてのS-1+CDDP併用療法のfeasibility試験」 | | |
| 研究分類 | Ph II | | |
| 研究代表者 | 濱口 哲弥 | 所属 | 国立がんセンター中央病院 消化器内科 |
| 研究担当者 | 高張 大亮 | 所属 | 国立がんセンター中央病院 消化器内科 |
| 施設数 | 5施設 | 予定症例数 | 50例 |
| 研究期間 | 登録期間:6ヶ月、追跡期間:登録終了後1.5年、総研究期間:2年 | | |

H19年度データマネジメント報告

1 実施内容

- ① CRFクリーニング(CRF23枚)
- ② クエリー作成
- ③ データ入力
- ④ データ集計

2 実施件数

- ① CRFクリーニング
25例/50例
- ② クエリー作成
25例/50例
- ③ データ入力
25例/50例
- ④ 9票作成
学会用データ集計
2008/02

3 試験進捗

25症例の登録時点で一時登録を中止

-以上-

| | | | |
|-------|--|-------|--------------------|
| 研究課題名 | 切除不能進行・再発大腸癌初回化学療法患者を対象とした、フルオロウラシル/イリコボリン、オキサリプラチン、ペバシズマブ併用療法 (mFOLFOX6+BV) stop and go臨床第II相試験 | | |
| 研究分類 | Ph II | | |
| 研究代表者 | 加藤 健 | 所属 | 国立がんセンター中央病院 消化器内科 |
| 研究担当者 | 沖田 南都子 | 所属 | 国立がんセンター中央病院 消化器内科 |
| 施設数 | 13施設 | 予定症例数 | 50例 |
| 研究期間 | 登録期間:2007年8月より8ヶ月、追跡期間:1年間、総研究期間:1年8ヶ月 | | |

H19年度データマネジメント報告

1 実施内容

- ① CRFクリーニング (CRF17枚)
- ② クエリー作成
- ③ データ入力
- ④ データ集計

- ⑤ プロトコルレビュー
- ⑥ CRF作成

2 実施件数

- ① CRFクリーニング
18例/50例
- ② クエリー作成
18例/50例
- ③ データ入力
18例/50例
- ④ データ集計
未定
- ⑤ プロトコルレビュー
レビュー後プロトコル制定
- ⑥ CRF作成
作成CRFはIRB通過

3 試験進捗

試験実施中

-以上-

| | | | |
|-------|--|-------|----------------------|
| 研究課題名 | Clinical Stage 0-I期直腸癌に対する腹腔鏡下手術の妥当性に関する第II相試験 | | |
| 研究分類 | Ph II | | |
| 研究代表者 | 渡邊 昌彦 | 所属 | 北里大学医学部外科 |
| 研究担当者 | 山本聖一郎 | 所属 | 国立がんセンター中央病院 第1領域外来部 |
| 施設数 | 40施設 | 予定症例数 | 350例 |
| 研究期間 | 登録期間:3年、追跡期間:登録終了後5年、総研究期間:8年 | | |

H19年度データマネジメント報告

1 実施内容

- ① CRFクリーニング(CRF13枚)
- ② クエリー作成
- ③ データ入力
- ④ データ集計

⑤ 事務局業務の一部

- ⑥ プロトコルレビュー
- ⑦ CRF作成

2 実施件数

- ① CRFクリーニング
7例/350例
- ② クエリー作成
7例/350例
- ③ データ入力
7例/350例
- ④ データ集計
未定

⑤ 事務局業務の一部
症例登録後のCRF等発送

- ⑥ プロトコルレビュー
レビュー後プロトコル制定
- ⑦ CRF作成
作成CRFはIRB通過

3 試験進捗

試験実施中