

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 (それのみのアームなど)、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 としてフルダラビンを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。しかし FDA は、protocol specific に報告が不必要な 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も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ったリストで、これまでは ASAEEL という名称であった)。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った。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る。

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。

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²、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(CT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膵内科では、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。錠菓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.

2 サブグループに対する 治療開発のための臨床試験 デザイン

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)

P o i n t

- 個別化治療とは、何らかのマーカーに基づいてサブグループを定義し、そのグループに対して効果のある治療を行うことである。個別化治療であっても、それが有効であるかどうかは臨床試験による検証が必要となる。
- Phase III Trialでall comers design、
 - (a) ある薬剤が対象者全体の全生存期間を延長するような標準治療となり得るかどうかだけでなく
 - (b) 試験薬投与前に腫瘍組織で調べた薬剤に対する反応(マーカー)によって治療効果が予測できるかを主要な目的として研究デザインすることも可能である。
- 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(linear仮説)や、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(threshold仮説)、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。

分子標的薬などの開発が進むにつれ、個別化治療という言葉をよく耳にするようになった。個別化治療とは、何らかのマーカーに基づいてサブグループを定義し、そのグループに対して効果のある治療を行うことである。個別化治療であっても、それが有効であるかどうかは臨床

試験による検証が必要となる。個別化治療の臨床試験デザインとして、大きく分けて図1の3つが提案されている¹⁾。このうちのall comers designについて、本稿では深く掘り下げてみたい。

All comers designは、まずはじめにマーカーを調べるが、マーカー結果によらず、ランダム化

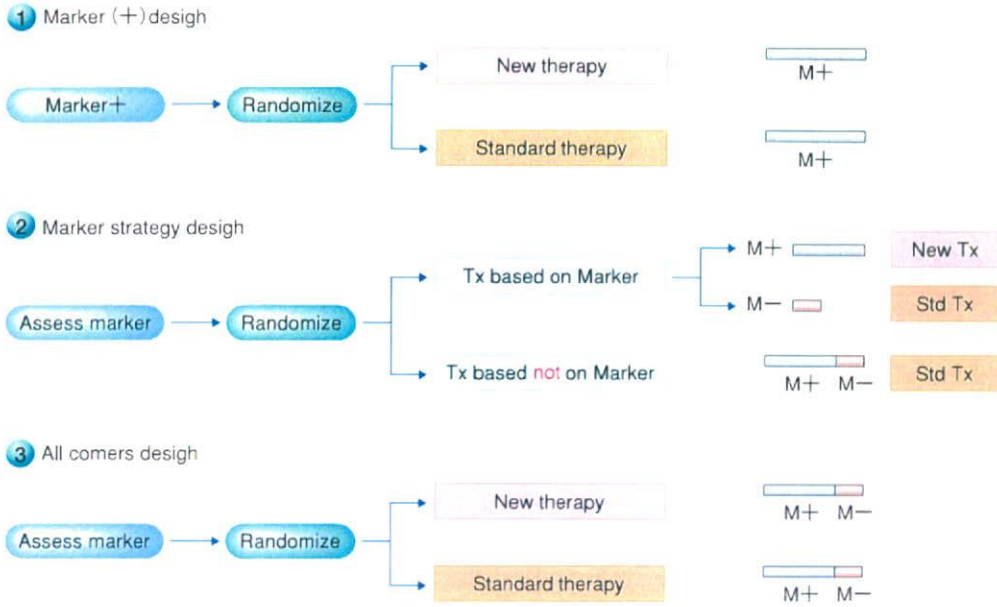


図1 ●個別化治療のための臨床試験デザイン

して一方には新治療、もう一方には標準治療を行うものである。マーカー結果によらず両群を比較することによって、対象集団全体に対し新治療と標準治療の治療効果のプライマリな比較ができるだけでなく、マーカー(+)群、(-)群のそれぞれで新治療の治療効果を調べることができる。つまり、マーカー結果と治療の間に交互作用があるかどうか、言い換えるとマーカーが治療効果の予測因子かどうかを調べることができる。全体で治療効果をみた後に単にサブグループアナリシスをするのではなく、さらに進めて、検証的なサブグループアナリシスを組み込むにはどうすればよいか、感受性仮説などの proof of concept をどのように行っていけばよいか、について統計的な試みがいくつか提案されているので、本稿ではそれらの方法を議論するこ

とによって、個別化治療の臨床試験デザインについて理解を深めていただきたいと思います。

Simon²⁾は、事前に規定したマーカー結果によって定義されるサブグループでの検定に対して、 α エラー^{*1}のコントロール、つまり検定の多重性^{*2}を調整するようなデザインを組めばいいの

*1 α エラーと β エラー

統計的検定において、帰無仮説が正しいときに誤ってこれを棄却してしまう確率を α エラーとよぶ。 α エラーの水準は5%とすることが慣習的である。また、対立仮説を誤って採択できない確率を β エラーとよぶ。 $1-\beta$ エラーが検出力となる。

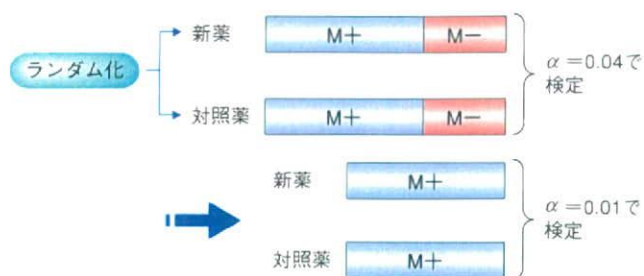
*2 検定の多重性

1つの研究のなかで検定を何度も繰り返すと α エラーが大きくなることをいう。検定の多重性を調整して試験全体の α エラーを名義水準通りに抑えることを多重性の調整という。臨床試験における検定の多重性とその調整の例として、中間解析と最終解析における複数回の検定、複数のエンドポイントの検定、などが挙げられる。

図2 ● Simonの提案

(Simon R, JCO 2005. より引用)

- 事前に規定したサブグループに対して多重性を調整
- 「マーカーと治療の交互作用を調べるデザイン」で2つの仮説を検証する
 - ・ 全対象者での群間比較: $\alpha=0.04$
 - ・ マーカー(+)のサブグループでの群間比較: $\alpha=0.01$



ではないかという提案を行った(図2)。これはマーカーと治療の交互作用を調べるデザインで、2つの仮説を検証する。つまり、はじめに全対象者をランダム化して、2群間比較の検定を行い、さらにマーカー(+)だけで群間比較の検定を行う。ポイントは、全対象者について行う検定は $\alpha=0.04$ でやっておいて、そのあとマーカー(+)のサブグループだけで検定するときには $\alpha=0.01$ で行うというものである。この論文のなかでは、厳密な統計的な議論はしておらず、簡単な提案のみ行っている。

これをさらに進めたものとして、Freidlin & Simon³⁾は“Adaptive Signature Design”という方法を提案している(図3)。この方法は、マーカーを試験開始前に特定できない場合に特に有効である。この方法は次の3ステップからなる。全員をランダム化し、新治療群と対照群に分けて治療を行うことは同様である。

a) まず、割付を行った全対象者に対して $\alpha=\alpha_1$ (<0.05)で2群間比較をする。全体の対象者で差があるかをみる部分である。

b) 試験前半(1st stage)で登録された対象者のデータに対し、さまざまなマーカーを用いて治療効果の異なるサブグループを探索的に探し出す(training set)。次に後半の対象者(2nd stage)のなかから同一のサブグループを抽出する(test set)。すなわち後半の対象患者のなかでこの薬剤に対して“感受性の高い”と考えられる対象者のサブグループを同定する。

c) サブグループを探索した1st stageのデータと独立である2nd stageのサブグループを用いて治療効果が異なるかどうかの検証を行う。すなわち、2nd stageのサブグループに対し、 $\alpha=\alpha_2$ で治療効果の群間比較の検定を行う(ただし、 $\alpha_1+\alpha_2=0.05$ となるように α_1 、 α_2 を定める)。

彼らは全対象者に対する検定を行う際の α を全体の α の80%、つまり $\alpha=0.05$ であれば $\alpha_1=0.04$ とし、探し出した“感受性の高い”と思われるサブグループに対する検定を行う際の $\alpha_2=0.01$ とすることを勧めている。この方法を用いると、試験全体の α エラーは0.05で抑えることができる。 $\alpha_1=0.04$ にすることによって、全体の対象者

- “Adaptive Signature Design”
- マーカーを試験開始前に特定できない場合
- 全被験者を対象に $\alpha = 0.04$ で群間比較して有意差なしの時
 - ・試験前半で登録された被験者を対象として、予測因子を探索
 - ・試験後半で登録された(予測因子の探索とは独立な)被験者を対象に、予測因子に基づくサブグループにおける群間差を有意水準 0.01 で検定

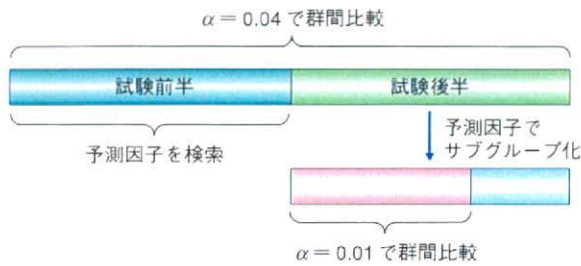


図3 ● Freidlin & Simonの提案

(Freidlin B & Simon R, Clin Cancer Res 2005. より引用)




に対する検出力は少し落ちるが、2nd stageの“感受性の高い”サブグループに対する検定を加えることによって、試験全体の検出力は増加することが期待される。“感受性が高い”(=十分な治療効果が期待できる)ことが本当であれば、そのサブグループでの群間差は全対象者での群間差よりも大きくなることが期待されるため、 $\alpha_2 = 0.01$ でもそれなりの検出力をもつであろうということである。また、 $\alpha_1 = 0.04$ で行う全対象者に対する検定の検出力を $\alpha = 0.05$ で検定する際の検出力と同様となるようサンプルサイズを多少増加させることもできる。その場合には“感受性の高い”サブグループに対する検定を加えた2つの検定を行うことによって、通常のデザインよりも有効な薬剤を誤って捨てる確率は少なくなることが期待できる。新規薬剤が多くの対象者に効果がある場合には、通常のデザインと同じ検出力をもち、マーカーで規定される少数のサブグループにしか効果がない場合にもそのサブグループでの感受性が高く治療効果が大きければ群間差を検出でき、試験全体としての検出力の増加が

見込めるためである。

National Cancer InstituteのこのSimonらのグループは、microarrayの解析やpharmacogenomicsの解析に積極的に取り組んでおり、そのなかでどのようにして予測因子を探索するかということに多くの経験をもっている。1st stageの対象者のデータを用いてさまざまなマーカーによって規定される感受性の高いサブグループを探索的に探す、というアイデアはその中から出てきたものと思われる。感受性の高いサブグループが見つかるかどうか、そのサブグループの割合がどうかといった結果に依存する部分や、1st stageと2nd stageのサンプルサイズ比をどう設定するか、など試験計画のデザイン時に最適化できない部分も多いが、試験開始前に感受性の高いサブグループが十分同定できない場合には有効な方法であると考えられる。

次に筆者が所属する日本臨床腫瘍研究グループ(Japan Clinical Oncology Group; JCOG)で現在計画されている研究デザインを紹介する。これは進行癌に対する化学療法の効果を調べる

表1 ●薬剤感受性に関する仮説

治療と感受性の関係を3レベルで考える	
一様な効果(交互作用なし)uniform effect ・薬剤の治療効果は感受性によらない	
閾値モデル threshold model ・薬剤の効果はマーカー値がある値を超えると発現する	
直線モデル linear model ・薬剤の効果はマーカー値と直線的に関連する	
ポイント	
治療とマーカーの交互作用として感受性仮説を検討	
想定した感受性仮説が正しければサンプルサイズは減少する	
感受性仮説の証明と臨床的意義は別	
帰無仮説は、治療効果なしと交互作用なしの両方が定義できる	

phase III trialを計画する際に検討されたデザインである。研究の主要な目的は、

a) ある薬剤が対象者全体の全生存期間を延長するような標準治療となり得るか。

であるが、

b) 試験薬投与前に腫瘍組織で調べた薬剤に対する反応(マーカー)によって治療効果が予測できるか。

ということも大きな関心である。もちろん、“腫瘍組織で調べた薬剤に対する反応”自体は単なるマーカーに過ぎず、生存期間延長といった治療効果と必ずしも一致するわけではない。bの仮説、すなわち“マーカーの値によって、治療効果が異なる”ことを本稿では“感受性仮説”とよぶことにする。

例となった実際のJCOG試験では、主要な目的であるa)を優先させるために、通常の2群間比較を数百例の規模で行い、副次的な目的として(多重性の調整は行わず)、b)の仮説を調べることにした。ここでは思考実験として、bの“感受性仮説”を主要な目的として検証することを考えてみる。治療への“感受性”との関連を期待するマーカーの値が連続的であり、3段階に分けて治

療効果との関係を調べることにする(表1)。3段階に分けると、さまざまな感受性仮説が検証しやすいからである。感受性仮説の例を以下に挙げる。

①linear仮説：マーカー値が「低」「中」「高」となるにつれ治療効果が大きくなる。

②threshold仮説1：マーカーが「中」以上であれば治療効果が発揮される。

③threshold仮説2：マーカー「高」であれば治療効果が発揮される。

感受性がない場合にも以下の2つの仮説が考えられる。

④uniform仮説：マーカー値が「低」「中」「高」に対し、治療効果が一定。

⑤全体の帰無仮説：どのサブグループに対して治療効果がない。

通常、臨床試験で行われる検定は、“⑤全体の帰無仮説”に対して“④uniform仮説”を対立仮説として行われる。繰り返しになるが、統計的に言うと感受性仮説は治療とマーカーの生存期間に対する交互作用のことを指す。つまり、治療効果があっても感受性がないことを示すuniform仮説は、統計的な交互作用がないことを示す。

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、次の標準治療はどうなるのだろうか(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

るといふ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しなければいけない。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、今までに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.

Effect of the introduction of minimum lesion size on interobserver reproducibility using RECIST guidelines in non-small cell lung cancer patients

Hirokazu Watanabe,^{1,6} Hideo Kunitoh,² Seiichiro Yamamoto,⁵ Shin Kawasaki,² Akira Inoue,² Katsuyuki Hotta,² Kazu Shiomi,³ Masahiko Kusumoto,⁴ Kazuro Sugimura¹ and Nagahiro Saijo²

¹Department of Radiology, Kobe University Graduate School of Medicine, 7-5-2 Kusunoki-cho, Chuo-ku, Kobe 650-0017;

²Department of Medical Oncology, ³Thoracic Surgery, and ⁴Radiology, National Cancer Center Hospital; and ⁵Research Center for Cancer Prevention and Screening, Statistics and Cancer Control, National Cancer Center, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

(Received September 17, 2005/Revised November 24, 2005/Accepted November 28, 2005/Online publication February 16, 2006)

We evaluated interobserver reproducibility for the response evaluation criteria in solid tumors (RECIST) guidelines and the influence of minimum lesion size (MLS) on reproducibility. The 110 consecutive patients with non-small cell lung cancer were treated with platinum-based chemotherapy. Five observers measured target lesions according to both the World Health Organization (WHO) criteria and RECIST. The percentage changes for unidimensional measurements (UD; RECIST type) and bidimensional measurements (BD; WHO type) were calculated for each patient. Interobserver reproducibility among five observers, that is 10 pairs, was expressed as the Spearman's correlation coefficient for the percentage changes, the proportion of agreement and the kappa statistics for response categories. The same analysis was carried out using MLS. BD was more reproducible than UD (Spearman rank correlation coefficient, 0.84 vs 0.81; proportion of agreement, 84.4% vs 82.5%; kappa value, 0.69 vs 0.61). When MLS was applied to UD, eligible cases decreased by 6.4% and the number of target lesions by 44.6%, whereas interobserver reproducibility for UD improved (Spearman rank correlation coefficient, 0.81–0.84; proportion of agreement, 82.5–84.2%; kappa value, 0.61–0.65). The introduction of MLS to UD could also improve intercriteria reproducibility between WHO and RECIST. It is important to apply the MLS when using RECIST for the comparable interobserver reproducibility attained with WHO. (*Cancer Sci* 2006; 97: 214–218)

Tumor response to chemotherapy was previously evaluated using the WHO criteria, which stipulate bidimensional measurement (BD; WHO type) of lesions.⁽¹⁾ With these standardized criteria for evaluating tumor response, valid and reproducible results could be obtained by all investigators. However, a number of modifications to the WHO criteria have been developed by different institutions, which made it difficult to compare response rates for screening new anticancer agents across different investigators. This has led to the introduction of a new system, the RECIST guidelines,⁽²⁾ which have been widely accepted as the new standard.

In order to standardize the methodology for evaluating tumor response, RECIST simplified the response evaluation through the use of unidimensional measurements (UD; RECIST type) instead of the BD used by the WHO criteria. Furthermore, the

MLS allowable for measurement at baseline study was defined as being no less than double the slice thickness on CT or MRI.

The validity and intercriteria reproducibility between the new RECIST guidelines and the previous WHO criteria have been investigated.^(2–7) However, to the best of our knowledge, no analysis of the influence of MLS on interobserver reproducibility, specified for measurability in tumor response evaluation according to the RECIST guidelines, has been published in the literature.

The purpose of the present study was therefore to evaluate interobserver reproducibility in tumor response evaluation using RECIST, intercriteria reproducibility between WHO and RECIST, and whether this reproducibility is affected by the application of MLS.

Materials and Methods

Patient population

This is a retrospective study of the radiological findings for patients who underwent chemotherapy for advanced NSCLC. The subjects were patients treated during clinical trials at the Medical Oncology Division of the National Cancer Center Hospital in Tokyo, Japan, between January 1999 and January 2001. All clinical trials were conducted in accordance with the Helsinki Declaration and the protocols were approved by the institutional review board. Written informed consent was obtained from each patient for the treatment protocols, which included the secondary use of treatment-associated documents. Patients were staged according to the Union Internationale Contra le Cancer TNM classification of malignant tumors.⁽⁸⁾ The 110 eligible patients included those histologically or cytologically diagnosed with NSCLC. Patients were required to undergo CT scans periodically for evaluating tumor response prior to and once after

⁶To whom correspondence should be addressed. E-mail: h-watanabe@kcc.zaq.ne.jp
Abbreviations: BD, bidimensional measurements; CR, complete response; CT, computed tomography; MLS, minimum lesion size; MRI, magnetic resonance imaging; NSCLC, non-small cell lung cancer; PD, progressive disease; PR, partial response; RECIST, response evaluation criteria in solid tumors; SD, stable disease; TNM, tumor node metastases; UD, unidimensional measurements; WHO, World Health Organization.