

である。DCTD の治療開発の管理・推進方法に関してはすでに述べたが、DCTD は同様の方法で診断法開発の管理・推進も行っている。また、がん予防とがんコントロール方法の開発は DCP が DCTD と同様の方法を用いて管理・推進を行っている。イソフラボン抽出物などを用いたがん化学予防薬の開発や、支持療法の開発などがそのターゲットとなる。地域がん登録といった国家的規模で行わなければならない記述疫学研究の管理や推進、その他の疫学研究の管理、推進は Division of Cancer Control and Population Sciences (DCCPS) がこれを担っている。基礎研究の extramural research は Division of Cancer Biology (DCB) が担当する。

Office of the director は office of communication (Physicians Data Query といった治療ガイドライン作りや臨床試験情報のデータベースを作ったり、それらを発信する Cancer Information Service がある部門) などのように、基礎・予防・臨床といった従来からの研究の枠に収まりきれない分野の extramural research を担当する所長直轄部門である。

3.2.2 NIH の名の下での研究費の管理

Extramural division の仕事は、一言で言うと研究費の管理及び執行であり、NIH の名の下でがん関係の extramural 研究費を実質的に管理している。NCI 全体の研究費は約 5000 億円であるが、9 割が extramural division と office of the director で管理され、1 割が後述の intramural division で用いられる。といっても、研究費の採択は NIH のセンターの一つである Center for Scientific Review (CSR) が NIH 全体を通して担当しており、現役

研究者どうしの厳密な peer review で決められるので、extramural division は研究費採択の評価をする部門ではない。

NIH からの研究費にはいくつかの種類があるが、大まかに言うと研究費は 3 種類の形で研究者に対して募集される。

3.2.3 NCI のグラント

一つは grant というタイプで、研究者が自由に課題を考え、応募するものである。DCB が管理する基礎研究は grant 型のものが多く、研究者の自由な発想によって研究が行われるものが中心である。さまざまな着想から自由に研究を行うことによって、中にはノーベル賞として評価されるような大きな発見に結びつくものが生まれ、科学の裾野を広げる役割を持つ。

3.2.4 NCI のコントラクト

もう一つは contract と呼ばれるもので、NIH が内容まで決定して研究者を募集するものである。前述した中では、DCTD/DTP が管理する抗癌剤候補物質のスクリーニングや保管といったものや、office of the director が管理するがんに対する情報をコールセンターを通じて患者さんに発信する Cancer Information Service などが contract の例となる。研究という側面もあるものの、目的が明確で方法が定型化しており、事業的側面の強いものを NCI が研究者に委託する。

3.2.5 NCI の Cooperative Agreement

Grant と contract の中間が Cooperative Agreement (CA) と呼ばれるものである。CA は、NCI と研究者が相互に援助を行うという関係を反映した研究費であり、研究の実施に当たり、NCI が内容

Table 1 NCIの研究費の性質の違い

Grant	A financial assistance mechanism providing money, property, or both to an eligible entity to carry out an approved project or activity. Performance responsibility rests primarily with the recipient, and there is little or no Federal involvement or participation in the performance of activities.
Cooperative Agreement	An award instrument, reflecting an assistance relationship between the NCI and a recipient, in which substantial NCI programmatic involvement is anticipated during performance of the activity.
Contract (Research & Development)	An instrument used by the NCI to procure cancer research services and other resources needed by the principal purpose of the transaction is to acquire a specific service or end-product for the direct benefit of or use by the NCI. Contract (under a grant): A written agreement between a grantee and a third party to acquire routine goods or services.

的に実質的な関与を行うものを言う (Table 1)³⁾。DCTD/CTEP が管理する多施設共同臨床試験に対する研究費は CA として出ており、データセンター、オペレーションズオフィス、研究者グループセットで申請すること、がん治療の臨床試験を行うことといったことが要件として研究申請を募集する。第III相臨床試験やがん予防・がんコントロール研究を行うネットワークである CCOP も CA である。つまり、新治療の開発のように、明確なゴールがあり、それに至るには独自の発想も必要だが、方法論的にはある程度確立されていて実績や経験が必要なものは、CA 型の研究費で募集する。研究者 (グループ) の採択は peer review によって行うが、実際の研究計画については、前述のように NCI の DCTD/CTEP でコンセプトやプロトコル審査を行い、他のグループとの棲み分けや試験の計画や進捗のモニタリング、監査などを行うことによって、全米規模で効率よく治療開発が進むように管理・推進を行っている。

実際に筆者が1年半過ごした感覚では、NCI は決してプレーキをかけるのではなく、より早く目的地に着くようにハンドルを切ったりガソリンを入れたりしているという印象である。我が国では、mission-oriented 型、top down 型として、NIH の contract 型の研究費が紹介されることが多いようだが、contract 型で研究費を出してしまうと、臨床試験のように組織の特徴や性格、課題に合わせてプロトコルの内容を精緻化していく場合には対応しにくい。CA のように、研究費募集の際に詳細を決めるのではなく、ある程度の mission と実行可能な組織形態を確認するような募集を行い、その後の進捗管理の中でサポートと管理を行うほうが適していると考えられる。NCI の extramural division は grant, contract, CA といった仕組みを巧みに使うことにより、全米のがん研究を管理・推進・方向付けているといえる。

3.2.6 NCI のサイエンティフィックレビューアー

NCI/NIH のように研究費を管理する機関は funding agency (FA) と呼ばれるが、NIH は米国における生命科学に関するほぼ唯一最大の FA であり、基礎研究から実際の治療の開発まで、生命科学に関する非常に広い分野をカバーし、管理・推進を行っている。研究費を管理する NCI や NIH の extra-

mural division のスタッフは一般的に、Program Officer (PO)、Program Director (PD) と呼ばれ、科学者出身の行政官であるが、その役割は多岐にわたる。CSR で Peer review の管理を行うものもいれば、NCI などの各機関で配分などお金の管理を担当するものもある。中でも、前述の CTEP の例のように治療開発といった目的のために、研究課題を設定し、研究者を選定、プロトコルレビュー、進捗管理をするといった PO/PD は scientific reviewer と呼ばれ、個々の研究者よりもその分野に対する広い知識と視野、先見の明が必要である。このように自分では研究を行わない専門官が長きにわたり担当することによって長期的ながん治療の開発を推進している。わが国には、専門官として永続的に scientific reviewer の役割を明示的に担う職種がない。科学者出身の行政官の育成とキャリアパスの確立は一朝一石には進まないであろうから、scientific reviewer 的な役割を担う科学者と研究費管理に精通した行政官との協力を組織的に行うことが研究費の効率的な運営には不可欠であろう。

3.3 NCI の intramural division

Intramural division とは、所内研究部門と訳すこともあり²⁾、自ら基礎研究、疫学研究、臨床試験を行うところである。基本的には、NCI/NIH 外の研究所や病院と変わらず、また、わが国の国立がんセンターと同じように、基本的には個々の研究者の興味と関心をもとに研究を行う。研究費は、NCI/NIH 外の研究者と異なり、grant などの研究費に応募するのではなく NCI から直接研究費を交付され、業績の評価も内容的には所外の研究者とほぼ同じように行われるようである。NIH 内には Clinical center と呼ばれる臨床試験のみを行う病院があり、NCI の intramural division が主に第 I 相臨床試験を行っており、同じ建物で基礎研究者が実験を行っている。日本から多くの研究者が留学しているのは、この intramural division である。歴史的にはもともと、他の機関と同様、NCI も自ら研究を行う施設であったが、研究費の配分機能を担うようになって後述の extramural division や office of the director が設立されたようである。筆者の観察の範囲内であるが Intramural division の研究テーマが「国立機関」らしい国全体を見据えたものであるということもなさそうである。NCI が自分たちで持っている、

普通の研究施設・研究病院であると考えればよいであろう。

4. 最後に

NCIのがん治療薬開発は、アカデミック発の創薬という観点から見ると、個々の研究者が自分たちですべてを担わなければならない日本とは大きな差がある。予算の額や国の事情も異なるため、同じことをする必要もないし、できないが、新規候補薬剤から臨床開発まで組織的にサポートを行うという考え方は参考になる。わが国のがん治療開発システムは米国に20～30年遅れているといわれる。米国で多施設共同臨床試験グループ設置のプログラムができたのが1958年であるが、わが国においても近年急速に多施設共同臨床試験グループがいくつか整備され、医師主導治験が実施できるようになってきた。しかし、臨床試験の実施を通してがん標準治療を世に広め、予防研究を実施するための組織であるCCOPに対応する仕組みはまだない（CCOPが米国にできたのは1986年）。また、NCIにおいて、治療開発の管理・推進を行うCTEPが設立されたのは1981年である（personal communication）。NCIは治療開発のみならず、診断法や予防薬開発にもDTPやCTEPと類似の組織・システムを持ち、総合戦略として国家的にがん研究を推進している。わが国にはこのように系統的かつ恒常的にがん撲滅の研究を統括する部署は国立がんセンターにも厚生労働省にも存在しなかった。しかし、2006年のがん対策基本法が成立し、それに応じた組織が厚生労働省（がん

対策推進室）と国立がんセンター（がん対策情報センター）に設置された。国立がんセンターがん対策情報センターは、わが国ではじめてできたがん研究分野のextramural divisionである。

米国と同じシステムを作ることはできなくとも、日本全国のがん研究者が協力して、わが国にあった形のがん予防・治療開発の戦略をとっていく必要がある。組織的な治療薬開発プロセスとCCOPに対応するネットワークを作る具体的なアクションを早急に取り始めないとアカデミック発の創薬、標準治療に対するエビデンスの発信という観点からの医療への貢献は、これまで以上に諸外国に遅れて行くことになるであろう。これには行政と協力した国立がんセンターがん対策情報センターの強力なリーダーシップと、日本全国のがん研究者が信頼することのできる透明かつ効率的なシステム作りが不可欠と考える。

文 献

- 1) 黒川幸典, 福田治彦: 米国 NCI における TR 支援体制. *Drug Delivery System* 22-1, 54-59 (2007).
- 2) Gallin JI 編, 井村裕夫監修: NIH 臨床研究の基本と実際, 丸善, 東京, 2004.
- 3) GRANTS PROCESS. Everything You Wanted to Know About the NCI. But Were afraid to Ask. National Cancer Institute. National Institutes of Health, U.S. Department of Health and Human Services. Revised August 2005.

Round Table Meeting

バイオマーカー研究の現状と展開

司会

近畿大学医学部ゲノム生物学教室教授

西尾 和人

Kazuto Nishio

国立がんセンター研究所がん情報・統計部室長

山本精一郎

Seichiro Yamamoto

国立がんセンター中央病院消化器内科医長

山田 康秀

Yasuhide Yamada

ABSTRACT

分子標的治療薬の臨床試験が上皮成長因子受容体(epidermal growth factor receptor: EGFR)-チロシンキナーゼ阻害剤(tyrosine kinase inhibitor: TKI)を中心に、国内外で積極的に行われている。2004年にEGFR遺伝子変異が発見され、EGFR遺伝子変異がEGFR-TKIであるゲフィチニブやerlotinibの奏効率と関連性の強いことが示された。それ以来、EGFR遺伝子変異をはじめとする、治療薬への感受性および疾患の予後を予測するバイオマーカー研究への関心が高まっている。

そこで、本座談会では「バイオマーカー研究の現状と展開」をテーマに、近畿大学医学部ゲノム生物学教室教授の西尾和人先生の司会のもと、国立がんセンター研究所がん情報・統計部室長の山本精一郎先生と国立がんセンター中央

病院消化器内科医長の山田康秀先生に、それぞれのご専門の立場から、分子標的治療薬におけるバイオマーカー研究の現状と今後の方向性についてお話をうかがった。

近年、遺伝子変異などのバイオマーカーを活用して患者に関する情報を取り入れた臨床試験では、有効な結果を得られやすいことが示されている。臨床試験を実施する際にはバイオマーカーを用いて患者を選択することが重要であり、かつペバシズマブなどの血管新生阻害剤の使用において生命に関わる有害事象を回避するためにも、バイオマーカーの活用が望まれる。今後、効果予測因子の解明をさらに進めるとともに、予後因子と副作用マーカーとの関連性など、バイオマーカー研究の発展が期待される。

予後因子・予測因子

西尾 本日は、現在の分子標的治療薬のバイオマーカー研究の現状と展開をテーマに、生物統計学に詳しい山本先生と、消化器がんのみならず分子標的治療薬の第Ⅰ相試験の経験をおもちで、基礎にも精通しておられる山田先生にお集まりいただきました。臨床の現場あるいは統計学的視点からみた臨床試験デザインについての、レビューおよび議論がいただけたと思います。また、2007年の米国癌治療学会議(ASCO)における分子標的治療薬に関する動向を交えながら、バイオマーカーあるいはバイオマーカーを取り入れた臨床試験について、事例を提示しながら、その問題点や方向性についてご議論いただければと思います。

さて、昨今、バイオマーカーを取り入れた臨床試験のデザインについて、prognostic か predictive かということが議論されるようになりましたが具体的な事例でご説明していただけますか。

山本 このところ、たとえばゲフィチニブを用いた単群での試験で、上皮成長因子受容体(epidermal growth factor receptor; EGFR)変異陽性の群で予後がよく、陰性の群で予後が悪いという議論が学会などでしばしば行われています。しかし、それはEGFR変異の有無が予後のよい群と悪い群とを分けるバイオマーカーであっただけで、薬剤の有効性については分かりません。そのため、概念を整理して、予後因子と予測因子をより正確に分けることが必要になっています。

予後因子とは、performance status(PS)のように、何も治療を受けなかった場合や標準治療を行った場合でも因子の水準によって予後が変わるものをいいます。一方、予測因子は、因子の水準によって治療効果が異なるもののことです。個別化治療開発とは、その治療が有効な集団を探す、という意味で予測因子を探す試みであるといえます。

予測因子を探すための統計的なデザインとして提案されているものは、大きく3つに分けることができます。まず、マーカー(+)の患者だけを集めて行う「marker(+)design」という、いわゆる enrichment

と呼ばれるものがあります。マーカー(+)の患者のみで新治療あるいは標準治療を行うのですが、これでは治療が有効なことはわかっても、無効な場合についての詳細がわからないという欠点があります。

また、標準治療の群とバイオマーカーに従って治療を変える群とをランダム化して比べるという「marker strategy」と呼ばれるデザインもあります。このデザインの場合は、マーカー(+)の患者がどの程度いるかによって全体としての治療効果が変わってしまいます。つまり、マーカー(+)の患者に対して非常に有効であったとしても、マーカー(+)の患者数が少なければ結果に差が出ないということになってしまいます。

それに対して、はじめにバイオマーカーを測定してからランダム化し、治療の効果をマーカー(+)の患者とマーカー陰性の患者で比較する方法があります。予測因子を探すには、この「all comers design」が最適であると考えられています。

バイオマーカーと臨床試験デザイン

山本 最近、Journal of the National Cancer Instituteに掲載されたプロテオミクスの例があります(Taguchi F, et al: J Natl Cancer Inst 99(11): 838-846, 2007)。これは、matrix-assisted laser desorption ionization (MALDI)を用いて予測因子・予後因子について検討したものです。

西尾 プロテオミクスのアプローチで、EGFR-チロシンキナーゼ阻害剤(tyrosine kinase inhibitor; TKI)の効果を予測できるバイオマーカーを同定するという研究ですね。

山本 そうです。この田口先生たちの論文は、主題はプロテオミクスによって治療効果の予測が可能かということですが、そこで使われた統計的な方法が今後のアイデアにつながるものでした。彼らのアプローチは、まずEGFR-TKIによる治療集団をトレーニングセットとして、プロテオミクスで予後が予測できるようなバイオマーカーを同定しました。次に、別のEGFR-TKIによる治療集団に対してそのバイオマ-

西尾 和人 氏

Kazuto Nishio

近畿大学医学部ゲノム生物学教室教授

PROFILE

1986年和歌山県立医大卒，同大学付属病院で内科研修を経て，国立がんセンター研究所薬効試験部で抗がん剤耐性，シグナル伝達，分子標的治療の研究を始める。中央病院計画治療病棟支援施設でのTR研究を経て，2006年より近畿大学医学部ゲノム生物学講座教授。WJOG TR委員長，東京医大，北里大学客員教授など。



カーを用いたところ，そこでも同様に予後予測ができ，一連の validation 作業によってそのバイオマーカーが予後因子だということが確認されました。

西尾 検証の際には，Eastern Cooperative Oncology Group (ECOG) やイタリアのデータも用いられていますね。

山本 はい。いくつものデータを使って validation が行われています。その後，EGFR-TKI を投与されていない集団に対してこのバイオマーカーを用いたところ，予後予測ができませんでした。そのため，このバイオマーカーは予後因子ではなく，予測因子であるという論理の展開をしています。

この論文は，治療予測因子を探す方法として2つの研究を組み合わせるといったアイデアを与えてくれたと思います。たとえば，ある治療を行った集団をよく調べたところ，そのなかで予後の違う集団が規定するマーカーをみつけることができたといいます。これで終わるのではなく，次にそのマーカーで規定される予後の良かった集団を enrich してその治療法のランダム化比較をすれば予後因子だったのかがわかる，といった方法が考えられます。ただし，どこかの段階でランダム化比較は必要だと思いますし，マーカーで規定されるある集団に治療を行うかどうかはそのマーカーが治療の予測因子か予後因子かではなく，その集団に対して治療効果があるかどうかによって決めなければなりません。

分子標的治療薬とバイオマーカー

1. EGFR-TKI

西尾 では次に，分子標的治療薬とバイオマーカーの今後の展開についてお話ししたいと思います。たとえば，保険適用になった遺伝子検査もあり，今後の動向は変わってくるだろうと思いますがいかがでしょうか。

山田 EGFR 関連では，EGFR-TKI である erlotinib やゲフィチニブの大腸がん，あるいは胃がんに対する第Ⅱ相試験がいくつか行われたのですが，全く効果がなく，奏効率は0～1%でした。大腸がんに関してはEGFR 変異自体がほとんどないということがわかっています。

一方，抗EGFR 抗体である cetuximab は，大腸がんに対して単剤で約10%，塩酸イリノテカンとの併用で塩酸イリノテカン不応例に対し20%ほどの奏効率が得られることがわかっています。そのため腫瘍縮小効果の得られやすい患者群の特定が行われているわけなのですが，その際に免疫染色法が役に立ちません。大腸がんの約80%の症例は免疫染色法で陽性ですが，陽性例でも陰性例でも抗EGFR 抗体の奏効率は10%程度であるという報告もあります。また，免疫染色の強度にも関係なく，全体の何割が染まっているか(陽性細胞率)ということにも全く関係ないということもわ

かっています。

少数例の検討ですが、FISH法でEGFR遺伝子の増幅がみられる症例に関しては、やや奏効率が高い傾向にあるのではないかとされています。EGFRの下流の話ですが、K-ras変異に関する興味深い2つの論文と2007年のASCOでの口演があります。K-ras変異を有する症例ではcetuximabの奏効率が、その3つの報告ともすべて0%で、K-ras変異は効果予測のきっかけになるかもしれないと考えられているのです。予後自体も、cetuximabで治療した場合に、K-ras変異を有する症例よりも変異をもたない症例のほうが良好であったことが併せて報告されました。

このほか、EGFR関係ではpanitumumabとnimotuzumabなどもあります。最もバイオマーカーの研究が進んでいるのはcetuximabで、その他の抗体に関してはそれほど情報は得られていないというのが現状だと思います。

西尾 Cetuximabについてはわれわれも他の日本のグループも、抗体依存性細胞傷害反応(antibody-dependent cell-mediated cytotoxicity; ADCC)の関与など、抗体のバイオマーカーという点で研究を進めています。

山田 おっしゃるとおり、ADCCの影響は非常に大きいのではないかと考えられています。しかしながら、cetuximabが免疫グロブリンG(immunoglobulin G; IgG)-1抗体で、panitumumabはIgG-2抗体ですが、単剤での奏効率がいずれも10%程度なのです。ADCCがほとんど関与しないといわれているIgG-2抗体のpanitumumabでも、単剤での奏効率がcetuximabと変わらないというところが気になりました。

リツキシマブでは、リンパ球のFc受容体の遺伝子変型により、リツキシマブの効果が変化するという報告があります。Cetuximabでも同様のことがあるか否か確認する必要があると思います。

西尾 もし日本人で白人と比べ奏効率が違っていると、人種差もまた調べていかなければならないかもしれませんね。ADCCは免疫担当細胞で起こります。トラスツマブでもそう思われていたのですが、直接的に抗体ががん細胞に結合して増殖を抑制する作用と、その免疫の細胞を傷害することで抗腫瘍効果を示すと

いう作用があります。

山本 それもマーカーでわかるのですか。

西尾 現在、それを測定しています。Fc受容体の一塩基多型(single nucleotide polymorphisms; SNPs)の解析も行っています。また、抗体糖鎖のフコースの状態によってADCCが少し変わるので、ヒトの体内でフコースの除去・付加を行う酵素の活性を測定しています。実際に、トラスツマブのフコースの量の半定量化を国立がんセンターのチームと共同で行っています。

山本 それはマーカーとセットで調べることになりませんか。

西尾 同じタイプの抗体を使ったとしても同じように効かないかもしれないのでそうなります。

山田 コストが高い薬剤では特に効果を予測するためのバイオマーカーが重要です。Cetuximabは塩酸イリノテカンとの併用でも奏効率は20%です。多くの患者には副作用が出るだけで効果がありません。そのため、治療にかかるコストの問題もありますが、われわれ科学者としては、治療が有効な症例をしばらくこんでいくことが非常に重要であり、それがテーラーメイド医療につながると思います。

2. 血管新生阻害剤—ベバシズマブ—

西尾 コストといえばベバシズマブもそうですね。ベバシズマブのバイオマーカーは必要だと思いますか。

山田 現在のところ効果を予測するバイオマーカーはありません。有害事象の面からみても、バイオマーカーは必要だと思います。ベバシズマブでは、生命に関わる有害事象が報告されています。頻度は1.5~2%と低率ですが、心筋梗塞、肺塞栓、脳梗塞のような動脈系の塞栓症を起こします。消化管穿孔の頻度も1.5%で、これも緊急手術をして救命できるとしても、生命に関わる場合があるかもしれません。ですから、ベバシズマブが有効な症例の特定と、有効であった症例がいつ無効となるのかを知ること、すなわち無駄な投与期間を減らすことは、生命に関わるような有害事象の頻度を最小限にするためにはきわめて重要なことだと思います。まだこのようなバイオマーカーは特定され

山本精一郎 氏

Seiichiro Yamamoto

国立がんセンター研究所がん情報・統計部室長

PROFILE

生物統計家としての専門を活かし、1996年からJCOGデータセンターの統計チーフとして、がん臨床試験のデザインや解析に従事。臨床試験方法論の研究にも取り組み、JCOG臨床試験の質的向上に貢献。また、厚生労働省多目的コホート研究の事務局として、全国14万人規模の前向き観察研究データの収集、解析を行っている。生物統計家として他分野の研究者と多くの共同研究を行い、自身の方法論的研究ならびに他分野研究者との共同研究によってがんの予防、診断、治療の各分野において独自の貢献をしてきた。



ていないため、臨床研究を継続的にやっていくべきであると思います。

西尾 副作用のマーカーですね。効果と経済的な観点からすれば、不必要な患者にはあまり投与しないことも必要ですし、投与を止める時期も大切ですね。

山田 ベバシズマブなどの血管新生阻害剤による高血圧や蛋白尿などの有害事象と、その効果や予後との関係は今のところ全くいわれていません。血管新生のバイオマーカーの探索に関してまさに日本の研究者が行うべき余地は、まだまだ残されています。

西尾 ベバシズマブについても、バイオマーカーのアプローチは、ASCOをみてもほとんどありません。わが国で何かできればと切に思います。

ASCOでのバイオマーカーに関するトピックス

1. AZD2281

西尾 ポリ(ADP-リボース)ポリメラーゼ(poly(ADP-ribose) polymerase; PARP)阻害剤であるAZD2281は、第I相試験で対象をBRCA1/2変異陽性の卵巣がんにとった結果、有効性もみられています。

山田 第I相試験とは思えないほど腫瘍縮小効果が高かったですね。

西尾 第I相試験で有望な結果を示すことができたのは、特定の患者集団に絞ったからです。はじめ

に小さな特定の集団で試験をするというアプローチは、非常に理想的です。

山田 特定の対象から進め、有効であることがわかれば、少しずつ対象を広げていくということですね。

西尾 そうなのです。BRCA1/2変異陽性ということは、むしろ乳がんのほうが主な対象となりそうです。第I相試験でこういうデータを見ると、はじめに特定の対象に絞って試験を行い、徐々に適応を拡大するという、従来とは逆の戦略が非常に必要だと思います。

2. AZD2171

西尾 もう1つASCOで注目したのは、脳腫瘍に対するAZD2171です。脳腫瘍ですぐに効いているのです。脳浮腫のところ非常にphenotypeとしてみやすく、血管を正常化することがproof of concept(POC)として証明されているのです。既存の腫瘍血管の正常化というものは、血管新生阻害剤が有効であるかどうか、そういう作用があるのかはあまりはっきりしていなかったのですが、それが証明されたというのは面白かったですね。

AZD2171による血管の正常化が臨床で確認されたことにより、こうしたアプローチは今後増えていくでしょう。たとえば、Nature Medicineに掲載されていたベバシズマブに関する論文では、間質圧の上昇を実際に測っており、非常に興味深いと思います。

山田 直腸がんが対象で、内視鏡を使った研究です

山田 康秀 氏

Yasuhide Yamada

国立がんセンター中央病院消化器内科医長

PROFILE

1989年3月弘前大学医学部卒業。1996年4月癌研究会附属病院化学療法科、1998年4月より国立がんセンター中央病院消化器内科へ勤務し、現在に至る。2006年には Memorial Sloan-Kettering Cancer Center へ留学。所属学会は日本内科学会、日本癌学会、日本癌治療学会、日本臨床腫瘍学会、日本消化器病学会、がん分子標的治療研究会、ASCO、ESMO など。



ね。確かに興味深いですね。われわれもぜひ内視鏡やMRIなどの画像も含めた形でさまざまな方向からの研究を試みたいと思っています。

西尾 EGFRなどの発現解析や蛋白レベルの解析、ゲノムレベルでの gene application などにおいて変異が注目されており、それらが予測因子である可能性があります。ですから、量的な変化ではなく質的な変化をマーカーとして使っていこうという方向性があると思うのです。

山本 薬剤の開発をするときには、マーカーの開発も一緒にするんですよね？

西尾 変異や遺伝子の変化があるような細胞株をライブラリーとして集めていく、いわゆる cell based screening というのですが、そういうアプローチをわれわれも始めています。それによって、特定の患者サブpopulationに該当する細胞株を用いた効果予測・創薬ができるのではないかと考えています。診断薬と治療薬を同時に開発しようというコンセプトがありますが、それには臨床あるいはヒトのサンプルが必要になります。そのためには、組織バンクなどの整備が日本でももっと行われなければ、現状ではなかなか難しいのです。

山本 ゲフィチニブのとき、結局第Ⅲ相試験を行ってからマーカーを探した形になっていましたね。

西尾 その反省があるからこそ、たとえば AZD2281 の第Ⅰ相試験で良好な結果が出たと考えています。

山田 優れたマーカーを事前にどれだけみつけさせるかというチャレンジですね。

西尾 そうです。患者をバイオマーカーで選んでいくというアプローチのほうがいいという流れになってきていると思うのです。そしてメシル酸イマチニブのように、徐々に適応拡大を図っていくのが望ましいですね。

3. メシル酸イマチニブと sunitinib malate

山田 メシル酸イマチニブでは、消化管間質腫瘍 (gastrointestinal stromal tumor; GIST) のKIT変異症例で有効であるということがわかった時点で、さまざまながん種の cell line のKIT変異が調べられましたが、結果的にはみつかりませんでした。効果予測のバイオマーカーを探していく段階で、KITの場合は初期にエクソン11が関与していることがわかりました。そして、メシル酸イマチニブ耐性となったら sunitinib malate となります。

当初、血液を用いて可溶性KITなども調べていたのですが、効果予測因子とはなりません。そのため、メシル酸イマチニブではこのような研究はほとんど行われていません。

西尾 Sunitinib malate の状況はどうか。

山田 Sunitinib malate に関しては、KIT変異の解析が行われ、エクソン11変異の場合は有効であるというデータがあるのですが、その真偽はわかっていま

せん。

エクソン11の二次耐性の機序は、KITのチロシンキナーゼ領域に二次変異が起きたためメシル酸イマチニブが結合できなくなるというものが最も多いようです。エクソン11の変異をもつ症例ではそのような二次性の変異がメシル酸イマチニブを長く投与している間にできてしまうため、sunitinib malateも結合できなくなってしまいます。ただし、sunitinib malateはエクソン11二次性変異症例には効かないかもしれませんが、メシル酸イマチニブ投与前にファーストラインでsunitinib malateを投与すれば違った結果になるかもしれないと考えている研究者もいます。

一方、エクソン9は比較的メシル酸イマチニブが暴露している時間が短いということもあり、二次変異が起きにくくなっています。もともとの自然耐性の影響が強いため、エクソン9にはまだsunitinib malateが結合しやすいと考えられていますが、はっきりとした事実はわかっていません。しかし、メシル酸イマチニブはファーストラインでの奏効率が非常に高く、マルチターゲットであるためさまざまな副作用が出てくるsunitinib malateに比べると、忍容性が高く、毒性をマネージメントしやすい薬です。

4. その他のトピックス

西尾 二次変異については、たとえばEGFR以外にもc-METに関する報告が今回のASCOでありました。MET阻害剤ということになれば胃がんに対する期待は大きく、EGFR-TKIとの併用を考えているグループもありますが、私は単剤でも効くと思います。

山田 c-METは以前から免疫染色法によって胃がんの悪性度との関係が日本ではいわれており、しかも今のところc-Metを標的とした治療はほとんどありません。未分化型腺がんにもc-Metが発現しているという報告もあり、非常に魅力的で、ASCOでは発表が2題ありました。

西尾 ありましたね。

山田 いずれもがん種は別として治療に対する反応がみられ、胃がんに対する開発も国内外で始まっているところもあると思いますが、早期に結果が知りたいですね。

西尾 今回のASCOで非常に注目されていましたね。

山田 そうですね。インパクトがありました。

西尾 今回のASCOではほかに何か気づいたものはありますか。

山田 マルチターゲットのTKIで、circulating endothelial cell(CEC)やcirculating endothelial progenitor cell(CEP)、あるいは可溶性トランスフォーミング増殖因子(transforming growth factor; TGF)受容体2などの報告があるのですが、一定の結果が出ているようには思えませんでした。採取するタイミングや薬剤の影響なのかもしれませんが、現在のところどのように理解すればいいのでしょうか。

西尾 Circulatingの話は、2005年、2006年は非常に盛上がったのですが、今回はあまり印象に残りませんでした。POCの証明として早期の試験で用いるには有効かもしれないのですが、予測因子として用いるにはまだ非常に感度が低いという印象があります。可溶性TGFについては、マルチで解析すると少量の血清で測定できるものがあります。たとえば今回のASCOでのAZD2171の発表で、CCのインスリン様成長因子(insulin-like growth factor; IGF)などを一緒に測定しており、単独ではなくいくつかの因子と組み合わせで発表されていました。そういったアプローチは有望だと思います。

血管新生は病態がよくわからないので、マーカーは難しいかもしれませんが、たとえばベバシズマブに耐性となったときにVEGFの代わりに他のリガンドが増えてくるというシフトの話があり、薬剤の効き方などに関わってくるかもしれません。したがって、これらを探していくアプローチは必要だと思います。

バイオマーカーを利用した臨床試験の方向性と問題点

西尾 では最後に、バイオマーカーを利用した臨床試験の今後の方向性と問題点についてお話をお願いします。

山田 もう一度、細胞傷害性抗がん剤の効果の予測を検証しておきたいと考えています。大規模な試験の

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、そういった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 開発とは異なる戦略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探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¹⁾²⁾。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基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。

Schaid⁶⁾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、Storer⁷⁾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。

Ellenberg⁸⁾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⁾, Scher¹⁰⁾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ただけ)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はないよう、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小さいといえる。

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い、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と誤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。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、より精確な推定値、すなわちより精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戦略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戦略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、今となってはももとの「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.

** Seichiro 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、ストップングルール、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ふ。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前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

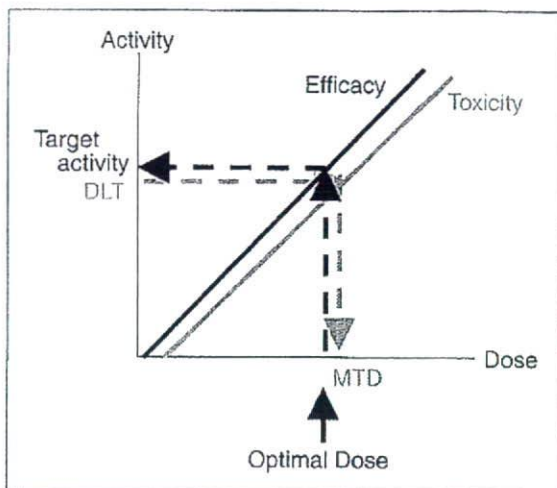


図1 用量-毒性-効果の関係(毒性と効果が比例)

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⁾。

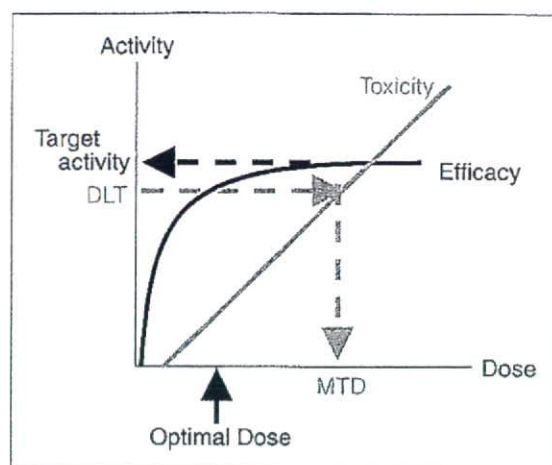


図2 用量-毒性-効果の関係 (MTDより前に効果がプラトー)

- 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