

もし感受性仮説が証明されたら、新しい標準治療の決定につながるのか？
一様な効果(交互作用なし) ・新治療がすべての患者さんに対し標準治療となる
閾値モデル or 直線モデル ・それぞれのマーカーサブグループに新治療を導入するかはそれぞれのサブグループでの治療効果の大きささしだい
少なくとも「治療効果はマーカー値によって異なる」という仮説は証明された ・標準治療とするためには、次のステップとして新たなRCTが必要

表2 ● 感受性仮説に基づいて標準治療を決定

uniform仮説でなく、治療と感受性の交互作用を示す(上記のような)感受性仮説を対立仮説とすると、感受性有り無しについての検証を行うことができる。帰無仮説は棄却したい仮説により、④とすることも⑤とすることもできる。

感受性仮説と臨床的意義について考えてみよう。例えば⑤を帰無仮説として、上に挙げたそれぞれの仮説が証明された場合、次の標準治療はどのようなのだろうか(表2)。uniform仮説が証明された場合は、もちろん新治療がすべての患者さんに対して標準治療になる。次にlinear仮説が証明された(マーカーが高いサブグループほど治療効果が大きい)場合や、threshold仮説が証明された(マーカーが高いサブグループにだけ治療効果がある)場合、どのサブグループに新治療を導入するかは、そのサブグループに対する治療効果がどの程度あるかということに依存する。つまり、感受性仮説が証明されただけでは、あるサブグループに対して十分に臨床的に意義のある治療効果があることを証明したことにはならない。標準治療となるためには、感受性仮説のみならず、それぞれのサブグループに対して臨床意義clinical significanceがあるかどうかを検証することが必要となる。

今回のJCOG試験では、感受性仮説の検証を副次的な目的とした。副次的な目的であっても治療効果が腫瘍に対する薬剤の反応によって異なる

というproof of conceptができれば、次のステップとしてそれを取り込んだ治療開発を行うことができる。仮説を立てるために利用するデータとそれを検証するためのデータは独立であるべきという科学の一般論からも、感受性のあるサブグループを探索するステップと臨床的意義があるかどうかを検証するステップは分けて考えるべきである。具体的には、以下のようなデザインが考えられる。

- ①：phase IIIで副次的にproof of conceptを行い、次のphase IIIで検証する。
- ②：前述のSimonたちのAdaptive signature designのように、1つの試験を感受性のあるサブグループを探索する部分と、それを検証する部分に分ける。
- ③：標準治療をアームに含んだrandomized phase II design (例えばランダム化スクリーニングデザイン⁴⁾)で感受性の探索やproof of conceptを行い、次のphase IIIで検証する。
- ④：randomized phase II / phase IIIを用いて、③のrandomized phase II と phase IIIをシームレスに行う。

本稿では、SimonやSimon & Freidlinの提案や、われわれのアイデアを紹介した。さらに進んだ方法として、linear、threshold、uniformといった複数の対立仮説から最もデータにフィットした対立仮説を選択するという統計的方法も考え

られる。あるいは、uniformとlinearを組み合わせることによって、より広い対立仮説に対して高い検出力を保つ方法も考えられる。統計的にはいろいろな工夫が可能であるが、臨床的な意義や行いたいproof of conceptを十分に反映させるようなものにならなければいけない。意義のある仮説とそれを効率よく検証できるよう試験をデザインするためには、今までも増して臨床家と統計家のコラボが必要と言えるだろう。

謝辞

本稿の内容は筆者が所属するJCOGデータセンターのメンバーやJCOGの研究者との議論のなかで整理され、考え出されたものであり、その方々との共作といえる。本研究はがん研究助成金指定研究(17指-5)の補助を受けた。

文献

- 1) Sargent DJ, et al. Clinical Trial Designs for Predictive Marker Validation in Cancer Treatment Trials. *J Clin Oncol* 2005; 23: 2020-7.
- 2) Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005; 23: 7332-41.
- 3) Freidlin B, Simon R. Adaptive Signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res* 2005; 11: 7872-8.
- 4) Rubinstein LV, Korn EL, et al. Design Issues of Randomized Phase II Trials and a Proposal for Phase II Screening Trials. *J Clin Oncol* 2005; 23: 7199-206.

研究者主導臨床試験における データマネジメントのアウトソーシング

¹国立がんセンターがん予防・検診研究センター、²国立がんセンター中央病院、

³京都府立医科大学分子標的癌予防医学、⁴東京大学医学系研究科・NPO日本臨床研究支援ユニット

山本精一郎¹、福田 治彦¹、濱口 哲弥²、奥坂 拓志²、
牧本 敦³、石川 秀樹³、大橋 靖雄⁴

1. 研究の背景と目的

がんの治療開発は、主に製薬企業による治験によって行われるが、化学療法以外の治療法（手術や放射線など）やこれらの組み合わせ治療の開発は製薬会社に頼るだけでは達成されず、研究者主導の臨床試験を実施することが必要である。臨床試験は実施施設とインフラの点から、安全性の綿密なモニタリングが必要なために単（少数）施設で行う第Ⅰ相～第Ⅱ相試験などの早期開発、より実臨床に近い状況で行うことが必要のために多施設共同研究グループ（Cooperative Group）などで実施する第Ⅱ相～第Ⅲ相試験などの後期開発に分けられる。また、法規制という面からは、治験とそうでないものに分けられる。しかし、それが人間を対象とした実験である以上、どのような臨床試験であっても、科学性と倫理性を担保するためには、研究者グループ、データセンター、独立監視機構といった機能が必須となる。研究者主導の臨床試験では、日本臨床腫瘍研究グループ（JCOG）や西日本胸部腫瘍臨床研究機構（WJTOG）など独自に恒常的なデータセンター機能や独立監視機構を有しているところもあるが、試験単位、あるいは施設単位でそのような機能を実現することは難しい。その場合、一つの可能性として、商業的あるいは非営利のCRO（Contract Research Organization, 開発業務受託機関）にデータセンター機能や独立監視機構を委託することが考えられる。本稿では、

自ら研究者主導臨床試験を実施する場合に、CROにそのような委託をした経験のある研究者、あるいは委託先であるCROの方、あるいは自分たちで苦勞しながら臨床試験をやっている研究者と研究者主導臨床試験データマネジメントのアウトソーシングの経験と今後の可能性について議論したい。

以下の議論では、主に①行っている（委受託している）研究の種類と内容、②スタッフおよび役割分担（データマネジメントやデザイン、解析を誰がやっているか、委託しているか）、③抱えている問題、などに焦点を当てる。

2. 研究者グループのみで臨床試験を行う際の問題点

① 行っている研究の種類と内容

本稿著者の一人である濱口が属する国立がんセンター中央病院消化器内科では、製薬会社の治験やJCOG臨床試験のほかに、自分たちで自主臨床試験を行っている。研究の種類としては、各癌種（食道癌、胃癌、大腸癌）のPhase I/II、Phase IIと、研究所との共同研究として行っている抗癌剤感受性におけるバイオマーカーに関する探索的研究（マイクロアレイ）3種（食道癌・胃癌・大腸癌）などこれまで10以上の自主研究を行ってきた。

② 内部スタッフと委託、臨床試験における役割分担

スタッフとして医長2名と医師3名がおり、1日25-50人の外来患者を担当している。チーフレジデント・レジデントは2-5名おり、スタッフとともにチームで入院患者10-15名を担当している。また、研究補助員が5名おり、3名が自主研究とJCOG研究のデータマネジメント・登録およびグループ秘書兼務、残り2名がJCOG研究のCRF (Case Report Form) 記入を行っている。その他には、実験補助員4名でトランスレーショナル・リサーチの検体処理などを担当している。治験に関しては病院の治験管理室のCRCが治験ごとに2-3名サポートしてくれている。外部への委託は行っていない。

研究計画・実施の流れとしては、スタッフやチーフレジデントがコンセプトを出し、グループカンファレンスで討議した後に、主にチーフレジデントが中心になってプロトコルを作成してスタッフが指導するという体制である。病棟業務がスタッフとチーフまたはレジデントとの2人1組のチーム制であるため、そのペアでプロトコル作成からデータのまとめまでを一緒にやることが多い。プロトコル作成はPhase IIに関しては自分たちが中心となって行ったJCOG研究など過去のプロトコルを元に作っている。研究事務局は、チーフもしくはレジデントと研究補助員がやっており、登録とCRF管理(データベースへの入力まで)は研究補助員、CRF記入およびデータ解析は担当のチーフもしくはレジデントがやっている。学会発表にあわせてデータをまとめることが多いといえる。

③ 抱えている問題点など

多忙な臨床をしながらの研究であり、治験やJCOG試験のようにCRCやデータマネージャーがいるわけではないので、さまざまな点でチェック機構が働いていない。

他の治験やJCOG試験などが複数同時進行

して混乱してしまうこと、とくに外来治療がメインとなった最近ではプロトコルを確認しながらオーダーするだけの余裕がないことから、採血項目が抜けていたりCTが規定通りにとられていない(C T予約を取りにくいことも災いしている)、減量規定なども守られていないこともある。また、データ入力したものの二重チェックやロジカルチェック、PR例のレビューも十分できていないのが現状である。また、固有の問題として、担当チーフレジデントが修了して他院で勤務することになった場合に論文化されずそのままになってしまう、という問題もある。

3. 研究者主導臨床試験におけるCRO委託の経験

① 行っている研究の種類と内容

本稿著者の一人である奥坂が属する国立がんセンター中央病院肝胆脾内科では、これまで多施設で行ってきた臨床試験のいくつかにおいてその業務の一部分をCROに委託した。委託した研究は「ラクトフェリン含有食品のC型慢性肝炎に対する臨床試験」、「小肝細胞がんに対する肝切除術と経皮的局所壊死療法」などがある。ここではその一部を紹介する。

② 内部スタッフと委託、臨床試験における役割分担

日本CRO協会パンフレットによると、CROの業務は様々あるものの、表1に挙げたものが主なものとなっている。このうち、「患者調査票等書式作成(デザイン)及び施設への送付」、「試験錠菓の割付及び施設への送付」、「登録事務局(登録受付、終了時受付)」、「データマネジメントに関わる業務(入力システムの作成・入力、データセット作成)」、「中間解析に係る業務(データセット作成)」などを試験ごとにそれぞれ別のCROに委託した。

研究班の中に統計家がいたため、研究計画書作成や統計解析、中間解析審査などはCROに委託せず研究班主体で行ったが、データマネジ

表 1. CRO (開発業務受託機関) の主な業務

- 1) 臨床試験・市販後調査などのプロトコール・治験(調査)薬等概要書・症例報告書(調査票)書式の作成や、作成に際しての専門的アドバイス業務
- 2) 臨床試験・市販後調査などの実施医療機関の選定や、医療機関に対する治験(調査)依頼・治験(調査)契約締結に関する業務
- 3) 実施医療機関における、臨床試験・市販後調査などのモニタリング(調査)業務、治験(調査)薬の交付・管理・回収、症例報告書(調査票)の回収とチェック業務
- 4) 実施医療機関に対する治験(調査)終了・中止に関する業務
- 5) 症例検討などのデータマネジメント業務や、症例の統計解析業務
- 6) 治験総括報告書・投稿論文・安全性定期報告(案)などの作成や、作成に際しての専門的アドバイス業務
- 7) 治験(調査)監査業務

日本CRO協会パンフレット(<http://www.jcroa.gr.jp/>)より

メント業務の主な部分は委託した。各施設とCROとの関係などの臨床試験オペレーション業務(クリニカルトライアルマネジメント)は研究代表者とその秘書による研究事務局が行った。

④ 抱えている問題点など

委託契約費に関しては、公的研究費の総額や委託費上限(研究費の50%以下)という制約の中では、CROにこれ以上の業務を委託することは不可能であった。錠薬の割付や患者さんの要求に合わせた登録業務など、研究者では対応できないようなことをしてもらえた反面、結果的にはCRFデザインなど研究者側でも同じレベルのものができた、こちらが期待したとおりのデータマネジメントが行われていなかった、各施設と研究事務局、CROとの間の関係などの臨床試験オペレーション業務などについて、予想していたよりも研究事務局の負担がずっと多かった、同じ業務を頼んでも業者によって価格設定がかなり違う、などの感想を持った。

これらの予想と現実との乖離は、委託費に制約があったこととも当然関係するが、臨床試験実施に対する研究事務局側の知識が不十分であったことにも由来する。CRO側が研究者主導臨床試験に不慣れであったこともあるかもしれな

い。いずれにしろ、CROに業務を委託する前に必要な臨床試験業務を洗い出し、CROと研究事務局で業務をどう分担するか、委託する部分については内容を詳細に詰め、自ら担当する部分についてはその対応をあらかじめ周到に準備する必要がある。また、委託業務の拡大が可能となるよう委託費上限の見直しや、公的研究に適した新たな契約形態の開拓、研究者主導臨床試験に適したCRO業務の内容を明らかにしていくことも重要な課題である。

4. 小児がん多施設共同臨床試験を支援するデータセンターの活動

① 行っている研究の種類と内容

小児がん領域の臨床研究は、長く臨床実践データの調査研究レベルを超えず、仮説検証に基づく治療開発を行うための十分な方法論と支援を行う機構が存在しなかった。平成14年度厚生労働科学研究費補助金(効果的医療技術の確立推進臨床研究事業)において、小児分野に初めて大型補助金が交付された事をきっかけに、本稿著者の一人である牧本を中心に主に小児固形腫瘍に対する臨床試験の支援を行うための「小児がんデータセンター」が設立された。

現在受託している研究は、高リスク横紋筋肉腫、限局性ユーイング肉腫、急性リンパ性白血

病、肺癌の4つで、いずれも Phase II 単アーム試験である。

② 内部スタッフと委託、臨床試験における役割分担

小児がんデータセンターが行っている業務は、臨床試験プロトコル作成支援、症例登録業務、データ収集と蓄積、データの品質管理（モニタリング）、統計解析、試験結果の公表における支援などである。看護師のデータマネージャー2名、薬剤師のデータマネージャー2名、生物学士のシステム担当1名、事務2名（医長秘書兼務）を擁しているが、常勤の統計家はおらず、日本臨床研究支援ユニット（J-CRSU）へ委託する形で、統計等の専門知識を得ている。上記データマネージャーも J-CRSU で訓練を受けた後は独学で実践を行いつつ学んでいる状況である。

ハード面は、4台のパーソナルコンピュータを使用し、SAS をベースに機能する臨床試験管理ソフト「DEMAND」と進捗管理ソフト「KIWI」を使用している。基本的には紙ベースのCRFを使用しているが、一つの臨床試験では、USBメモリ内に内蔵された電子CRFを使用して、インターネット経由のデータ授受を行うペーパーレスの品質管理を試みている。

現在の受託している4つの研究は、異なる研究費によってスポンサーされており、それぞれが独立に人材を雇用し、データセンターに集約して役割分担するという形を取っている。

上記に加えて、データセンターでは臨床試験オペレーション業務も行っている（一部は予定）。具体的には、施設管理とパフォーマンス管理、広報、印刷、中央病理診断、二次利用のための検体保存、薬事安全管理（+研究計画管理）、財務管理、などである。これらの業務は米国の小児がんの多施設臨床試験グループである Children's Oncology Group のオペレーション・オフィスを見学したことも参考になっている。

表2. 小児がん臨床試験組織に必要なオペレーション機能

- 1) 研究グループ代表者からなる運営委員会
- 2) オペレーションセンター
- 3) 委員会事務局(会議計画、財務、出版、広報等)
- 4) 施設メンバー管理・パフォーマンス管理
- 5) 研究計画管理、薬事安全管理
- 6) リファレンスセンター・組織バンクの充実

③ 抱えている問題点など

上記データセンター業務・オペレーション業務のうち、統計業務や中央病理診断、検体保存などは別の研究支援組織にさらにアウトソーシングを行っている。また、小児がんの臨床試験は、疾患グループ毎に班研究がベースになっているが、いくつかの疾患グループに対しては、データセンター業務自体をアウトソーシングし、グループ間の標準化や橋渡しを小児がんデータセンターが担っているといえるものもある。

これらは、小児がんの疾患特異的な問題から発生している部分がある。例えば、稀少疾患であるため症例集積が困難、長期生存が望める故に観察期間が長期、多種多様、かつ多臓器に発生するため、試験データ管理の標準化が困難、研究組織の未熟性の問題、稀少疾患であるため研究費が僅少、といった理由である。結果として、疾患特異的・班研究ベースでグループ化が進み、独自の活動・非効率性・リソースの分断につながってしまっているともいえる。これを打開するためには、組織化・効率化された全国多施設のネットワークが必須である。これは、データセンターの充実と公的認知とともに、表2に示す各種オペレーション機能の充実によって実現されると考えられる。

5. 主に研究者自主研究のサポートを業務とする研究支援会社の紹介

① 行っている研究の種類と内容

「メディカル・リサーチ・サポート」は、臨床研究者が行う自主研究のサポートを目的とした会社である。本稿著者の一人である石川が行

う臨床研究や、全国の大学、研究所に所属する研究者が実施している食品成分や薬による発癌予防試験や、新しい内視鏡や手術機材などの比較試験、適正な内視鏡検査間隔の比較や、癌術後のコホート研究などのデータマネジメントを請け負っている。現在、進行中のものが大小含めて7つ、今年中に開始予定が5つある。新薬開発などの治験に関する業務の受託は行っていない。

② 内部スタッフと委託、臨床試験における役割分担

業務内容は、データの入力や集計などのデータマネジメント業務、受診勧告などの連絡、ランダム割付、プロトコルや配付資料の作成業務、統計解析、各種委員会の設定などである。現在、スタッフの数は事務員が4人、管理栄養士が2人であり、実施する試験に応じて、事務員は遺伝カウンセラーなどの勉強も行っている。

組織に専任の研究者がいないため、試験などの管理については受託している研究に石川が共同研究者として参加し、業務の指導や管理を行っている（石川はこの会社から給料他金銭的サポートを受けていない）。利点は、きわめてスリムな組織のため比較的費用が安くてすむことである。また、石川に加え組織のメンバーがこれまで大腸領域の研究を継続して行ってきたため、これまでの経験を活かしたプロトコル作成や試験実施のサポートができることが大きな利点として挙げられる。また、商業CROと異なり利潤追求を目的としていないため、オペレーション業務を含め、契約にとらわれず臨床研究者サイドに立ったサポートを行っていることも大きな特徴といえる。

③ 抱えている問題点など

もともと、特定の研究者個人が行う臨床研究をサポートすることを主な目的として作られた組織であるため、組織のパフォーマンスがその特定個人に大きく依存してしまう部分がある。例えば、他の研究者から個人的に業務を依頼さ

れた場合、契約内容や料金設定が両者の力関係によって決まってしまうたり、業務の質が個人の知識によって左右されてしまう可能性があることである。研究者サイドに立った組織であるため、委託する研究者にとって非常に有利な組織であり、依頼が殺到して特定個人に負担がかかりすぎると、それ以上の依頼を受けることができなくなる。しかしながら、将来、専任の研究者を擁する企業として発展し、なおかつ安定経営できるようになると、研究者主導臨床研究をサポートする組織の理想的な形態の一つとなるかもしれない。

6. 研究者が設立した臨床研究支援のための専門組織

- ① 行っている研究の種類と内容
- ② 内部スタッフと委託、臨床試験における役割分担

NPO 日本臨床研究支援ユニット (<http://www.crsu.org>) は、研究者主導の疫学研究・臨床研究を支援する目的で、本稿著者の一人である大橋が中心となって2001年に設立した。現在は、専任・パートのNPO職員、専任の派遣あるいは出向スタッフ、そして研究者（20%以上のエフォート）からなる混成部隊が3部門に分かれ、表3のような研究の支援を行っている。

NPOは他に教育事業（日本メディカルライター協会・SoCRA、スタッフ数3）と総務部門（2名）を抱え、事業規模は平成16年度で収入ベース約2億5千万である。支援の内容は（臨床試験の場合は）症例登録・データ管理・統計解析が中心であり、派遣モニタリング機能は持たない。研究グループに初期から参加しプロトコル作成にも積極的に関与していることが民間CROとの違いであろう。医師主導治験においては事務局業務を担当し調整役を担っている。現制度下で大きな負担となっている重篤有害事象の報告は別のCRO（J-DSRU）の協力を仰いでいる。

表 3. NPO 日本臨床研究支援ユニットによる研究の支援

- 1) CSPOR(乳癌臨床研究支援事業)
 - 大規模補助療法試験3、ヘルスアウトカム研究7、第II相試験2、国際共同試験2(ATLAS、IBISII)
 - スタッフ数:総括2(データセンター全体も管掌)、DM7、登録/進捗管理/入力3、解析1
- 2) 血液ユニット
 - 医師主導治験(造血幹細胞移植時の免疫抑制剤)2、厚労省班研究8
 - スタッフ数:総括1、DM3、登録/進捗管理/入力2
- 3) その他
 - 循環器大規模コホート研究と高齢者高血圧患者追跡研究(動脈硬化予防研究基金)、前立腺癌PSAスクリーニング研究、腎臓移植全例登録追跡、乳癌発症登録、骨粗鬆症大規模臨床試験(ATOP)1、臨床試験登録割付18、厚労省班研究による臨床試験(糖尿病2、小児腎症3)、膀胱癌診断研究
 - スタッフ数:16(総括1含む)

(<http://www.crsu.org>)

表 4. 単施設の臨床試験組織

- 1) 基本的に多施設の臨床試験組織と同じ
 - ・単施設であるため多施設にわたる標準化が不要
- 2) データマネジメントやオペレーション業務について多科でのリソース共有
 - ・オペレーション業務の透明性と公正さ
 - ・業務手順の共有と標準化
- 3) 医師、データマネージャー、CRCなどすべての人材に対する院内教育制度
- 4) Electric Data Capturingや病棟からのRemote Data Entryの活用
- 5) 自由度の高い研究費や資金の獲得
- 6) CRCやDMなどの安定雇用の仕組み
- 7) CROなど外部組織の利用と育成

③ 抱えている問題点など

運営上の最大の課題は、研究予算に制限がある中での人材の確保と教育であり、民間CROとの協力関係を模索している。東京大学のARO構想の中で、データ管理・プロジェクト管理の教育の場として機能させることが将来の目標である。

7. まとめ

単施設であろうと、多施設であろうと、質の高い臨床試験に必要な要素は基本的に同じである。要素によって単施設では実現しやすいものもあれば、しにくいものもあるため、多施設の臨床試験組織を参考にしながら、それぞれの要素をどのように担っていくかを実現することと

なる。表4に単施設の臨床試験組織を考える際に必要なポイントをまとめた。

臨床試験にデータマネジメントが必要であるということはもはや多くの研究者の理解するところとなった。データマネジメントを行うために、研究者が自前で行うにせよ委託するにせよ、とにかく人が必要であり、そのためにこそお金が必要であるという認識もできつつある。次には、人や業務を含めてデータマネジメントの質の管理、チェック体制が必要であるということも実際に臨床研究のデータマネジメントに携わったものでは理解できるところである。

しかし、臨床試験に必要なものはデータマネジメントだけではない。様々な第三者的監視を行うための委員会機能などが必要であることも

理解されつつある。しかしこれに加えて、臨床試験推進のためには、オペレーション機能が必須である。いろいろなプロジェクト・マネジメントも含め、試験の管理も必要であるし、ヘルプデスク的な機能、参加施設との対応も必要で、これなくしては臨床試験の円滑な推進はありえない。これらの機能をよく理解し、どれを自分たちで担うのか、どれを委託するのか、委託する場合には具体的に何を委託するのかを明確に特定しなければならない。また、これらの業務を委託する組織と友好的な関係を保つとともに、自らそれらを育てていく努力も必要であろう。

8. 謝 辞

この研究は本稿著者らが属する組織、研究班での活動を通して得たものであり、本稿著者のオリジナルというより、それらの経験を代表してまとめたものと言える。これまでの研究に協力してくれたすべての方々と、我々にこれらの経験を与えて頂いた多くの研究参加者の方に心から感謝し、この研究結果が今後の臨床研究の発展のために少しでも役立つことを祈ります。本研究は厚生労働省研究費（がん研究助成金、厚生労働科学研究費）の補助を受けた。

Data management outsourcing for investigator-initiated clinical trials

Seiichiro Yamamoto¹⁾, Haruhiko Fukuda¹⁾, Tetsuya Hamaguchi¹⁾,
Takuji Okusaka¹⁾, Atsushi Makimoto¹⁾, Hideki Ishikawa²⁾ and Yasuo Ohashi³⁾⁴⁾

¹⁾National Cancer Center

²⁾Department of Molecular-Targeting Cancer Prevention,
Kyoto Prefectural University of Medicine

³⁾University of Tokyo

⁴⁾Japan Clinical Research Support Unit

Key Words: clinical trial, data management, single institute, contract research organization

It has been well established that data management is an essential component in conducting clinical trials. To obtain data management of satisfactory quality, a certain amount of funding is needed, regardless of whether investigators have their own datacenter or a contract with a CRO (clinical research organization) for their data management. In addition to data management, an operations office is also necessary for smooth conduct of clinical trials. This office is responsible for the management of trials, institutions, investigators, budgets, and various projects. The principal investigators of a study should therefore familiarize themselves with all the functions necessary for clinical trials and decide who or what organization will be responsible for them.

米国 NCI 臨床試験の有害事象報告について

山本精一郎

国立がんセンターがん予防・検診研究センター情報研究部
JCOG データセンター統計部門チーフ

はじめに

筆者は 2003 年 7 月から 2004 年 12 月まで、米国 National Cancer Institute (NCI) の Division of Cancer Treatment and Diagnosis (DCTD) に在籍していた。DCTD は NCI の治療に関する extramural research を行う部門であり、自分たちは研究を行わない。NCI/DCTD はスポンサーとして国内外の研究施設や多施設共同臨床研究グループ (Cooperative group) が実施するすべての臨床試験のプロトコルレビュー、研究進捗確認、有害事象 (Adverse Event [AE]) 報告や有効性に関する情報の管理を行い、がん治療開発臨床研究の方向性をリードする。

その中で中心的な役割をする Clinical Trial Evaluation Program (CTEP) や Biometric Research Branch での経験から、NCI 試験において、研究者からの有害事象報告がどのように行われているかについて紹介したいと思う。

NCI/CTEP における Monitoring と Reporting

NCI/CTEP の CTMB (Clinical Trial Monitoring Branch) では、NCI/DCTD がスポンサーであるすべての試験 (以降 CTEP 試験と呼ぶ) の有害事象報告の管理と監査 (audit) を行っている。そこで以下の事項を中心にどのようなことが行われてい

るか順に紹介する。

1. CTEP 試験における有害事象報告 (expedited, routine reporting, AdEER) について
2. CTEP 試験に対する audit について
3. 以上の点における Food and Drug Administration (FDA) の規制との関係
4. 移植の試験など、AE が非常に多い場合に重要でない報告を減らせる可能性はあるか

まず前提として、CTEP 試験の多くは IND (Investigational New Drug) trial である。つまり、その後の「New Drug Application (NDA: 新薬承認申請) のために必要な要件=FDA の要求する要件」を満たしている必要がある。その意味で製薬会社のもと同じ質、少なくとも FDA の要求水準以上を満たしていると考えられる。

CTEP 試験の有害事象報告と audit の方法は、AE の Grade、AE は un/expected (予期されない/される) のどちらか、研究の相、薬剤が IND かどうかによって異なっている。簡単に言うと、IND を用いた CTEP Phase I trial は CTMS (Clinical Trial Monitoring Service) によって、それ以外は CDUS (Clinical Data Update System) によってモニタリングされる。AE 報告のサマリーは NCI guidelines: Expected adverse event reporting requirements for NCI investigational agents を参照されたい。

Monitoring/Reporting は外部委託機関が行っている

CTMS とは、CTEP の外部委託機関（現在は Theradex 社）が行うモニタリングのことである。CTEP はこの委託機関と 4 つの契約を結んでいる。1 つ目の契約（契約 1）は、CTEP の IND Phase I trial のモニタリングを行うことである。具体的には以下の通りである。

- ・各施設研究者より、2 週間ごとに AE に関する CRF (Case Report Form) が送られてくる
- ・委託機関はそれに対して不明点や missing に対する query/cleaning を行い、月に 1 回 CTEP に対し、レポートを送る
- ・CTEP の drug monitor はそれを検討する

これらを通常報告 (routine reporting) と呼ぶ。要は、on-site でなく in-house monitoring である。

2 つ目の契約（契約 2）は、各施設に対し、年に 3 回 audit を行うことである。1 回は委託機関のスタッフである PharmD/MD によるもので、あとの 2 回は同スタッフである PharmD/CRA によるものである。

各 audit において行うことは、sampled subjects の全報告の SDV (Source Document Verification) である。Unexpected のみとか、Grade 3-4 のみとかではなく、全て行う。もちろん毒性だけでなく、response や eligibility のチェック、IRB 書式関係の確認なども行う。ここで特筆すべきは、CTEP の IND Phase I trial の CRF の書式は委託機関がすべて作成している（というか、委託機関が作成したものを使っている）ことである。もちろん研究ごとに適格規準などの adaptation は行うものの、他の項目については基本的に同じで、研究によっては関係ない項目も含まれた CRF を用いている。これによって、2 週間ごとの報告の管理や画一した audit が可能となるとともに、施設研究者の手間を省き、かつ CTEP の IND Phase I trial の質を担保していると言える。また、この CRF の

electronic version として ACES というソフトがあり、施設研究者はこれに直接入力し、それを送ることもできる。また、施設で独自の clinical trial database を持っているところは、そこから ACES へのデータ変換を行うことによって、electronic submission をしているそうである。さらに施設独自の clinical trial database の中には、診療データベースの Lab データを自動抽出できるものもあるそうだ。

3 つ目の契約（契約 3）は、Phase II や Phase III を実施する米国の Cooperative group が自ら行う audit に同行することである。全体の 10% にあたる audit に参加し、問題があったときに行う re-audit にも参加する。

4 つ目の契約（契約 4）は、Non-Cooperative group の audit を行うことである。Cooperative group や契約 4 の audit の内容は契約 2 の audit (sampled subjects、全ての項目の SDV など) と同じだが、試験でなく施設を対象とし、3 年に 1 回行う点が異なっている。

CTEP の IND Phase I trial 以外の in-house monitoring を行う CDUS は、CTEP 試験と DCP (Division of Cancer Prevention) が行う臨床試験の primary data source のことである。つまり、CTEP 試験と DCP でサポートする全ての臨床試験の基本データがここに集められている。試験の相や治療試験かどうか、IND を使うかどうか、研究実施主体 (Cooperative group かどうか)、スポンサー (DCTD か、DCP か) によって、簡易 (abbreviated CDUS) か完全 (complete CDUS) かの違いがある。

CTMS trial は CTMS から CDUS へデータが送られるので、CDUS を行う必要はないが、他の試験はこのシステムを通して、通常の定期的な有害事象報告が行われている。CTMS と CDUS が NCI trial に対して NCI が管理する基本的な定期的な reporting/monitoring システムである。

有害事象の急送報告

次に CTEP 試験の急送報告システムである AdEERS (Adverse Event Expedited Reporting System) について説明する。これは CTEP IND trial の AE 急送報告の electronic submission system である。急送報告の対象となるのは unexpected で attribution が possible 以上の Grade 2-3 とすべての Grade 4-5 で、いつまでに報告すべきかというのが Phase I と Phase II-III で異なっている。登録時から AE があった人はそれ以上悪くなるかいったん回復してさらに悪くなった場合以外、報告しなくてよいこと (厳密には should NOT report) になっており、commercial agent (市販薬) についても別扱いになっている。Commercial agent と IND の combination の場合は IND 扱いであるが、commercial agent の場合には (そのみのアームなど)、FDA に MedWatch を通して報告し、そのコピーを NCI に送ることになっており、AdEERS は使わないようである。ちなみに、Cooperative group の班会議に行くと、NCI から人が出向いて AdEERS の使い方の講習会を行っている。

IND の場合、すべて FDA の規制に則っている。企業の治験は、全ての audit が施設でなく試験が対象となることと、in-house でなく、site-visit monitoring をしていることが大きく異なる点で、monitoring や audit でチェックすることは原則同じである。企業の場合、site-visit monitoring の際に全項目について SDV を行うのが原則であるので quality は当然高くなると思われるが、NCI が IND でも in-house monitoring しか行っていないということは FDA がそれでよいとしていると解釈できる。IND のうち、unexpected や重篤なものについては AdEERS を通して情報を得たあと、FDA に報告する。それとは別に IND の義務として、情報をまとめて annual report を FDA に提出する。Commercial agent の AE については前述の

通り (施設研究者から MedWatch 直接でなく、CTEP を介しているのかどうかは不明)。NCI/CTEP の CTMB の担当官に FDA の audit との違いについて聞いたところ、彼らはいろいろな audit を行っているし、施設研究者との電話の記録が残っているか、あるいはインキュベーターの温度管理表が残っているかとか “off the wall” な感じだ、との印象を持っていた。直接 FDA の担当官に聞かないと詳しいことはわからないが、NCI のこれまでの方法で特に問題ないようである。

重篤でない有害事象報告を減らす工夫

ここまでの一般論に加えて、造血幹細胞移植の臨床試験など毒性が頻発し、そのうちの重篤でないものに対して作業を減らせる可能性があるかについての NCI の考え方も紹介したい。

まず、ガイドラインであるが、NCI CTC ver. 2 では BMT trial について別の毒性規準があったが、CTCAE ver. 3 では削除された。理由は FAQ に記載されている (JCOG/JSCO の日本語訳には少ししかないが、改訂 FAQ でより詳しく記載されている)。

具体例を示してみよう。現在計画中の移植の CTEP 試験では、IND agent として vaccine、commercial agent としてフルダラピンを用いるものがある。その場合、commercial agent には関係あるが IND agent には関係のない AE の報告は行わなくてもよいかと FDA と交渉したが、FDA はどの AE がどの治療に attribute するかを決めるのは難しいので unwilling と言われたそうだ。しかし FDA は、protocol specific に報告が不要な AE を事前に決めればよい、とは言っている。ただ、たくさん試験を行う NCI としては “protocol-specific” だと管理しきれないことから、protocol-specific に記載することは行っていないようである。そこで妥協案として、IND が commercial agent による expected event を悪化させた場合 (頻度や重

症度が増した場合)のみ急送報告することを次回の改訂 AE guideline に含めようとしたらしいが、FDA との合意が得られず、2005 年 1 月 release の改訂版からは削除された。Protocol への事前記述に関しては後述する。

Protocol への予期される有害事象の記述

予期される AE に添付文書や治験薬概要書以外のもので論文報告があったものなどを含める、という考え方について解説したい。いったん FDA に承認され、妥当な研究や論文があれば、off label でも保険会社が reimburse するという彼らの EBM 土壌において、「添付文書」にあるものだけが予想できるもの、という考え方はそぐわない。もっとも、添付文書が right time に right way に update されているなら話は別である。CTEP 試験では、このような迅速な update が実現されている。

実際の protocol の記載例で確認すると、protocol に事前を書くことによって作業を減らしているのは程度の軽い expedited reporting であり、routine reporting ではないことがわかる。Routine reporting についての(一括)軽減は FDA が No と言っているようである。Expedited reporting については、expected であれば報告しなくてよいわけであり、expected list に expected なものを手間をいとわずたくさん載せればよいわけである。どのような場合 expected と扱うのかであるが、端的に言うと NCI が知りうる限りの情報を使って日々 update しているリストに載っているものが expected ということになる。CTEP では CTMS や AdEERS によって過去の試験も on-going trial の AE も全て管理している。そこで IND に対して新しく挙がってきた(あるいは新しい頻度) AE を CTEP のスタッフが review し、attribution を評価し、必要な場合はただちにそれを CAEPR (Comprehensive Adverse Events and Potential Risks

List) に追加する(これは私が在籍当時に移行中であつたリストで、これまでは ASAEL という名称であつた)。もちろん論文情報も同様である。protocol を作る際には、expected としてこのリストに載っているものをコピーして転載することになる。研究途中でもリストが update されれば protocol 上の記述を update する。この update の前後で当然 expedited reporting すべきかどうかが変わってくる。試験実施者側が勝手に expected かどうかを update していくことを FDA は認めているのかについては、“More on the other side”であり、もっと早く update すべきとのことであつた。Make sense である。

Expected と報告する元々の意義は、予期される AE を事前に予想して、実際に起こったときにしかるべき対処が取れるようにしておくためのもののはずである(報告するものを減らすためではない)。この意味で、すでに情報があるのに、治験薬概要書や添付文書に載っているもののみが予想される AE という考え方があるとすれば、それこそ本末転倒と言えるのではないだろうか。

治験の省力化とは

我が国ではどうであろうか。治験ではスポンサーによる site visit monitoring が行われており、医師主導治験でも site visit monitoring を実施している。Site visit monitoring が必要な場合もあると思うが、over quality でお金がかかりすぎるという批判も常にある。On-site と in-house monitoring を併用したり、NCI のように治験全体、あるいは医師主導治験全体に対し、外部委託のようなシステムを導入しての省力化は不可能であろうか。

厚生労働省・文部科学省が平成 15 年に公表し、17 年 4 月 1 日から施行されている「臨床研究の倫理指針」には治験以外の臨床試験に対しても、「研究責任者は、臨床研究機関の長に対し、重篤な有害事象その他の臨床研究の適正性及び信頼

性を確保するための調査に必要な情報を報告しなければならない」「研究責任者は、他の臨床研究機関と共同で臨床研究を実施する場合には、当該他の臨床研究機関の研究責任者に対し、臨床研究に起因する重篤な有害事象を報告しなければならない」とあり、研究機関の長はこれら AE 報告が提出された場合には、「臨床研究の実施又は継続の適否その他の臨床研究に関し必要な事項について、速やかに倫理審査委員会の意見を聞かなければならない」となっている。また、厚生労働科学研究費補助金やがん研究助成金では、研究の過程で健康危険情報を把握した場合には厚生労働省に報告することになっている。個々の研究者の力量に頼る形でこのような情報を集めるだけでは、個々の研究者にとって過大な負担となり、報告漏れにもつながってしまう。それよりも、国としてシステムティックに管理する仕組みが必要であろう。国としてシステムティックに安全性や有効性を把握しておくことは、研究開発の流れを把握することにもつながる。

似た様な観点ではあるが、予期される AE をリストにしておくのも素晴らしい考えだと思う。もちろんこれは、AE 情報を一括管理して、必要なときに利用するということから始まったのであろう。管理はできても、それを必要なときに利用することまではなかなか難しい。多施設共同臨床研

究グループで研究を行っているような場合には、新しい研究計画を考えたり protocol 作成のためだけでなくさらに一歩進んで、予期される AE をリスト化し、開示するなりして、皆で共有するようにはどうであろうか。そうすれば protocol に載せる情報の管理もその update も効率化できる。将来、効果安全性評価委員会の AE リストから予期される AE リストを更新、ということが routine になればよいと思う。

おわりに

臨床試験は、臨床的・科学的に重要な研究をすることがもっとも大切だが、それらから正しい結果を導き出すためには、いろいろなインフラの整備が必要であることは論を待たない。これには費用も手間、人もかかる。なるべく効率よく共有できるような仕組みが必須と考えるが、他人任せで待っていても進まない。人の仕事だと思わずに研究者自らが、グループそれぞれがシステムティックな把握システムをどのように構築するかを自分の周りの問題として捉えていただきたいと思う。

この原稿を書くにあたっては、厚生労働科学研究費補助金、並びに特定非営利活動法人日本臨床研究支援ユニットの移植医療臨床研究支援事業若手研究者派遣プログラムの支援を受けた。心より感謝申し上げます。

山本精一郎 Seiichiro Yamamoto

東京大学大学院医学系研究科保健学専攻博士課程を修了後、平成 8 年より、国立がんセンターにて、厚生労働省がん研究助成金指定研究班を中心とする多施設共同臨床試験グループである JCOG (Japan Clinical Oncology Group) 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。データの取り扱い方により無再発生存期間の推定にバイアスを与えることを明らかにした研究 (Jpn J Clin Oncol 2002) 等の成果は JCOG 臨床試験の方法論的整備、質的向上に貢献した。また、厚生労働省多目的コホート研究として、みそ汁やそれに含まれるイソフラボン摂取が女性の乳がんリスクの減少と関係があることを明らかにした研究 (J Natl Cancer Inst 2003) 等を行ってきた。

平成 15~16 年に米国 National Cancer Institute に客員研究員となり、米国でのがん臨床研究、疫学研究、予防研究の NCI の funding agency としての機能についての研究も行った (New Engl J Med 2005)。

臨床試験の方法論的研究並びに他分野研究者との共同研究、基盤整備に関する研究によって、がんの予防、診断、治療の各分野で、今後も独自の貢献を目指している。



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) を決めるといふ, 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つ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。いまずぐにでもphase IIIを行いたいが、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)²⁾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