

ものであるが、その他にも(b)標準治療では成績が十分でない集団を選別して、その集団に追加治療の有効性を検討する、逆に(c)標準治療のみで成績が十分によく、過剰治療(over treatment)の可能性がある集団を選別して、その集団に治療が手控えられるかどうかを調べる、といった場合にも適用できる。ただし、エンリッチメントデザインではマーカー(-)に対するデータは得られないので、マーカーが真の意味で治療効果の予測因子となるかの評価は行えない[マーカー(-)でも新治療が有効である可能性を否定できないため]。先行研究のデータや明確な生物学的根拠などから、マーカー(-)を試験に含めることが合理的ではないと判断できる状況に適したデザインである。

試験開始時にマーカーのバリデーションができていない場合は、マーカー検査の結果にかかわらずに登録を行い、全体効果(overall treatment effect)の検証およびマーカーの検証を同時に実施することが求められる。すなわち、マーカーの妥当性評価も行う解析計画が必要になる。たとえば、以下のようなデザインが考えられる⁹⁾。

2) 交互作用の検定に基づく解析

このデザインでは、治療法(新治療, 標準治療)×サブグループ(マーカー(+), (-))の交互作用の検定をまず実施する。交互作用の検定が有意とならなければ、全例を対象として新治療 vs. 標準治療の検定を行う(全体効果の検証)。交互作用の検定が有意となれば、マーカー検査の結果別に(すなわちサブグループ別に)新治療 vs. 標準治療の検定を行う。これは二元配置分散分析の伝統的なアプローチに従った方法といえる。

3) 有意水準の分割

Simon & Wang⁷⁾によるアプローチでは、全例を対象とした新治療 vs. 標準治療の検定を有意水準 α_1 (たとえば $\alpha_1 = 0.04$) でまず実施する。この検定が有意となれば、新治療が全体に対して有効であると結論する。有意とならない場合にはマーカー(+)に対して、新治療 vs. 標準治療の検定を有意水準 $\alpha_2 =$

$\alpha - \alpha_1$ (たとえば $\alpha_2 = 0.01$) で行い、この検定が有意となれば、新治療はマーカー(+)に対して有効であると結論する。2回検定を行うことになるが、有意水準の分割によって試験全体の有意水準を $\alpha_1 + \alpha_2 = \alpha$ ($0.04 + 0.01 = 0.05$) 以下に抑えることができる。すなわち、有意水準の分割によって検定の多重性を避けることができる。このデザインでは、新治療の効果を広い対象に対して期待しているが、万一のときに(全体効果が示されなかったときに)検証的に機能するサブセット解析をあらかじめ準備しておくものである。全体における検定とマーカー(+)における検定の相関を考慮したより正確な有意水準も研究されている⁹⁾。

しかし、現実には試験開始時にマーカーに関する情報がほとんどない、あるいはマーカーとなりうる連続量の測定値は利用できても(例えば遺伝子発現量など)、その閾値[マーカー(+)または(-)に2群化するためのカットオフ値]が不明であるため、マーカー(+)の集団さえ定められない、という事態に直面しうる。こういった状況に対処するため、1つの試験のなかでマーカー(ないしその閾値)を確立し、かつ(そのマーカーが定める)マーカー(+)における治療効果の検証までいっぺんに行ってしまうデザインも提唱されている^{9,10)}。

4) 適応的 signature デザイン

このデザイン(adaptive signature design)では、所定の追跡期間が終了した後、試験前半部分で登録された症例(図1のA)を対象に、適切な統計手法を用いて予測マーカーの構築を行う。そのうえで、主たる解析は以下の2ステップをとる。全例を対象とした新治療 vs. 標準治療の検定を有意水準 α_1 で実施する(図1のA+B+Cで新治療 vs. 標準治療の検定を行う)。この検定が有意となれば、新治療が全体に対して有効であると結論する。有意とならない場合には構築されたマーカーを用いて、試験後半部分で登録された症例(図1のB+C)の中からマーカー(+) (図1のB)の集団を選別し(エンリッチメントし)、この集団で新治療 vs. 標準治療の検定を有意水準 $\alpha_2 = \alpha - \alpha_1$ で行う(図1のBで新治療 vs. 標準

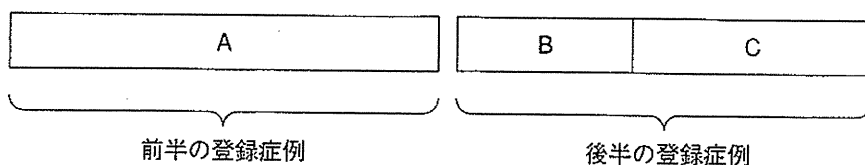


図 1 適応的 signature デザイン

治療の検定を行う)。この検定が有意となれば、新治療はマーカー(+)に対して有効であると結論する。

すなわち第1ステップとして全体効果の検証を行い、そこで有効性を証明できなければ、引き続き第2ステップとしてマーカー(+)における治療効果の検証を行う。上記 Simon & Wang によるアプローチと同様の手順であるが、利用可能なマーカーの存在を前提としているかどうかの点でまったく異なる。マーカーの構築も同じ試験内で行ってしまおうというのが適応的 signature デザインの主旨である。もし試験開始前にすでにいくつかのマーカー候補が存在する場合には、そのなかからいずれか1つを統計的に選択して(マーカーを新たに構築するという最初の手順を、候補マーカーのなかから1つ選択するという手順に置き換える)、第2ステップでその選択されたマーカーを用いることもできる。薬剤設計のために標的分子は同定していても実際の患者標本では何をマーカーとして測定すべきかが明確ではなく、たとえば点突然変異や遺伝子増幅、タンパク過剰発現の有無など、マーカー候補が複数存在しているケースは多いので、そうした場合にもこのデザインは有効である。

デザイン上重要なのは、「マーカー構築のためのデータ」と「マーカー(+)での治療効果を検証するためのデータ」を分離している点である(図1でマーカー構築のための集合Aとマーカー(+)での治療効果を検証するための集合Bは別である)。これらを分離しないと、マーカーの予測精度を過大評価してしまうことが知られている。一方、全体効果の検証ではマーカーに依存した検定を行うわけではないので、全例を対象にして解析することに問題はない。

5) 適応的バイオマーカー閾値デザイン

試験開始時にマーカーとなる連続量の測定値は利

用可能だが、閾値が定まっていない場合を想定したデザイン(biomarker-adaptive threshold design)も提案されている。1つの試験内で、第1ステップで全例を対象とした全体効果の検証、第2ステップで最適な閾値を探索し、その閾値から適応的に定めたマーカー(+)に対する検証も計画するというものである。適応的 signature デザインと同様、1つの試験内でマーカーやその閾値を決定し、その妥当性評価まで行ってしまうことの是非はあるが、治療開発の速度を上げる有効な方法となりうるものである。

複数のマーカーを組み合わせる治療戦略を決定し、その戦略自体の妥当性を評価したいような場合には、以下のデザインが有用である。

6) マーカーベース戦略デザイン

このデザイン(marker-based strategy design)は、ランダム化された2群のうち、一方の群はマーカーに基づく治療戦略をとり、もう一方の群は一律に標準治療を行うといった状況を想定している(当然、後者の群がコントロールとなる)¹¹⁾。たとえば、複数マーカーを組み合わせる治療選択する戦略の検証に利用できる。進行期非小細胞肺癌を対象とした MAdEIT 試験¹²⁾は、ゲムシタビン、カルボプラチンの感受性にそれぞれ関連のある RRMI, ERCCI の発現レベル(high, low)を組み合わせ、都合4通りのサブグループ毎にレジメンを変える治療戦略をとる群と、一律に(標準治療として)ゲムシタビン+カルボプラチンを行う群を比較する研究として実施されており、マーカーベース戦略デザインと考えられる(図2)。ただし、このデザインはエンリッチメントデザインと同様、マーカーの妥当性自体を直接示すためのものではない¹³⁾。マーカーが規定するサブグループに対して最適と思われる治療法を定める“治療戦略”そのものを評価しようとしているデザ

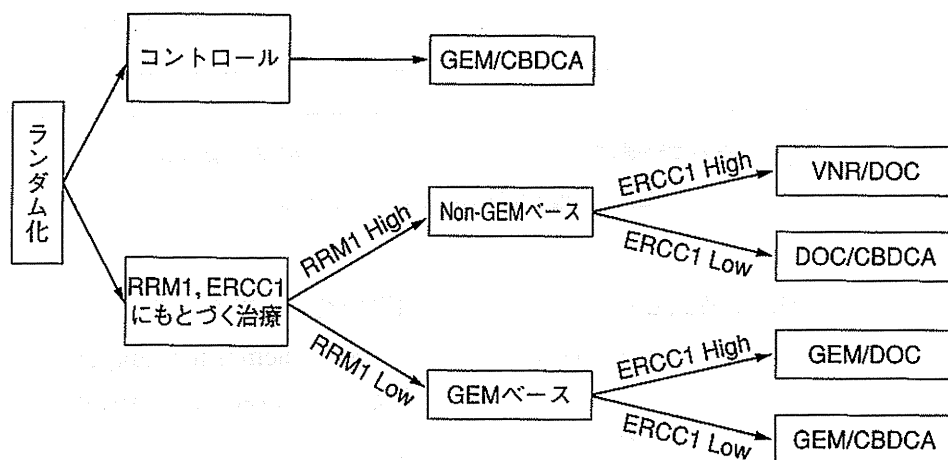


図2 遺伝子発現レベルに基づくマーカーベース戦略デザイン

インである。

2. 第Ⅱ相試験

分子標的薬の開発が進められるなか、いかに第Ⅱ相試験を実施して第Ⅲ相試験の成功へとつなげるかに関して様々な議論がなされている。「第Ⅱ相試験の方法論には多く変更の余地があり、より効率的に薬剤開発を行うためにいかに改良していくか?」という議論であり、バイオマーカーの役割についても予測因子やエンドポイントなど複数の観点から提案がなされている¹⁴⁾。これまでの分子標的薬の第Ⅱ相試験(ただし単剤試験に限る)のレビュー結果に基づき、奏効率をエンドポイントとした非ランダム化試験での実施を推奨する向きもあるが^{15,16)}、個々の試験が抱える状況にも依存するため画一的な議論は難しい。しかし、(是非は別として)従来どおりの単純な設定で行うにしても、今後はバイオマーカーを積極的に取り入れた試験実施が必要となるであろう。第Ⅱ相段階でマーカーを用いた患者の絞り込み(エンリッチメント)まで実施するのは困難であるにしても、候補マーカーを評価して次相で活用できるような情報を得るための試験計画は可能である。たとえば、マーカー(+)と(-)では奏効率が異なることをあらかじめ想定して古典的なサイモンの2段階法(two-stage design)を拡張するデザイン¹⁷⁾や、全体に対する奏効率が基準をクリアできなかった場合に

マーカー(+)に対する奏効率の評価へと移行するタンドムデザイン(tandem design)¹⁸⁾などが提案されている。

ベイズ流デザインの適用への注目も高まっている¹⁹⁾。たとえば、バイオマーカーに関する情報をランダム化の際の割付け確率に逐次的に反映させるアダプティブランダム化を採用したBATTLE試験²⁰⁾などのような事例も出てきている。BATTLE試験は既治療の進行期非小細胞肺癌を対象としたランダム化第Ⅱ相試験で、EGFR変異・K-ras変異・VEGF発現など、複数マーカーの組み合わせによって患者集団を計5つのサブグループA~Eに分割する。サブグループ別に4つの分子標的薬治療の評価を行うことが目的であるが、各サブグループで4つの治療法の効き目は異なることが予想される(たとえば、AグループはEGFR変異を有するのでエルロチニブが効きやすいetc.)。そこで各サブグループにおいて効果がより期待できると思われる治療法により多くの患者が割付けられるように割付け確率を逐次変更してランダム化を行うのである(したがって、登録実施中に割付け確率は変化し、またA~Eごとの割付け確率が異なってもよい)。割付け確率の変更はベイズ流に、すなわち、割付け確率を変更しようとする時点までに得られているアウトカム情報をもとに変更する。この手順で試験を進めてデータを集積し、A~Eごとに4治療法の有効率(BATTLE試験では1時点での無増悪生存率)を推定することが目的である。従来のデザインを改良し

て、より効率的な実施を試みる目的で、今後、ベイズ流デザインの早期フェーズへの適用は増加していくことが予想される。

おわりに

一部の患者サブグループにベネフィットをもたらす治療法の開発には、標的集団を選別するためのバイオマーカー、およびそれを明示的に組み込んだ臨床試験のデザインが必要である。ここではいくつかのデザインの整理を試みたが、少なくとも本稿の執筆段階では上記に述べた方法を含め、標的集団の存在を意識したデザインによる試験の経験はほとんど積まれていない。分子標的薬時代の到来がわれわれに要請する「治療法とバイオマーカーの前向きな同時開発」は新たな試みであり、試験の複雑性が増す、さらなる方法論の研究、それを可能にする試験実施体制の充実、試験実施を通じた経験の蓄積が求められる。

●文献

- 1) Saijo N, Takeuchi M, Kunitoh H. Reasons for response differences seen in the V15-32, INTEREST and IPASS trials. *Nat Rev Clin Oncol*. 2009; 6: 287-94.
- 2) 柴田大朗. 選択的治療を可能にする方法論 生物統計学の側面から. *Mebio Oncology*. 2008; 5: 4-9.
- 3) Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med*. 2001; 344: 783-92.
- 4) Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin Cancer Res*. 2004; 10: 6759-63.
- 5) Maitournam A, Simon R. On the efficiency of targeted clinical trials. *Stat Med*. 2005; 24: 329-39.
- 6) Simon R. The use of genomics in clinical trial design. *Clin Cancer Res*. 2008; 14: 5984-93.
- 7) Simon R, Wang SJ. Use of genomic signatures in therapeutics development in oncology and other diseases. *Pharmacogenomics J*. 2006; 6: 166-73.
- 8) Song Y, Chi GY. A method for testing a prespecified subgroup in clinical trials. *Stat Med*. 2007; 26: 3535-49.
- 9) 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.
- 10) Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J Natl Cancer Inst*. 2007; 99: 1036-43.
- 11) Sargent DJ, Conley BA, Allegra C, et al. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol*. 2005; 23: 2020-7.
- 12) Simon GR, Begum M, Bepko G. Setting the stage for tailored chemotherapy in the management of non-small cell lung cancer. *Future Oncol*. 2008; 4: 51-9.
- 13) Hoering A, Leblanc M, Crowley JJ. Randomized phase III clinical trial designs for targeted agents. *Clin Cancer Res*. 2008; 14: 4358-67.
- 14) McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into phase II trials. *Clin Cancer Res*. 2009; 15: 1898-905.
- 15) El-Maraghi RH, Eisenhauer EA. Review of phase II trial designs used in studies of molecular targeted agents: outcomes and predictors of success in phase III. *J Clin Oncol*. 2008; 26: 1346-54.
- 16) Booth CM, Calvert AH, Giaccone G, et al. Design and conduct of phase II studies of targeted anticancer therapy: recommendations from the task force on methodology for the development of innovative cancer therapies (MDICT). *Eur J Cancer*. 2008; 44: 25-9.
- 17) Jones CL, Holmgren E. An adaptive Simon Two-Stage Design for Phase 2 studies of targeted therapies. *Contemp Clin Trials*. 2007; 28: 654-61.
- 18) Puzstai L, Anderson K, Hess KR. Pharmacogenomic predictor discovery in phase II clinical trials for breast cancer. *Clin Cancer Res*. 2007; 13: 6080-6.
- 19) Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov*. 2006; 5: 27-36.
- 20) Kim E, Herbst R, Lee J, et al. Phase II randomized study of biomarker-directed treatment for non-small cell lung cancer (NSCLC): The BATTLE (Biomarker-Integrated Approaches of Targeted Therapy for Lung Cancer Elimination) clinical trial program. *J Clin Oncol*. 2009; 27(15s): abstr 8024.

〈山中竹春 山本精一郎〉



研修医からの質問 Q & A

今回の質問者：文田壮一(近畿大学医学部内科学講座腫瘍内科学部門)
吉田えり(三沢市立三沢病院腫瘍内科)

● 最新の臨床試験結果はどこで仕入れるものなのでしょうか？どのような場合に実際に使用できるエビデンスと言えるのでしょうか？

■ A 最新の臨床試験結果は国際学会や学術雑誌で発表され、学会などでコンセンサスを得ることによって実際に使用されるエビデンスとなります。

最新の臨床試験結果は、国際学会などで発表されます。最も大きなものは米国臨床腫瘍学会(American Society of Clinical Oncology, ASCO)で、最近ではASCO-GI(消化器がん)などその分科会でも新しい研究結果の発表が行われます。血液がんでは米国血液学会(American Society of Hematology, ASH)、乳がんではサンアントニオ乳癌シンポジウム、なども代表的な発表の場です。これらの学会やピアレビュー雑誌に新しい治療が有効であるという信頼できる結果が掲載されると、学会もしくは、さまざまな場で議論が行われ、診療ガイドラインなどでそれらが新しい標準治療としてコンセンサスが得られると実際に使われるようになります。通常、信頼できる結果とはこれまでの標準治療を対照群としたランダム化比較試験の成績です。広く使われるためには、保険でカバーされるようになる(されている)ことも条件となります。

● 回答者：山本精一郎(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)

特集

婦人科がん臨床試験参加に必要な知識

2. がん臨床試験の生物統計学

山本精一郎*

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*

要旨

Phase1 臨床試験の目的は、試験治療の最適用量を決めるために毒性によるスクリーニングを行うことである。Phase2 試験の目的である Phase3 へ進むかどうかの決定には、毒性の情報やほかの開発薬剤の状況なども加味するので、統計的判断は有効性の足切りの目安として使うのが一般的である。Phase3 試験の論文を読む際には、primary endpoint に関する結果が重要であり、有意かどうか、どこまでいえるか（何が検証的で何が探索的か）を理解しておくことが必要である。

Key Words 生物統計学, 臨床試験デザイン

はじめに

本稿では、がんの臨床医ががんの臨床試験の論文を読むのに必要な知識、おもに統計学的側面について知っておくべきことを述べる。これらはもちろん、臨床試験に参加する場合にも知っておくべき知識であり、自ら臨床試験を計画する研究者はもっと深く知る必要がある。また、論文を読むには統計だけでは不十分なので研究デザインについて理解することが必要であるため、デザインについても説明する。逆に、統計の概念自体を本稿のみで説明するのは難しいため、参考図書も合わせて学習していただきたい。

Phase1 臨床試験論文を読むのに必要な統計

Phase1 臨床試験の目的は、試験治療の最適用量を決めることである。細胞障害性抗がん剤の場合には、毒性と有効性が比例することを利用して、毒性によるスクリーニングを行うことによって、推奨用量を決定してきた(図1)。この目的のために最もよく用いられるデザインは3例コホートである。これは、3例ずつ同一用量を投与し、毒性で評価した最大耐用量となるまで用量を増加する方法である(図2)。3例コホートデザインは、それを用いて実際に臨床試験を計画し、実施する際にも、また、3例コホートを用いた臨床試験論文を読むためにも統計的な知識は特に必要ない。3例コホートは、2/6の対象者に用量制限毒性(Dose Limiting Toxic-

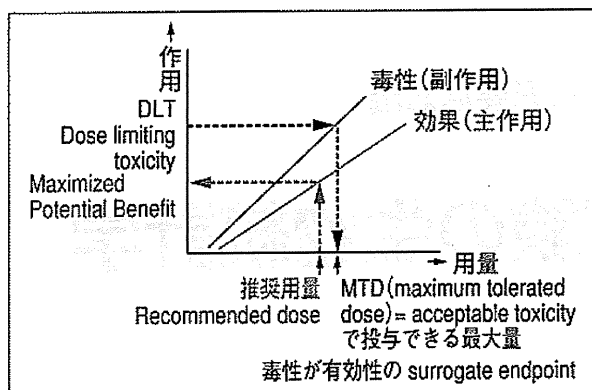


図1 細胞障害性抗がん剤 p1 デザインの rationale “More is better” with acceptable toxicity

ity: DLT) が発現すると中止するようなデザインとなっているため、抗がん剤としては DLT が 33%あたりが限界ラインと想定しているのかもしれないが、統計的に考察すると、実際には 33%よりも早めに中止となることが多いと考えられる。3 例コホートに統計的な根拠はあまりないが、この方法でこれまで開発を行ってきた、特に問題が生じず、やりやすくわかりやすいからということで、現在でもよく用いられる経験則的な方法であるといえる。3 例コホートは細胞障害性抗がん剤のために開発されてきたデザインであるが、分子標的薬のように、必ずしも毒性と有効性が比例するとは限らない場合にも、(明らかに優れた方法がほかにないため) 標準的に用いられているようである。

3 例コホートはかなり低い用量から開始し、順に増量するため、なかなか最適用量に達しない、最適用量付近でのサンプルサイズが小さい、といった問題が指摘されている。そこで、この点を解消するために提案されている方法が CRM (continuous reassessment method) である。この方法は、1 例ごとに統計的に用量反応曲線を推定し、それに基づいて最適と思われる用量を投与、その結果に基づいて用量反応曲線を推定し直し、次の対象者の用量を定める、を繰り返す方法である。ベイズ流統計学の手法を用いているが、3 例コホートに比べ最適用量付近が

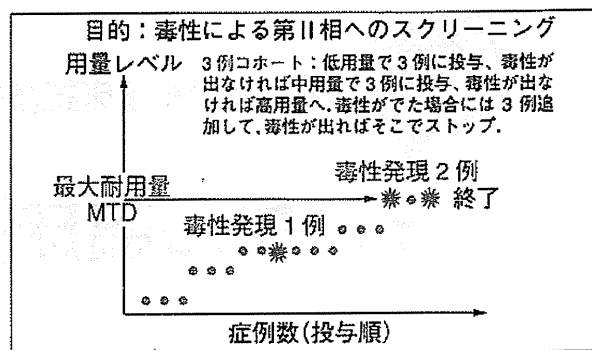


図2 第I相臨床試験の典型的なデザイン [By courtesy of Dr. Ishizuka]

選択されやすい優れた方法であることを知っていればよい。実際に実施するには、3 例コホートに比べ、手順がかなり複雑になることもあって、それほど頻繁には用いられていない。

Phase2 臨床試験論文を読むのに必要な統計

Phase2 臨床試験の目的は、試験治療の開発を続け、標準治療を対照とした第 III 相試験 Phase3 を行うかどうかの判断を行うことである。Phase3 へ進むかどうかの決定には、毒性の情報やほかの開発薬剤の状況なども加味するので、統計的判断は有効性の足切りの目安として使うのが一般的である。シングルアーム試験やランダム化選択デザインなど通常 100 例以下のデザインが用いられることが多いが、検証的な結果ではなく、いずれにしろ、その先に phase3 試験が必要であることを理解すべきである。

おもなデザインは、閾値・期待値を用いたシングルアームのデザインである。これは、primary endpoint (奏効割合 response rate など) に対して、それ以下なら開発を中止すると考える値を閾値として設定し、真の response rate がある値 (期待値) であった場合には、実際に得られるデータによって閾値以下であることが高い確率で統計的に棄却できるようにサンプルサイ

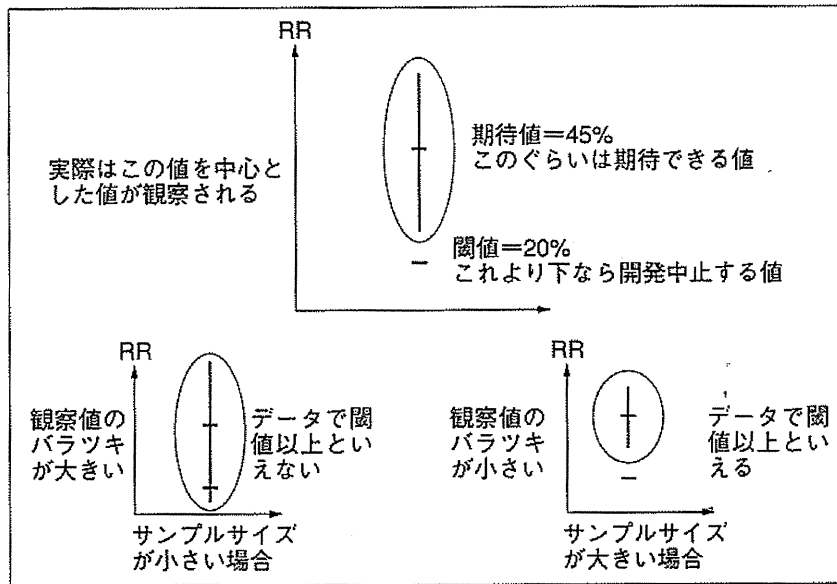


図3 閾値、期待値を用いたサンプルサイズ設計

ズと decision rule を設定するデザインである (図3)。閾値以下であるということが棄却されれば統計的には有効性が期待できるということになる。このデザインのオプションとして、試験の最後まで待たずに、途中ですでに無効が証明された場合には登録を終了する 2 stage design が用いられることも多い。

phase2 臨床試験の論文を読む場合に必要なのは、試験に実際にエントリーされた対象者の成績が、対照としている治療の成績と比較可能かどうかをチェックすることである。たとえば、ヒストリカルコントロールとの比較の場合、実際に試験にエントリーされた対象者とヒストリカルコントロールの対象者がどのくらい比較可能かをチェックすることが重要である。適格規準に記載された対象者が必ずしも満遍なく試験にエントリーされるわけではないため、実際に試験に登録された対象者がどのような属性の分布をもっているか、それがヒストリカルなデータとの間で大きく異なっていないかを検討する必要がある。特に time-to-event (生存や無増悪生存期間) をエンドポイントとする場合には、ヒストリカルコントロールのデータがよほど

stable でないと比較できない。time-to-event のデータは、試験に組み入れられた対象者がどういいう集団かによって大きく異なるからである。より比較可能性を担保するために、ヒストリカルコントロールでなく、Phase2 試験であっても、ランダム化した対照群を設定する場合がある。この場合に注意すべきなのは、ランダム化した対照群があったとしても、サンプルサイズが小さいので検証的でなく、あくまで Phase3 へのスクリーニングと考えることである。

対照群を置くためではなく、新治療同士をランダム化比較するランダム化選択デザイン (randomized selection design) が用いられることもある。これは、患者を2つ以上の試験治療にランダムに割り付け、最も高い奏効割合が得られた治療を選択し、Phase3 試験の候補とするものである。つまり、2つ以上の試験治療がある時に優先順位をつけるデザインということができる。このデザインでは、非常に promising な治療法があった場合、その治療法を高い確率で選択できるようにサンプルサイズと decision rule を決定するものである。図4の例のように、真の response が Drug A で 35%、Drug B で 20%

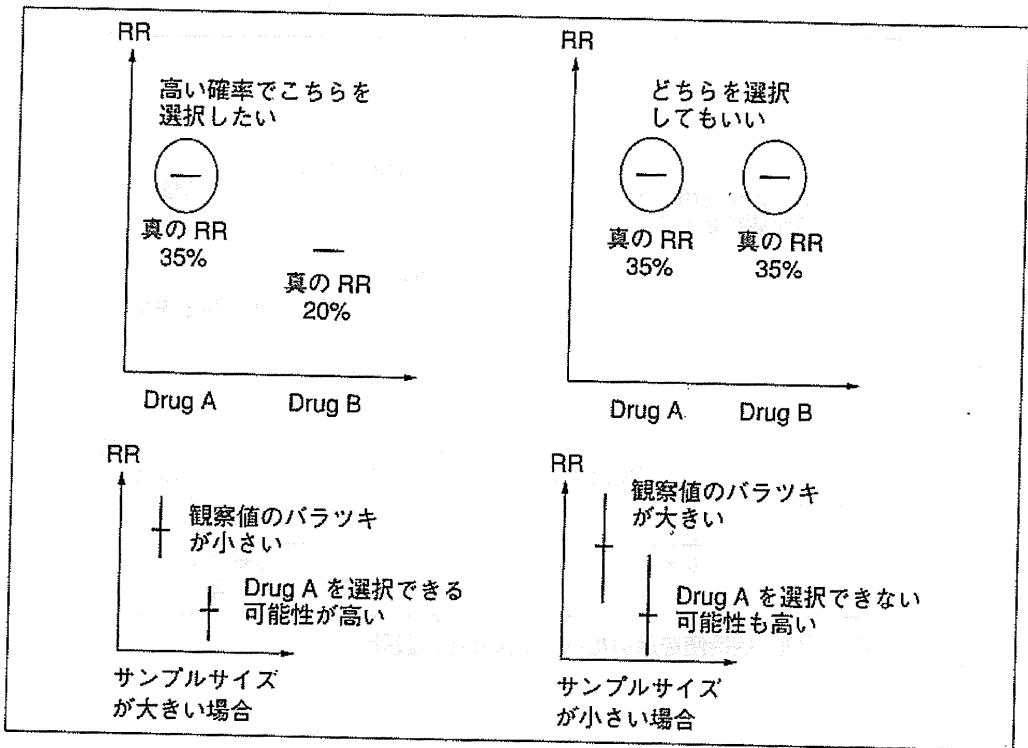


図4 randomized Phase2 デザインの考え方

である場合には、Drug A を 90% の確率で選択できるためには 29 例が必要となる。このデザインで注意すべきことは、どちらも同じ有効性をもつ治療法の場合には、50% の確率で強制的にどちらかを選択してしまうことである。差がない場合にどちらかに優劣をつけることが間違いとするならば、 α エラーが 100% といってもいいだろう。すなわち、まったく検証的試験とはいえないデザインであり、有効性に大きな差がない場合には、有効性の高い治療法を正しく選べない確率も高くなってしまふ。

繰り返すが、Phase2 臨床試験の結果によって、Phase3 へ進むかどうか（閾値・期待値）、どのレジメンを選択するか（選択デザイン）の決定するには、毒性の情報や他の開発薬剤の状況なども考慮する必要がある。統計的判断は有効性の足切りの目安として使うことが一般的である。サンプルサイズの目安として、シングルアームだとだいたい数 10 名（ <50 ）であることが多く、患者数が少ない疾患などで、次に

phase3 を実施することが現実的に不可能で、やむを得ず Phase2 試験に検証的な意味をもたせたい場合には 100 名以上とすることもある。ランダム化比較している場合には、だいたい 100 名程度であり、ランダム化選択デザインだと 100 名以下であろう。いずれにしろ、検証的な結果ではないことを理解することが大切である。

Phase3 臨床試験論文を読むのに必要な統計

Phase3 の論文を読む際には、まず abstract で研究の概略と質を把握することが効率的である。研究仮説は何か、primary endpoint は何か、デザインはどうか、試験が予定通り完遂しているか、などが評価のポイントである。研究仮説として、比較する治療は何と何か、どちらの治療が標準としているかを把握する。適切でない治療を標準治療としてしまっている試験は結果の解釈が困難である。次に、primary endpoint として最も確実である全生存期間（overall sur-

vival: OS), すなわちあらゆる死因による死亡までの期間, を用いているか(死因別死亡はあまり使われない)を調べる, progression-free survival (PFS), disease-free survival (DFS) や time-to-treatment failure (TTF) を用いている場合はそれらがその疾患領域で確立されているエンドポイントであるか知ることが重要である。そうでなければ, 必ずしも検証的な研究結果としてとらえてはいけなことになる。デザインは優越性試験か非劣性試験か, 新しい治療が toxic new なら優越性, less toxic new なら非劣性デザインを使う必要がある。非劣性試験の場合, 臨床的に意味のある secondary endpoint が定義されていて, それが標準治療に勝っていることが必要となる。試験が完遂しているかも大きなポイントである。予定通りのサンプルサイズで終了しているか, 途中終了の場合, 理由は何か, また, 追跡は十分かなどもチェックするポイントである。abstract にこれらの記載がない場合は, 本文の中でチェックすることになる。

introduction では, 研究を行うに至った背景, 特に対象と標準治療の決め方, なぜその新治療が有効であると考えられるかの説明の妥当性を見る。それらを踏まえ, どのようなデザインが適当かを自分で考えてみることはよい勉強になるであろう。

methods では, どのような試験グループが行っているか, eligibility の記載は正確か, 統計記述は十分かなどをチェックする。たとえば, sample size 設計の根拠がきちんと書かれているかといったところである。優越性試験であれば, 期待する治療効果が大きすぎないかをチェックすべきであるが, これは研究の結果, 統計的な差が認められなかった場合に検出力不足のせいかどうかを検討するためである。非劣性試験であれば, 設定している許容域が広すぎるかどうかをチェックすべきで, 広すぎる場合

には, 必要以上に統計的に有意になりやすいことになる。中間解析のやり方を含め, 解析方法がきちんと書かれているかも重要なポイントである。また, データマネジメントについて記述があるかなども研究の質を推測する手掛かりになる。最近の論文では, CONSORT 声明に従った研究の流れ図が記載されている論文も多いが, これは途中で脱落した人がどのくらいいるか, 治療のコンタミネーションがどのくらい起きているか, Intention-to-Treat (ITT) 解析をしているか, などを容易に把握することができ, 研究の質を評価するのに有用である。自分でこの流れ図を書いてみるのもよい勉強になる。

results では, まず背景因子の比較が示されることが一般的である。ここでは, 重要な予後因子に関して, 群間に大きなアンバランスがないかどうかを確認する。もし大きなアンバランスがあれば, 結果の解釈の際に注意すべきであるし, 統計モデルでアンバランスな要因を調整した解析結果なども参考にすることになる。背景因子ごとに p 値が示されていることも多いが, 割り付け調整因子になっていない場合には, 20に1つは統計的に有意にアンバランスになってしまうので, p 値自体にあまり神経質になる必要はない。ランダム化で期待しているのは, 「同じ治療をすとした場合の予後が等しくなる」ことであって, 個々の背景因子のバランスではない。アンバランスがある場合でも, 最も重要な結果はランダム化に基づいた解析であり, 統計モデルで調整した解析はあくまで確認のための結果であると考えべきである。

背景因子の比較の次には, 有害事象・毒性や治療のコンプライアンス, 有効性の secondary endpoint の結果等が順に示されることが一般的である。もともと, 有害事象に関してサンプルサイズ設計を行っているわけではないため, 有害事象が両群で統計的に有意に異なるかどうかの p 値は参考程度にすべきであって, 有意かど

うかだけで判断すべきでなく、より臨床的な知識を用いてその差を検討すべきである。治療のコンプライアンスについては、経験上、正確に定義することと、それにあったデータを収集するのが難しいため、そのことに注意して解釈する必要がある。有効性の secondary endpoint として response rate が記載されている場合も多いが、Phase3 では response rate はあくまで参考値とすべきである。なぜなら、response rate はもともと、おもに Phase2 において新治療の有効性をほかの治療法と比較するために用いることを目的としたエンドポイントであり、phase3 では全員について計測されていないことや central review されていないことも多いからである。QOL について報告されている場合もあるが、QOL 自体を primary endpoint にした研究でなければ、これも参考程度と考えるべきである。patient report の QOL が重要であることは間違いないが、測定方法やデータの収集、適切な解析など全部をクリアし、正しい結論であることを保証することが非常に難しいからである。

最も重要なのは primary endpoint の結果である。前述の通り、primary endpoint としては全生存期間 OS が最も望ましく、PFS が primary endpoint である場合でも、OS は大きな意味をもつ。用いられる統計手法は、Kaplan-Meier curve による生存曲線の比較が一般的であり、打ち切りがどのくらいあるか（ひげはついていない場合もある）がデータの maturity の目安となる。生存曲線の比較には、Logrank 検定を用いることが多い。Generalized Wilcoxon 検定を用いることもあるが、これは logrank 検定に比べ、初期のイベントに大きな重みをおいている方法である。検定結果が有意であるかどうかとともに、効果の大きさ（ハザード比や Median Survival Time (MST) の差) についても着目する必要がある。Cox 回帰は、単変量、多変量ともに、ハザード比を出すのに用いられ、前述のように、

多変量調整した結果はあくまで補助的なものと考えらるべきである。

discussion では、結果の解釈について議論される。優越性デザインの場合、試験がしっかり計画、実施、解析された場合には、primary endpoint で有意な結果が得られた場合、新治療が勝っていると判断する。その際、効果の大きさについても考慮することが重要である。primary endpoint で有意な結果が得られなかった場合には、対象とした集団全体では新治療が勝っているとはいえない。この場合にも、探索的にサブグループ解析をして、次につながる仮説を立てることが重要である。また、論文を読んでみて、試験がしっかり計画、実施、解析されていない場合と考えられる場合には、結果の解釈に注意が必要である。サンプルサイズが小さい、途中でリクルートできず終わった試験などは、有意であっても偶然の可能性も高いため、探索的な試験として考えるほうがよいといえる。

非劣性試験の場合には、結果の解釈に注意が必要である。非劣性試験とは、試験治療の primary endpoint (たとえば OS) が、標準治療と比べて「許容できる差」以内にあるかどうかを検証するデザインである (図5)。許容できる差以上に劣っている、という仮説を棄却することによって、許容できる差以内であることを検証する。この場合、有害事象が少ないなど、ほかにメリットがあるからこそ、OS は同等であればよいといえる。実際には、OS で同等以上が期待できる新治療に対し、ほかにメリットがあるために非劣性デザインを組むことが多い。非劣性デザインでは、試験がしっかり計画、実施、解析された場合には、primary endpoint で有意な差、かつ secondary endpoint で新治療の有効性が示されている場合、新治療が優れている、あるいは標準治療のオプションになりうるかと判断することになる。この場合、非劣性の許容域が

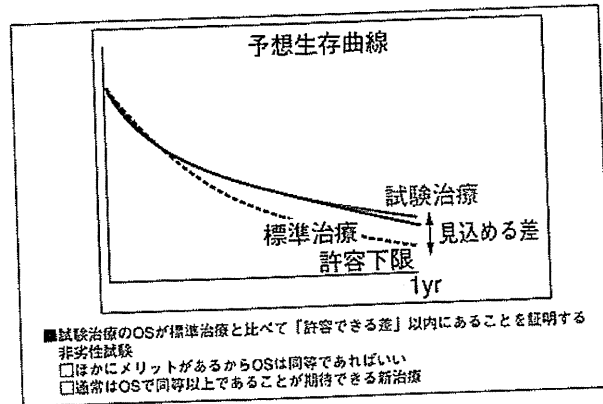


図5 非劣性試験とは

広すぎないこと（5年生存率で10%やハザード比1.5など）に注意する必要がある。primary endpointでは有意であったが、secondary endpointで新治療の有効性が示されていない場合、標準治療のオプションになり得るかは判断が難しい。また、primary endpointで有意な結果が得られなかった場合には、新治療は新しい標準治療、もしくはそのオプションとはなれないと考えるべきである。ただし、信頼区間で解釈して許容域を再定義することは解釈のうえで有用である。たとえば、5年生存率で3%劣っていないという許容域は棄却できなかったが、5%は劣っていないことは信頼区間による検討からいえる場合がある。この場合、5%に臨床的意義がある場合には、ある程度有効であるという判断をしてもよいであろう。もし、試験がしっかり計画、実施、解析されていないと考えられるような場合には注意が必要である。なぜなら、非劣性試験は、きっちり行わないと有意になりやすい（非劣性が証明されやすい）からである。これは、治療の不遵守で両群が同じ治療をしたため、測定の変異があると差が消えていって同じ成績に近づいていくからである。

試験のprimaryな結果に加えてdiscussionでよく議論されるのは、サブグループ解析の結果である。サブグループ解析をしているいろいろな検討をすることは非常に重要なことであるが、結

果の解釈には慎重になるべきである。2007年に出版されたNew England Journal of Medicineのサブグループ解析の報告に関するガイドラインが解釈の参考になる。これには、サブグループの結果をabstractに報告してよいのは、それらがprimary endpointに対するものである場合、かつ、事前に規定したサブグループ解析を解析全体として解釈できる場合に限る、としている。また、methodsにサブグループ解析の数（報告数、実施数）を明記すること、事前に規定して実施したものとそうではなく事後的に実施したものを区別すること、エンドポイントと解析法の明記、特に興味のあるものをhomogeneityの評価と区別して記載、 α エラーへの影響の程度とそれに対する対応法の明記、治療効果の不均一性の評価には交互作用に対する検定を行うことが規定されている。discussionでは、過度の解釈を避ける、信憑性の評価、限界を記載、結果を支持/矛盾する先行研究結果を提示するなど、細かい規定が設けられている。一読することを勧めたい。

Phase3論文を読む際の注意をまとめると、abstractでは、研究の概略と質（研究仮説は何か、primary endpointは何か、デザインは適切か、試験が予定通り完遂しているかなど）を把握し、introductionでは、研究を行うに至った背景、特に対象と標準治療の決め方、なぜその新治療が

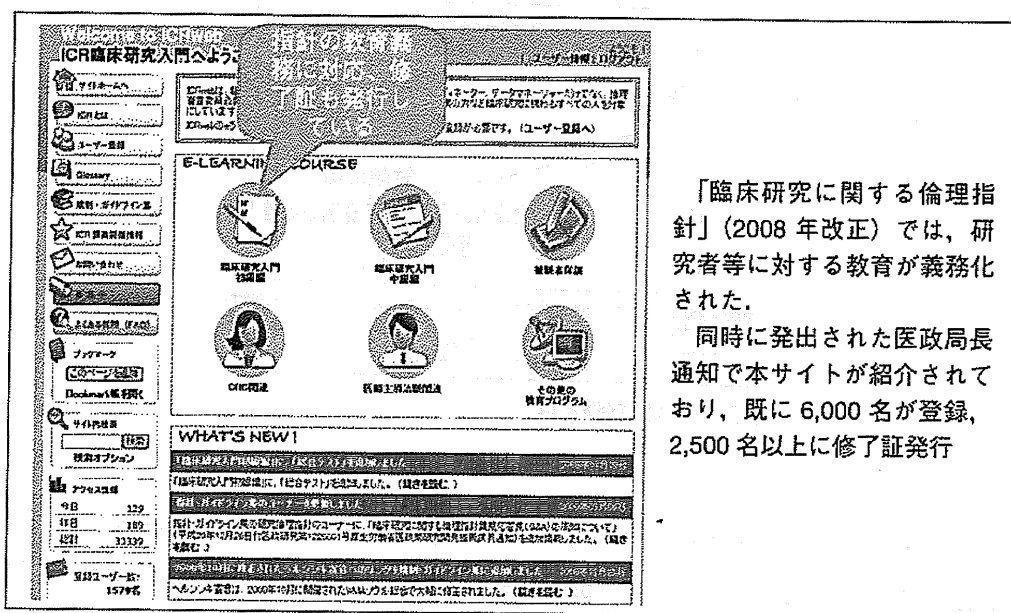


図6 臨床研究入門 e-learning サイト (略称: ICRweb) <http://www.icrweb.jp>

「臨床研究に関する倫理指針」(2008年改正)では、研究者等に対する教育が義務化された。

同時に発出された医政局長通知で本サイトが紹介されており、既に6,000名が登録、2,500名以上に修了証発行

有効であると考えられるかを把握する。methodsでは、どんなグループが実施しているか、eligibilityの記載は正確か、統計記述は十分か、データマネジメントについて記述しているかなどに注目する。その際、研究の流れ図を参考にすることが有用である。resultsとdiscussionでは、overall survivalの結果が重要であり、有意かどうか、どこまでいえるか(何が検証的で何が探索的か)を理解しておくことが必要である。

おわりに

ここまで、臨床試験の論文を読む際に注意することについて述べてきた。これは臨床腫瘍学会の教育セミナーBセッションで毎年話している内容で、臨床試験の論文を読むだけでなく、臨床試験に参加する際にも知っておくべき知識である。臨床試験に携わる際には、臨床試験だけでなく、非臨床試験や被験者保護、観察研究などについても理解しておくことが重要である。厚生科研費によるe-learningサイトICRweb(<http://www.icrweb.jp/>)は、これらの分野の基礎的概念を網羅しているので、ぜひご利

用いただきたい(図6)。当サイトは、厚生労働省の臨床研究に関する倫理指針改正に伴う医政局長通知のなかで、臨床試験を行う際に必要な教育を受けられるサイトとしても紹介されたものである。すでに6,000人以上の臨床研究者にご登録いただき、2,500人以上の方に初級編の修了認定を発行している。結構骨があるので、ぜひ挑戦していただきたい。

●文献

- 1) 福田治彦, 大橋靖雄: Japanese Journal of Clinical Oncology 投稿に際しての統計解析結果のレポートに関するガイドライン (<http://www.jjco.org/UsefulInfo/guideline.html>)
- 2) Wang R, et al: Statistics in medicine—reporting of subgroup analyses in clinical trials. *New Eng J Med* 2007; 357: pp2189–2194.
- 3) Green S, Crowley J, Benedetti: (翻訳福田治彦他) 米国 SWOG に学ぶがん臨床試験の実践—臨床医と統計家の協力をめざして. 医学書院 2004.
- 4) 大橋靖雄, 荒川義弘(編): 臨床試験のすすめ方. 南江堂 2006.
- 5) ICR 臨床試験入門: website (<http://www.cancerinfo.jp/ICR/index.htm>)

著者連絡先

〒104-0045
東京都中央区築地5-1-1
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
山本精一郎



解説

研究倫理と被験者保護： 国内外における現状と今後の方向性*

山本精一郎**

Key Words : research ethics, human subject protection, research ethics committee, institutional review board, ethics guideline

研究倫理をめぐる国内の状況

2009年4月より、厚生労働省の「臨床研究に関する倫理指針」(以下、臨床指針)が改正となり、研究者は臨床研究を始める前に臨床研究に関する倫理についての教育を受けること、臨床研究機関の長は教育の機会を提供することが義務となり、倫理審査委員会の委員も教育を受けることが努力目標となった。施行から1年が経ち、各施設でも試行錯誤の中、研究倫理に関する教育が始まっているようである。各施設で教育を行うことが義務と急に言われても、どのような教育を行っていいかわからないと想定されることから、臨床指針の改正とともに、医政局長通知として、教育機会の提供の例として、2つのe-learningサイトが紹介されている。1つは日本医師会治験促進センターによる「臨床試験のためのe training center (<https://etrain.jmacct.med.or.jp/>)」であり、もう1つはわれわれの研究班による「臨床研究入門ICRweb (<http://www.icrweb.jp/>)」である。われわれのサイトは、主に日本臨床腫瘍研究グループ(JCOG)での経験をもとに、JCOGで行われている研究倫理教育をより噛み砕いたものとなっている。それに加えて、われわれのサイトでは、米国の被験者保護局(Office for Human Research Protections; OHRP)(後述)が、研究機関の長、倫理審査委員長に対し受講を義

務づけている教材、米国NIH(National Institutes of Health, 国立衛生研究所)の倫理審査委員会委員向けの被験者保護教育コースの翻訳なども行っている。そこで本稿では、ICRwebで提供している研究倫理・被験者保護教育の内容を中心に紹介したい。

研究倫理に関する歴史

新しい治療法を開発するためには、ヒトを対象として観察・実験を行うことが必要である。なぜなら、その他の方法(動物実験や基礎研究など)だけでは、治療効果や副作用を予想することができないからである。臨床研究は治療法や公衆衛生を発展、改善させ、公共の利益につながる望ましいことであるからこそ、行うことが正当化される。正しい知識を得るためには、正しい方法で臨床研究が行われなくてはならない。そのためには臨床研究を科学的に行うことが必須となる。さらに、少数の個人が他人あるいは社会の利益のために、研究の被験者として負担やリスクを背負わされる(搾取の可能性)ために、倫理的配慮が必要となる。科学的でなければ正しい研究結果を導くことができないため、結果を世の中に還元できず、臨床研究を行うことが正当化できない。つまり、科学的であることは研究が倫理的であるための必要条件である。しかしそれだけでは倫理的には十分でない。臨床

* Research ethics and human subject protection in clinical research : history and current standard of Japan and USA.

** Seichiro YAMAMOTO, M.D.: 国立がん研究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); Cancer Information Services and Surveillance Division, Center for Cancer Control and Information Services, National Cancer Research Center, Tokyo 104-0045, JAPAN

研究は、科学的かつ倫理的に行う必要があるといえる。

しかしながら、このような考え方は昔からあったわけではなく、今までの歴史の中で人を対象とした非倫理的な臨床研究がたくさん行われてきた中で、その反省と再発防止のために整理されてきた考え方である。これまで行われてきた非倫理的な臨床研究の中で最も有名なものは、第2次世界大戦中のナチスドイツの人体実験であろう。超高度実験、低体温実験、マラリア実験など多くの人体実験を行って多数の犠牲者が出た。これに関して、第2次世界大戦後、ニュルンベルグ継続裁判で、約23名(医師20名)が裁かれた。この裁判記録の一部が、「ニュルンベルグ綱領」(1947年)として公表され、それが「ヘルシンキ宣言」(1964年)などの基礎となっている。

そのあと、アメリカでもいろいろな事件が起こった。なかでも有名なのは1932年から1972年にかけて行われたタスキギー梅毒研究である。この研究は、米国厚生省公衆衛生局(PHS)がアフリカ系の低所得者を対象として、梅毒の自然経過を観察するものであった。1932年の段階では有効な治療法がなかったので、最初は正当化されたのかもしれないが、1943年に梅毒の治療法(ペニシリン)が利用可能となったあとも、被験者に対してはまったくペニシリンを使用せずに無治療で観察を継続した。その間に、悪い血を抜くといったりして採血をしたり、無治療での観察を繰り返し行っていた。このタスキギー梅毒研究が、マスコミによって明らかにされ、非常に批判を受けることとなった。この大きな批判に対して、エドワード・ケネディ上院議員が1973年に「ヘルスケアおよび人対象試験の質に関する公聴会」を開催した。この公聴会では、タスキギー事件のみではなく、ウィローブロック肝炎研究などのさまざまな事件についての検討が行われた。そして、この公聴会の結果を受けて、アメリカでは国家研究法(National Research Act)が成立することとなった。この中で、いわゆるIRB(施設倫理審査委員会, Institutional Review Board)の設置が義務づけられており、これが、結果として米国でのIRBシステムの確立につながっていく。国家研究法には、人体実験規則

の公布の要求、つまり、法律を作ることが規定されており、これが連邦行政令第45編第46部(45CFR46)の制定につながっていく。さらに、被験者保護のための国家委員会を設立することが規定され、この国家委員会は精力的に活動して、17冊の報告書を刊行した。その1つが「ベルモント・レポート」である。

医療倫理の4原則

「ベルモント・レポート」は1978年に刊行された研究倫理に関する3つの倫理原則を提唱したということで非常に有名なレポートである。3つの原則とは、人格の尊重(respect for persons)、善行(beneficence)、正義(justice)である。さらに、「ベルモント・レポート」の作成にかかわったビーチャムらが、1979年に「生物医学・医療倫理の諸原則」を発表した。ここでは、医療全体に関して使用することができる4原則を提唱している。4原則とは、自律尊重(autonomy)、無危害(non-maleficence)、善行(beneficence)、正義・公正(justice)である。

自律尊重(autonomy)とは、患者の自律的な意思決定を尊重しようということである。この自律尊重(autonomy)の原則が、臨床研究におけるインフォームドコンセントの基礎となる。善行(beneficence)原則とは、患者に利益をもたらすということである。臨床研究に当てはめれば、ベネフィットを最大化ということになるであろう。無危害(non-maleficence)原則とは、患者に危害が及ぶことを避けよということである。臨床研究に当てはめればリスクの最小化となるであろう。最後の正義・公正(justice)原則とは、利益とリスク・費用を公平に分配せよということであり、臨床研究でいえば、被験者の公平な選定といったことになると考えられる。

4原則は、状況によっては対立してしまう。たとえば末期がんの患者を対象とした臨床試験での予後の告知などがその例に当たる。自立尊重の原則によれば、試験参加者は十分な情報を与えられるべきであるが、無危害原則によれば、予後告知によって精神的苦痛を与えるべきではないとなり、2つの原則が対立してしまう。原則が対立した場合には、どちらかの原則を優先せ

倫理要件		詳細要件	
Collaborative partnership 共同パートナーシップ	この 順 番 に 考 え る !	研究成果はコミュニティで共有 成果のみの搾取は防止 関係者は協調して研究を行う	
Social or scientific value 社会的/科学的価値		診断・予防・治療の向上に貢献, 疾患・健康に有用な知識を得る, すでにある知識や無駄な重複ではない	
Scientific validity 科学的妥当性		一般的に認められた科学的方法論 適切な統計手法・正しいデータ	専門家
Fair subject selection 適正な被験者選択		社会的弱者の保護・過大なリスクのある被験者の除外 被験者のリスクに見合う被験者/社会のベネフィット	市民委員
Favorable risk-benefit ratio 適切なリスク/ベネフィットバランス		リスクの最小化・ベネフィットの最大化 被験者のリスクに見合う被験者/社会のベネフィット	
Independent review 第三者審査		研究と利害関係を持たない独立した第三者による デザイン・対象・リスク/ベネフィットの評価	
Informed consent インフォームドコンセント		研究目的・方法・リスク/ベネフィット・代替治療の十分な説明 (information), 理解 (comprehension), 自発同意 (voluntariness)	事務局
Respect for potential and enrolled subjects (候補者を含む) 被験者の尊重		同意撤回の自由, プライバシー保護 開始後の新知見や研究結果の説明 継続的な被験者保護の監視	

図1 臨床研究の8つの倫理要件

ざるを得ない。そのような場合、ある状況では原則を適用しない(限定)、どちらが重要か場合によって判断する(比較考量)、もともと原則に優先順位づけしておく、といった方法があるようであるが、4原則という考え方をを用いることによって複雑な問題を解決しやすくできるからこそ、その考え方を導入する意味があると考え、みなさんも先の例で考えてみていただきたい。

「ニュルンベルク綱領」や「ベルモント・レポート」のような、臨床研究のガイドラインといえるようなものは、その後いくつも国際的に提唱されてきた。主なものでは、「ヘルシンキ宣言」がある。これは、「ニュルンベルク綱領」を基に世界医師会が作成したものであり、1964年の初版以来、繰り返しアップデートされている。わが国の「臨床研究に関する倫理指針」もこの内容を踏まえたものである。CIOMS(国際医科学評議会)の「WHOガイドライン」は、「ヘルシンキ宣言」を発展途上国に適用する目的で作成された。補償の明言や、弱い立場にある人々に対する配慮等が、「ヘルシンキ宣言」よりも踏み込んだ形で述べられている。日米欧の医薬品開発の統一規

準として制定された「ICH-GCP」にも、ヘルシンキ宣言などの考え方が取り入れられ、科学的で倫理的な臨床試験を行う際の必要な要件としてとらえることができる。

研究倫理の8要件

このようにさまざまなガイドラインが出る中で、米国NIHのエマニュエルらは、それらのガイドラインを統合した包括的かつ具体的な7つの倫理要件をJAMA誌に発表¹⁾し、のちにもう1要件加え²⁾、全体として8つの要件を提案した(図1)。これらの要件は研究が倫理的であることの必要十分条件ではないかもしれないが、研究が倫理的であるかどうかを検討するための有用な方法であるといえる。研究が倫理的であるかということ倫理という難しい概念としてではなく、臨床研究者の立場で理解できる具体的なチェックリストに落とし込んだ現場のマニュアルとしても使えるものなので、ここでぜひ紹介したい。2008年にエマニュエルが日本で講演した時には、8つを次の順番で検討することが重要であると紹介していたので、その順に述べたい。

まず、関係者間の協調、「共同パートナーシップ(collaborative partnership)」である(実際にはこれが最後に追加提案された)。臨床研究は研究成果の恩恵を受けられるコミュニティーを対象として行わなければならない、成果のみの搾取は行うべきでない、そのために関係者は協調して研究を行わなければならない、というものである。共同パートナーシップのない研究の例として、アフリカでエイズ薬の早期治験を行い、欧米で販売するといったことが挙げられる。早期治験を行ったアフリカでは、薬代が高くて使えないといった研究は共同パートナーシップがない。他の例として、欧米で抗がん剤の早期治験を行い、安全・有効とわかったら日本へ導入する、といった事例や、一部の富裕層しか使えないような高額な分子標的治療薬の開発もこれに当たるといえるだろう。このような例に対し、共同パートナーシップを実践するためには、早期段階からグローバル開発を行い、研究成果の恩恵を受けたい各国が参加することや、研究を実施したコミュニティーで研究成果をより多くの人が使えよう仕組みを考案することなどが考えられるであろう。たとえば、研究成果を健康保険のシステムに組み込めるよう工夫し、コミュニティーで共有するといったことや、研究の計画や実行、監督においてもコミュニティーが参加すること、対象となる患者集団やその代弁者、諮問委員会、倫理審査委員会、研究資金の支援者などの意見を聞くことなどが実践になると考えられる。臨床試験を行うことが、被験者が構成する社会において認められなければならないということである。

2つ目は、研究が倫理的であるためには、それが「社会的/科学的に価値がある(social or scientific value)」ものでなければならない、というものである。なんらかの意味で対象者の方に犠牲を強いるわけであるから、その結果ががん治療の進歩・未来の患者さんに貢献しなければ倫理的とはいえない。すでにわかっていることしか出てこない研究や研究の無秩序・無駄な重複、学会発表のため、業績作りのためだけの研究は倫理的でない。

3つ目は、「科学的妥当性(scientific validity)」

である。臨床研究は、科学的に正しいと認められた方法に基づいて実施する必要がある。デザインが適切でなかったり、被験者が少なく結論が出なかったり、データがきちんと管理されていないならば正しい結論を出すことができない。科学的に妥当な研究をするためには、医師が片手間に臨床研究を行うのではなく、生物統計家の支援体制の確保、データマネージャーによるデータ管理が必要である。もちろん、臨床医の負担を軽減したり、臨床医が臨床試験方法論を学ぶこと、教育の機会を提供することも重要である。

4つ目は「適正な被験者の選択(fair subject selection)」であるが、これは2つの側面から考える必要がある。被験者の選択は科学的な理由によって行われなければならない、①社会的弱者を臨床試験に不当に組み込んだり、②逆に科学的根拠もないのに不当に排除してはいけない。被験者にしやすいという理由で自分の会社の社員や学生を被験者に選ぶべきではないし、アジア人に対しても使用されるような薬剤の臨床試験には、アジア人を組み込まないとその対象者に対する結果を正しく評価できない。2つ目の側面である、根拠もないのに排除してはいけない(入れにくいから入れない)は見落としがちな観点である。

5つ目の要件は「適切なリスク/ベネフィットバランス(favorable risk-benefit balance)」である。つまり、被験者のリスクを最小化して、被験者の利益を最大化することが重要ということである。そのためには、リスクとベネフィットを試験開始前にしっかりと同定し、比較考量する必要がある。被験者のリスクと社会の利益のバランスを取ることも必要である。社会が大きな利益を受ける可能性があるからといって、被験者に大きな犠牲をもたらすことは正当化されない。試験の実施中には、被験者のリスクを最小化するために、治療変更規準をしっかりと遵守し、逸脱・違反をチェックすること、重篤な有害事象の情報を共有したり、適切な中間解析を行うことなどが必要となる。

6つ目の要件は「独立な第三者審査(independent review)」, 7つ目は「インフォームドコンセント

(informed consent)」である。これまでに述べた5要件を満たすように研究者は研究を計画し、それを評価できる第三者が確認し、被験者はそれを理解して臨床試験に参加してもらうことが必要である。研究者自身が上の要件が満たされているか判断することは、客観性がなく、自分たちに有利に判断しがちであり、利害関係(conflict of interest)のために正しく判断できない可能性もある。もちろん、被験者の自立性(autonomy)を尊重する必要がある。ただし、倫理といえは第三者審査とインフォームドコンセントというふうを考えられがちであるが、単に審査が独立であればよいわけではないし、被験者が同意すればどんな研究でもやってよいわけではない。上記5要件を満たしていることを第三者にきちんと審査してもらう必要があり、そのためには、倫理審査委員の教育も含めた倫理審査委員会の整備が必要といえる。第三者としては、被験者の利益を代表する市民委員といえるような人を倫理審査委員に含めるべきであり、有識者として弁護士や文学部の教授などを委員とすることが必ずしも適切とはいえない。インフォームドコンセントに関しては、適切な説明文書やインフォームドコンセントの取り方の教育・訓練をすること、CRC(臨床研究コーディネーター)の支援を受けることなどによって、よりきちんと実践できるといえる。

8つ目、最後の要件は、「候補者を含む被験者の尊重(respect for potential and enrolled subjects)」である。研究に参加した人のみならず、参加する可能性のある人は試験参加者のスクリーニングの段階から試験終了まで保護されていなければならない。それを実践するためには、十分な情報の提供、同意後の参加撤回の自由の保証、新たに判明した情報の提供、被験者保護の継続的な監視が必要となる。プロトコル治療中止後にも適切な医療を行うことや、研究結果のお知らせ、プライバシー/個人情報保護なども重要な要件となる。ただし、個人情報保護は重要な要件ではあるが、被験者保護の一側面にしかすぎないことも理解する必要がある。

これらの8要件は、先に述べた生命倫理の4原則をより具体的に記述したものといえる。図1

に示すように、研究者等の専門家、第三者(市民委員)、倫理審査委員会事務局など、それぞれある程度重点を置いてチェックすることも効率的であろう。自分が行う研究や、第三者として研究の倫理性を評価する場合にも、具体的なチェックリストとしてぜひ用いてみてほしい。

被験者保護のための体制

被験者保護は、個々人の努力だけで守ることは不可能であり、体制として守ることが必要となる。ここでは、米国と比較することによって、わが国の被験者保護体制を見てみたい。

米国における体制として、まず、連邦規則が挙げられる。各政府機関から研究費をもらって研究を行う場合には、コモンルールと呼ばれる各管理監督機関共通の45CFR46 subpart Aという連邦規則を守る必要がある。日本で言うところの治験は、食品医薬品局(FDA)の21CFR50, 56という連邦規則を守る必要があるが、これらの両方の連邦規則は内容的にほとんど同じであり、「治験」、臨床試験、臨床研究は基本的に同じルールに基づく。さらに体制として、臨床研究の管理監督機関であるOHRP(Office for Human Research Protections)が政府内に設置されている。OHRPでは、IRB登録、米国連邦保証制度、Federalwide Assuranceと呼ばれる施設承認、監査、教育などが行われている。これと連動する形で、スポンサーである研究費配分機関では、研究費配分の条件として施設のFWA承認を必要としている。FWAが取り消されると、施設への研究費差し止めとなる。FWA承認のためには、各施設で45CFR46を遵守すること、各施設でHRPP(Human Research Protection Program)と呼ばれる、教育、被験者保護の方針の策定、プロトコル審査、試験モニタリングなどの被験者保護プログラムを規定し、遵守すること、被験者保護担当者をおくこと、被験者保護担当者、施設長、IRB委員長はOHRPの教育プログラムを履修することが必要となる。米国では、「被験者は皆で責任を持って守ること(Human Subjects Protection is a Shared Responsibility)」とされ、ここでいう皆とは、HRPPにおけるキープレイヤーである施設長、IRB委員/委員長、IRB管理者、研究者、研

究スタッフ、他の内部スタッフ(試験薬係、放射線管理士、ほか)、外部(スポンサー、被験者、規制当局)ということになる。このように、連邦規則、監督機関、研究費配分機関が連動した体制をとることによって、関係者皆による被験者保護を実現するというシステムとなっている。

わが国では、規則として、治験に関しては薬事法による「医薬品の臨床試験の実施の基準に関する省令」(GCP)があり、その他の医学研究は、通知の位置づけの各種倫理指針による規制ということになる。臨床試験に関しては既述の厚生労働省「臨床研究に関する倫理指針」が対応する倫理指針となる。管理監督機関は、GCP/倫理指針ごとに異なり、厚生労働省、文部科学省、経済産業省などとなっている。スポンサー(研究費配分機関)としては、厚生科研費研究では「臨床試験登録」や「利益相反開示」を義務化しており、研究の質の担保のために生物統計学者や疫学者の参加を促しているが、どの程度採択や評価と関係しているかは必ずしも明確ではない。日本における「被験者保護」は「臨床研究に関する倫理指針」の前文によると、「…臨床研究においては、被験者の福利に対する配慮が科学的及び社会的利益よりも優先されなければならない。…被験者の人間の尊厳及び人権を守る…」と記載されている。キープレーヤーとして、GCPや臨床指針では、研究機関の長、研究者/研究代表者、倫理審査委員会、スポンサー(GCPのみ)の責務

が規定されている。

おわりに

臨床試験は、臨床指針の対象となるが、臨床指針は2009年4月に大幅改正された。改正の骨子は、倫理審査委員会の監視機能強化、有害事象報告、年次報告、終了報告、国としての管理体制、臨床試験登録、年次報告、重篤な有害事象報告、研究者の教育義務化、臨床研究倫理と研究に必要な知識、倫理審査委員会委員の教育推奨が研究機関の長の責務として盛り込まれたことなどである。わが国では、規則、監督機関、研究費配分機関が必ずしもうまく連動しているとはいえないかもしれないが、この臨床指針の改正により、徐々に体制が整ってきている段階にあるといえる。被験者保護は誰が守るということではなく、米国のように皆で守るもの、システムとして守るもの、という考え方が有用であると考えられる。

文 献

- 1) Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA 2000 ; 283 : 2701.
- 2) Emanuel EJ, Wendler D, Killen J, et al. What makes clinical research in developing countries ethical? The benchmarks of ethical research. J Infect Dis 2004 ; 189 : 930.

* * *

品質管理の観点からみた治験業務と臨床検査部の共生

国立がんセンター中央病院臨床検査部における治験業務の経験

小 関 満*1 塩 谷 香 奈*2 古 田 耕*3

Harmonization Procedures Directed toward Clinical Trials in Laboratory Medicine at the National Cancer Center Hospital

Mitsuru OZEKI*1, Kana SHIOYA*2 and Koh FURUTA*3

Recent advances in pharmacology and molecular sciences made it possible to develop drugs for patients with various maladies. Frustration has existed concerning the delayed provision of these drugs for routine practices in the clinical field. To correct this problem, the importance of clinical trials is increasing. Although there exists a strong demand for participation of clinical laboratories in clinical trials, an awkward attitude in clinical laboratories frustrates those performing clinical trials. We are attempting to correct this problem by introducing our experience with harmonization procedures directed toward clinical trials in laboratory medicine in general. First we described the current status of clinical trials in our hospital. Then we will show personnel in need for clinical trials. Finally we describe in detail our clinical trial procedures. We focus particularly on three aspects of participation in clinical trials: pre-analytical, analytical, and post-analytical. Additionally we describe the problems and perspectives in clinical trials by giving special reference to the clinical laboratories in general through discussion with various personnel and specialists. Our goal in the field of laboratory medicine is to benefit patients through the establishment of a harmony between clinical trials and clinical laboratories.

[Rinsho Byori 57: 827~833, 2009]

Corresponding author: Koh FURUTA, Department of Clinical Laboratory, National Cancer Center Hospital, Chuo-ku, Tokyo 104-0045, Japan. E-mail: kfuruta@ncc.go.jp

【Key Words】 clinical trial(治験), quality management(品質管理), standardization(標準化), pre-analytical(検査前手順), post-analytical(検査後手順)

I. 背 景

治験とは、何か。厚生労働省のホームページ¹⁾によれば、くすりの候補の開発の最終段階で、健康な人や患者さんの協力によって人での効果と安全性を

調べるいわゆる臨床試験のことであり、くすりの候補を用いて国の承認を得るための成績を集める臨床試験を特に治験と呼ぶという説明がなされている。一般人を対象としたホームページだけに非常にわかりやすくかつ注意深い説明がなされていると思う。

受付 2009 年 4 月 1 日・受理 2009 年 8 月 3 日

*1~3 国立がんセンター中央病院臨床検査部(〒104-0045 東京都中央区築地 5-1-1)

*2 独立行政法人国立病院機構宇都宮病院(〒329-1193 宇都宮市下岡本町 2160)