

Acknowledgement This study was supported by a science research grant (H21-005) from the Ministry of Health, Labour and Welfare (for research on infrastructure development for clinical trials).

Conflict of interest statement Masashi Ando—Research funding: Sanofi-Aventis, Wyeth, Novartis, and GlaxoSmithKline.

Chikako Shimizu—Honoraria: Daiichi Sankyo, Novartis, Chugai, Astra-Zeneca, Sanofi-Aventis, and GlaxoSmithKline.

Kenji Tamrua—Honoraria: Bristol Myers Squibb and Chugai; Research Grant: MSD.

Noriyuki Katsumata—Advisory Board: Astra-Zeneca; Honoraria: Sanofi-Aventis, Chugai, Daiichi Sankyo, Novartis, Kyowa-Kirin, and Ono Pharmaceutical; Consultant for protocol: Ono Pharmaceutical, Takeda Bio, and GlaxoSmithKline; Research funding: Eisai, GlaxoSmithKline, and Astra-Zeneca.

Yasuhiro Fujiwara—Honoraria: Taiho, Sanofi-Aventis, Eli Lilly, and Chugai; Research Funding: Pfizer, GlaxoSmithKline, Chugai, Eisai, Daiichi Sankyo, Taiho, Nihon Kayaku, Amgen, Novartis, Takeda, and Astra-Zeneca.

The other authors declare no conflicts of interest.

References

1. Ministry of Health, Labour and Welfare (2005) The expert review committee for unapproved drugs. Available from URL: <http://www.mhlw.go.jp/shingi/2005/01/s0124-9.html> [accessed 3 July 2010].
2. Ministry of Health, Labour and Welfare (2006) The expert review committee in order to rapidly supply effective and safe medicines. Available from URL: <http://www.mhlw.go.jp/shingi/2006/10/s1030-8.html> [accessed 3 July 2010].
3. Fukuhara H (2006) Period between world first launch and country launch. Office of Pharmaceutical Research. Research Paper No. 31, Office of Pharmaceutical Research, Tokyo. (in Japanese)
4. Tsuji K, Tsutani K (2008) Approval of new biopharmaceuticals 1999–2006: comparison of the US, EU and Japan situations. *Eur J Pharm Biopharm* 68:496–502
5. Farrell AT, Papadouli I, Hori A et al (2006) The advisory process for anticancer drug regulation: a global perspective. *Ann Oncol* 17:889–896
6. Pharmaceuticals and Medical Devices Agency (2010) Public assessment report. Available from URL: <http://www.info.pmda.go.jp/approvalSrch/PharmacySrchInit?> [accessed 3 July 2010].
7. Food and Drug Administration (2010) Drugs@FDA. Available from URL: <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm> [accessed 3 July 2010].
8. European Medicines Agency (2010) EPARs for authorized medicinal products for human use. Available from URL: <http://www.ema.europa.eu/htms/human/epar/> [accessed 3 July 2010].
9. Medicines and Healthcare Products Regulatory Agency (2010) Public assessment reports for medicines. Available from URL: <http://www.mhra.gov.uk/Onlineservices/Medicines/PublicAssessmentReports/A-Zlisting/index.htm> [accessed 3 July 2010].
10. Ministry of Health, Labour and Welfare (2006) Guidelines on methods of clinical evaluation of oncology drugs. Available from URL: <http://www.who.mhlw.go.jp/hourei/doc/tsuchi/171101-b.pdf> [accessed 3 July 2010].
11. Ministry of Health, Labour and Welfare (2007) Basic principles on global clinical trials. Available from URL: <http://www.pmda.go.jp/operations/notice/2007/file/0928010-e.pdf> [accessed 3 July 2010].
12. Ministry of Health, Labour and Welfare (2004) The expert review committee for combination chemotherapy. Available from URL: <http://www.mhlw.go.jp/shingi/2004/05/s0521-5.html> [accessed 3 July 2010].
13. Ministry of Health, Labour and Welfare (2002) Pharmaceuticals affairs law. Available from URL: <http://www.mhlw.go.jp/topics/2002/09/tp0910-2.html> [accessed 3 July 2010].
14. Ministry of Health, Labour and Welfare (2007) The expert review committee for discussing what registration should be. Available from URL: <http://www.mhlw.go.jp/shingi/2007/09/s0919-8.html> [accessed 3 July 2010].
15. Hayashi Y (2007) New 5 yearly clinical trial action plan. *Pharm Regul Sci* 38:658–663 (in Japanese)
16. Pharmaceuticals and Medical Devices Agency, Japan (2009) Profile of services fiscal year 2009. Available from URL: http://www.pmda.go.jp/english/about/pdf/profile_of_services.pdf [accessed 3 July 2010].
17. Pharmaceuticals and Medical Devices Agency, Japan (2008) Points to be considered by the review staff Involved in the evaluation process of new Drug. available from URL: <http://www.pmda.go.jp/english/service/pdf/points.pdf> [accessed 3 July 2010].
18. Fujiwara Y (2003) Approval system of oncology drug in Japan and surrounding problem. *Ka-ryou News* 11, Japanese Gynecologic Oncology Group, Tokyo.
19. Tillman K, Burton B, Jacques LB et al (2009) Compendia and anticancer therapy under Medicare. *Ann Intern Med* 150:348–350
20. Ministry of Health, Labour and Welfare (2010) The expert review committee for unapproved drug for unmet medical needs. Available from URL: <http://www.mhlw.go.jp/shingi/2010/02/s0208-9.html> [accessed 20 September 2010].
21. National Cancer Center. Trends in cancer statistics. <http://ganjoho.jp/data/public/statistics/backnumber/1isaa000000068m-att/fig18.pdf> [accessed 20 September 2010].

産科と婦人科 別刷

Vol. 77 No. 5 (2010年5月1日発行)

発 行 所 株 式 会 社 診 断 と 治 療 社

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-

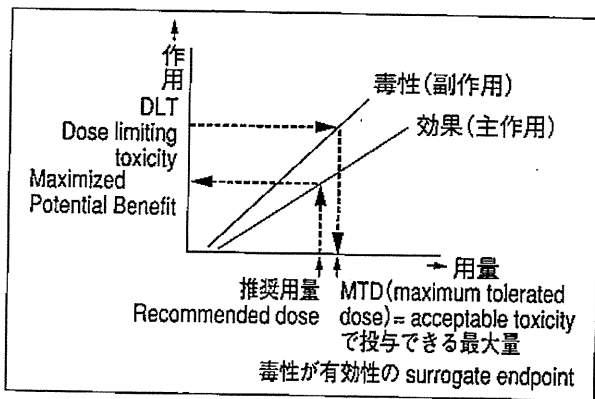


図1 細胞障害性抗がん剤 p1 デザインの rationale “More is better” with acceptable toxicity

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、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

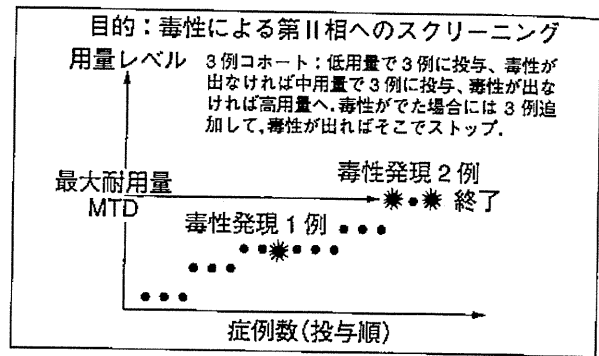


図2 第I相臨床試験の典型的なデザイン [By courtesy of Dr. Ishizuka]

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

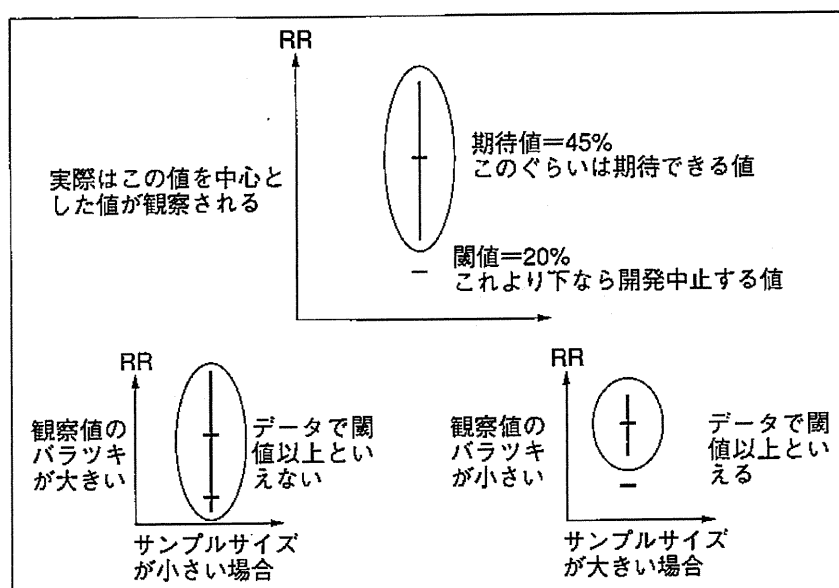


図3 閾値、期待値を用いたサンプルサイズ設計

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ふ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%

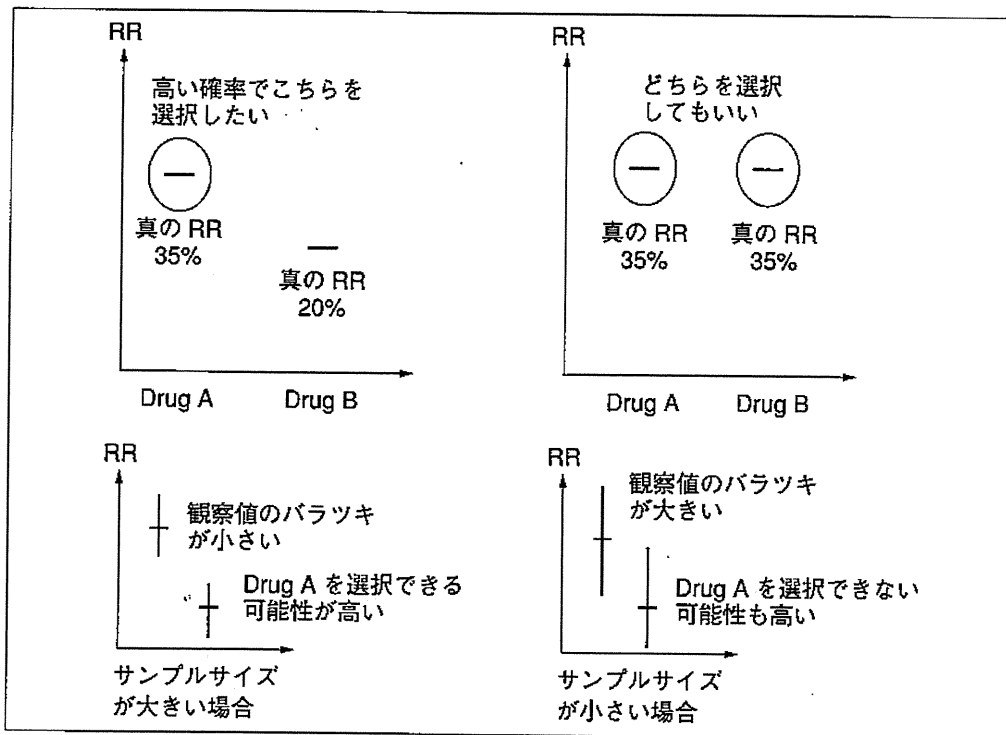


図4 randomized Phase2 デザインの考え方

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り、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いことになる. 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

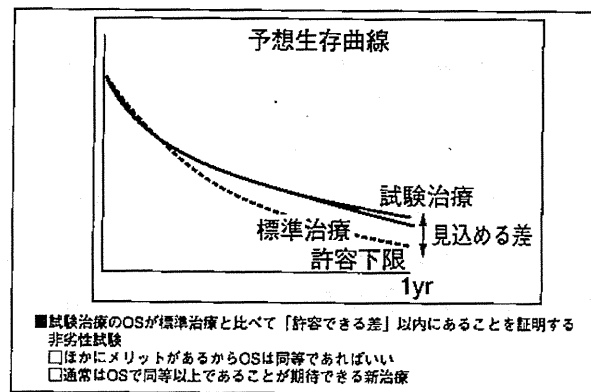


図5 非劣性試験とは

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り、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>

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際だけでなく、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。厚労科研費による 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

免疫療法の臨床試験 方法論、バイオマーカー

吉村 健一 (京都大学医学部附属病院 探索医療センター 検証部)

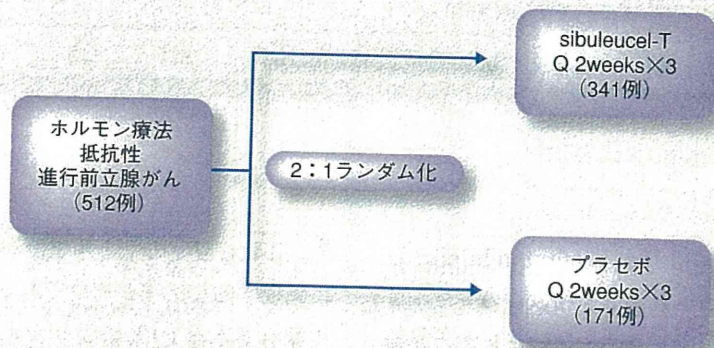
山本精一郎 (国立がん研究センター がん対策情報センター がん情報・統計部 がん統計解析室)

Point

- 免疫療法を開発するにあたって、新しい標準治療として広く普及させることを目的とする場合、臨床開発の最終段階として、標準治療との直接比較を行う第Ⅲ相試験の実施が求められる。
- 早期試験(第Ⅰ相試験や第Ⅱ相試験)は、通常は大規模となる第Ⅲ相試験で評価すべき有望な治療をスクリーニングすることを主たる目的として実施する。
- 早期試験では相対的に、検証性よりも効率性に重きを置くべきであり、免疫療法の特徴を十分に考慮することによって効率性を高められる可能性がある。
- 開発しようとする免疫療法について、現実的な治療域として想定される用量において、治療開発直後に重篤な、または生命に危険を及ぼすような毒性を生じさせる可能性がきわめて低いならば、第Ⅰ相試験において、従来のように低用量からの慎重な漸増を行う必要性はないかもしれない。
- 重篤な、または生命に危険を及ぼすような毒性を生じさせる可能性が一定以上存在することが否定できないならば、倫理的な観点からも、細胞毒性治療薬を開発する際に用いる従来型の漸増計画を伴う第Ⅰ相試験を実施すべきと考えられる。
- 第Ⅱ相試験は第Ⅲ相試験に先立って実施するものであり、免疫療法の開発においては、従来型のがん第Ⅱ相試験デザインを基本として、免疫療法の特徴をどのように考慮すればより効率的な試験となるかを考えるのが適切であると考えられる。
- がん患者を対象とした治療法を開発し、最終的にその有用性を検証することを考えると、免疫療法を開発するにあたって、これまでに発展してきたがん臨床試験の方法論の多くが引き続いて有用であると考えられる。

これまでの歴史的な背景により、がん臨床試験の方法論は、主に細胞毒性治療薬、近年では分子標的治療薬

をその対象とし、これらに対する効率性と検証性の双方の観点を考慮して体系化と最適化がなされてきたものであ



主要エンドポイント：overall survival (OS) HR=0.759、p=0.032
 副次エンドポイント：time to objective disease progression (ODP)

図1 IMPACT第Ⅲ相試験のデザイン

プラセボに比した優越性を検証することを目的として、大規模な検証的試験として計画された。

る。しかしながら生物学的メカニズムを根拠として安全性に問題がなく、さらには有効性があると主張し、全開発戦略を通じて適切にデザインされた臨床試験などを行わなくても、免疫療法というカテゴリーで一括りにしたうえで、日常診療として実施してよい、と判断してしまえば多大なリスクが生じうる。この意味において、実験性のある免疫療法がわが国の一部において必ずしも科学的評価を伴わずに用いられてしまっている現状は、倫理的観点・科学的観点の双方から憂慮すべき状況であるとも考えられる。真に有効な免疫療法が存在するのであれば、曖昧な状態で一定の科学的評価も得られない状態で実施し続けるのではなく、適切にデザインされた臨床試験により効率的にその有用性をできる限り早期に検証することこそ求められよう。以上より、開発の効率性と検証性を考えると、これまでに発展してきた方法論の多くは、免疫療法の開発および評価に

においても、個々の状況に応じて以下にも述べるような若干の修正を必要とするかもしれないものの、引き続いて有用なものであると考える。

標準治療を決定する 第Ⅲ相試験

免疫療法を開発するにあたって、現在の標準治療と比べてよいことを示すことによって新しい標準治療として広く普及させることを目的とする場合、その最終段階として標準治療になりうるかを評価するためには第Ⅲ相試験の実施が求められるということについては、まず異論がないところであろう。詳細な解説は成書(Greenら、2002¹⁾；Piantadosi、2005²⁾)に委ねて本稿では省略するものの、第Ⅲ相試験はランダム化を伴い、全生存期間(overall survival；OS)などの臨床的な有効性を直接的に反映するエンドポイント(真のエンドポイントtrue endpoint

という)において標準治療との直接比較を行うものであり、このような試験は検証的試験ともよぶ。ホルモン療法抵抗性進行前立腺がん症例を対象として、前立腺がん特異的抗原PAPを用いて活性化した免疫細胞製剤 sipuleucel-Tの効果を検証するために実施された第Ⅲ相試験(IMPACT試験)は、プラセボをコントロール群としたランダム化比較試験として実施され、主要エンドポイントである全生存期間における有意な延長効果を報告している(図1)。規制当局である米国食品医薬品局(Food and Drug Administration；FDA)は、本試験より前に実施された2つの比較的小規模な比較試験の結果に基づく承認申請については2007年に却下していたものの、全生存期間を主要エンドポイントとする大規模な比較試験として実施された本試験の結果を評価した2010年には初のがん治療ワクチンの承認を決定するに至っている。

表1 免疫学的反応率をエンドポイントとして最小有効量(minimum effective/active dose)を探索する第I相試験デザインの一例

期待する免疫学的反応率 (または必要とする免疫学的反応率)	20%	25%	33%	40%	50%
用量レベルあたりの必要症例数	11例	9例	6例	5例	4例

一方、第III相試験の前段階として、通常は大規模となる第III相試験で評価すべき有望な治療をスクリーニングすることを主たる目的とする早期試験(第I相試験や第II相試験)では相対的に検証性よりも効率性に重きを置くべきである。すなわち、免疫療法の特徴を十分に考慮したうえで、効率性の観点から、これまでに体系化されたがん臨床試験の方法をどのように用いるべきかについて再整理する必要性が考えられる。

第I相試験

免疫療法の特徴として、まず現実的な治療域として想定される用量において、治療開始直後に重篤な、または生命に危険を及ぼすような毒性を生じさせる可能性がきわめて低いことがあげられる。これが正しければ、細胞毒性治療薬の開発とは異なり、第I相試験において低用量を開始用量として用量制限毒性発現まで慎重に漸増する必要はないかもしれない。実際上は、実施可能性などによって用量が制限されることも多い。また、前臨床試験で得られた知見に基づいて用量を最適化せざるを得ないような状況も想定される(Salgallerら, 1996³); Cormierら,

1997⁴)。これより、first-in-humanの設定で行われる第I相試験の主たる目的が、単一用量における安全性の評価や細胞療法などにおける手技自体の実施可能性の評価となることも十分に考えられる。

ホルモン療法抵抗性前立腺がんを対象として実施されたPROSTVAC-VFワクチンの第I相試験として、単一用量における安全性および実施可能性を評価することを主たる目的としたもの(DiPaolaら, 2006⁵)、およびそれに加えて生体内分布ならびに動態を評価することを目的としたもの(Arlenら, 2007⁶)の2試験が後述する第II相試験(Kantoffら, 2010⁷)に先立って報告されている。

実施可能性や前臨床試験データなどを根拠に用量を定めるのではなく、第I相試験の設定において最小有効量(minimum effective/active dose)を探索し、これに基づいて推奨用量(recommended dose; RD)を決定することも可能であるかもしれない。例えば、用量レベルごとに症例単位で免疫学的反応の有無を評価し、これにより免疫学的反応率を求める状況を考える。RDにおいて期待する免疫学的反応率、または必要とする免疫学的反応率は状況に応じて決めるべきであるが、この値として20~50%を想定す

る。これら設定の下、用量レベルごとに免疫学的反応率がゼロであることを片側有意水準10%により否定できるような精度を担保することを前提に症例数設計したものを一例として表1に示す。例えば、免疫学的反応率として25%以上を期待した場合には用量レベルあたりの必要症例数が6例以下となり、がん第I相試験で慣習的な方法としてしばしば用いられてきた3例コホート法における3~6例に等しいか、あるいは若干多くなる程度の症例数となることが分かる。

用量があらかじめ定まってい、かつ第I相試験において安全性に関する問題が生じる可能性がきわめて低いことが事前に予想されるのであれば、後述する第II相試験の評価も同時に行えるような第I/II相試験としてシームレスに実施することも可能となる。これにより、開発の効率化を期待できる。

一方で、この節の最初に述べた前提が必ずしも成立せず、重篤な、または生命に危険を及ぼすような毒性を生じる可能性が一定以上存在する場合には、倫理的な観点からも細胞毒性治療薬を開発する際と同様のデザインにより実施するほうが適切となる。あるいは、免疫療法単独ではなく、他の薬剤や治療法と組み合わせて併用療法とし

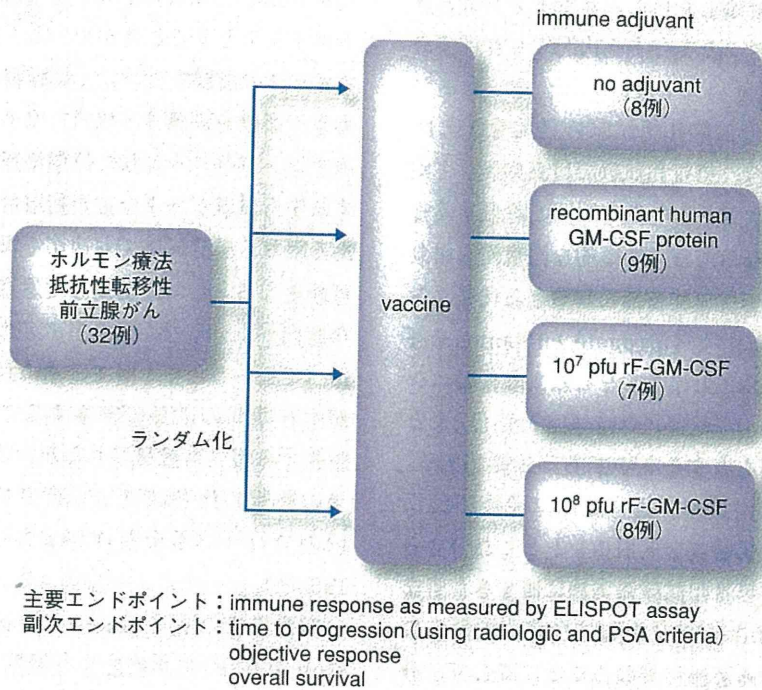


図2 PROSTVAC-VF ワクチンの免疫学的反応を評価するために実施されたランダム化第Ⅱ相試験

免疫学的反応の評価を主たる目的として、4つの治療法へのランダム化を伴う試験として計画された。

で開発する場合にも同様に考えるべきであるかもしれない。

第Ⅱ相試験

がん第Ⅱ相試験の主たる目的は、次相となる①第Ⅲ相試験における標準治療との直接対決に値する有望な治療法であるかを主たる目的として有効性において評価すること、②第Ⅲ相試験で評価すべき治療法を最適化すること、③第Ⅲ相試験を実施する集団(治療のターゲットとなる集団)を決めることである。便宜的に試験に登録可能な症例数が少ないなどのみを根拠として、小

規模な第Ⅱ相試験を実施することには倫理的な観点から問題が生じうる。歴史的背景からも強く示唆されるが、小規模な第Ⅱ相試験のデータを積み重ねていったとしても、それらから最終的に検証的な結果を産み出すことは困難を極める。

前述の通り、免疫療法であっても、治療開発の過程において最終的に実施するものが第Ⅲ相試験であることには変わりがないため、第Ⅱ相試験の目的とする①～③は免疫療法の開発においても共通するものである。これより、従来型のがん第Ⅱ相試験デザインを基本として、免疫療法の特徴をどのように考慮すればより効率的な第Ⅱ相試験

となるかを考えることこそ、ここでの議論の中心となる。

第Ⅱ相試験のエンドポイント

第Ⅱ相試験の目的が前述の②にあり、複数の処置の比較などによって免疫療法の最適化を行いたい場合、免疫学的反応を評価することは自然であるのみではなく効率的とも考えられ、これにより有用な情報が得られることになるかもしれない。Gulleyら(2010)⁸⁾は、32例のホルモン療法抵抗性転移性前立腺がんを対象とし、主要エンドポイントを免疫学的反応としたうえで

4つの治療法へのランダム化を伴う第Ⅱ相試験を実施している(図2)。

また目的を③とする場合にも、評価する免疫療法が特定のサブ集団においてのみ免疫学的反応をもたらすことが生物学的にも説明づけられる状況であれば、免疫学的反応を評価することが自然であるかもしれない。例えば特定のサブグループに属する対象者においてのみ免疫学的反応が生じることが生物学的に期待されるものの、特定方法が未知である状況では、免疫学的反応等を第Ⅱ相試験のエンドポイントとして、サブグループの特定に用いるバイオマーカーの開発を行うことは十分に考えられる。ただし、免疫学的反応などのエンドポイントにおいて個人間、あるいは測定施設間で大きなバラツキが存在する場合、ヒストリカルコントロールや他の研究との比較はしばしば困難となる。測定値の個体間差や施設間差が大きいような状況において、Rubinsteinら(2010)⁹⁾は、“第0相”試験を例としているものの、生物学的効果を評価するためのエンドポイントとして自己をコントロールとして個体レベルでの有意な反応を定義することを

提案している。免疫学的反応などの試験で評価したいエンドポイントに大きなバラツキを有する状況でも有用であろう。

一方、第Ⅱ相試験の目的を①とする場合、第Ⅲ相試験では真のエンドポイントである全生存期間を評価することになるため、効率的な観点からより早期に評価することができる代替エンドポイント(surrogate endpoint)を評価することになる。これより代替エンドポイントは、真のエンドポイントと相関を有するものである必要がある。免疫療法を評価する際に免疫学的反応を代替エンドポイントとしたいならば、第Ⅲ相試験の実施に値する有望な治療であるかの判断を免疫学的反応における評価に委ねることに関して十分な妥当性が求められる。例えば、免疫学的反応が真のエンドポイントである全生存期間との相関を有さないならば、この免疫学的反応をもって有望な治療であると判断してしまうことは、目的①に照らし合わせると必ずしも適切でない。

測定可能病変を有する集団を対象とできる状況では、従来型のがん第Ⅱ

相試験と同様に、積極的に腫瘍縮小効果(tumor shrinkage)や奏効をエンドポイントとすることができる。選択バイアスの影響を受けにくい腫瘍縮小効果や奏効を評価する場合、ヒストリカルコントロールとして標準治療に関する先行試験データなどが利用可能である限り、単群試験でも適切な解釈が可能となる。なお、奏効した患者の生存期間が奏効しなかった患者のそれに比べて長いことをもって、その治療法が生存期間の延長に寄与することを主張する報告も散見されるが、この結果の解釈の仕方は決して適切ではないので行うべきでない(Anderson, 1983¹⁰⁾)。

増悪までの期間(time-to-progression ; TTP)や無増悪生存期間(progression-free survival ; PFS)などのいわゆる無イベント生存期間は、その程度は疾患臓器にもよるものの、真のエンドポイントである全生存期間とのより直接的な関連を有するものとして位置付けられる。これらは、その目的を①とする第Ⅱ相エンドポイントとして適切な状況も多いかもしれない。ただし、TTPやPFSは選択バイアスに対して一般に脆弱である状況が多く、試験に実際に登録された症例の属性分布に大きく左右されうる。選択バイアスの影響を減じるためには、対照群を伴うランダム化試験を実施する必要がある。ランダム化試験を実施しようとする場合、一般に必要な症例数は大きく増大する。

*1… α エラー

真には帰無仮説が正しいという状況の下で仮説検定の結果として帰無仮説を否定してしまう誤り。試験デザインに沿った適切な統計解析法を用いる限り、 α エラーを犯す確率は有意水準以下に抑えられる。その定義から、帰無仮説が真には誤っている状況(あるいは対立仮説が正しい状況)ではこのエラーは生じない。また同様にその定義から、事後的に帰無仮説を否定できなかった状況でも生じない。

*2… β エラー

真には帰無仮説が誤っている(対立仮説が正しい)という状況の下で仮説検定の結果として帰無仮説を否定できない(対立仮説を採択できない)誤り。また同一の状況の下で仮説検定の結果として帰無仮説を正しく否定できる確率(1から β エラーの確率を減じたもの)を検出力とよぶ。その定義から、真には帰無仮説が正しい状況ではこのエラーは生じない。また同様にその定義から、事後的に帰無仮説を否定できた状況でも生じない。

単群第Ⅱ相 試験デザイン

単群試験のデザインは、エンドポイントが何らかの反応の有無などを表す二値変数(二値エンドポイント)である場合、通常用いられるものをそのまま利用できる。二値エンドポイントを用いた単群第Ⅱ試験デザインの詳細な解説は前述した成書^{1,2)}に委ねるが、例えば、ある二値エンドポイントに関して当該試験で否定したい割合(閾値という)を5%、試験治療に期待する割合、または有望と判断するにあたって必要とする割合(期待値という)を20%とそれぞれ設定する状況を考える。この設定の下で、従来型の第Ⅱ相試験でよく用いられるSimon(1989)¹¹⁾の最適2段階デザインにより、片側有意水準10%、検出力90%として評価を行う場合、まず第1段階で12例を登録し、その12例のうち1例でも反応が観察されれば第2段階としてさらに25例を登録し、最終的に第1段階の症例も含めた全37例のうち4例以上の反応が観察された場合にその試験治療を有望と判断することになる。第1段階で1例の反応も観察されない場合や、第2段階に進んだとしても最終的に全37例のうち3例以下しか反応が観察されなかった場合には試験治療が有望でないと判断することになる。ここでは一例を示したが、特定の閾値および期待値にあわせて適切な必要症例数および判断基準をそれぞれ求めることができる。

このデザインの特徴は2段階で実施することであり、試験治療が有望なも

のでないならば早期に試験を無効中止することができる。例えば、この例において、残念ながら試験治療における真の割合が閾値以下となる場合、54%以上の確率により第1段階で試験を早期無効中止することができる。無意味な試験を早期に中止することは倫理的観点から好ましい。エンドポイントが二値以外の場合にも同様に考えることができる。免疫学的反応などのエンドポイントが連続的な値(連続変数)として測定されている場合には、二値化によって情報損失が生じる。このため、安易に二値化すべきではないかもしれない。一般に、連続変数は二値化するよりも連続量変数のまま評価するほうが高い精度を期待できる(Karrisonら, 2007¹²⁾)。対象とする患者集団や治療法の性質などの状況にも大きく依存するが、免疫学的分析試料からもたらされるものが連続変数である場合にも同様に効率的な試験デザインを採用できる可能性が考えられる。

ランダム化第Ⅱ相 試験デザイン

第Ⅱ相試験でランダム化試験デザ

インを採用する典型的な状況として、(A)新治療の候補が同時に複数存在する状況でそのなかで開発の優先順位をつけたい場合(選択の問題)、(B)複数の用量や投与スケジュールの候補が存在する状況でそのなかで最適化を行いたい場合(最適化の問題)、(C)単群試験の結果と比較可能となる適切なヒストリカルコントロールが存在しない場合(コントロールの問題)があげられる。

それぞれの目的に応じて適切な試験デザインが用いられる。(A)および(B)の場合、標準治療を含まず、複数の候補治療法にランダム化するデザインが用いられる。特に(A)の場合には、優先順位をつけるという目的に対して適切となる選択デザイン(selection design)を用いることができる。(B)の場合、主たる目的に応じて、選択デザイン、用量反応性を評価するデザインなどを用いることができる。(C)の場合には、標準治療を含めてランダム化する同時対照デザインが用いられることになる。特に先にあげたTTPやPFSなどのいわゆる無イベント生存期間を評価する状況、あるいはそもそも代替エンドポイントとして適切なものが存在せずにスクリーニングが目的の第Ⅱ相試

*3…p値

実際に観察された結果の、帰無仮説が正しいという状況の下での不自然さを表したものであり、0から1の範囲の値をとる。値が0に近づくほどより不自然であるように構成されている。仮説検定の枠組みではこの値が事前に設定した有意水準を下回るか否かによって帰無仮説を否定するか否かを判断する。帰無仮説を否定した際に生じる最小の α エラー(smallest α error)。

*4…有意水準

試験結果から算出したp値がこれを下回った場合に帰無仮説を否定するために事前設定された閾値。実験デザインに沿った適切な統計解析法を用いて検定を行う限り、 α エラーを犯す確率はこの有意水準以下に抑えられることが保障される。医薬分野では多くの領域・臨床試験において、片側有意水準5%或いは両側有意水準5%(片側有意水準2.5%に対応する)が慣習的なコンセンサスとして用いられることが多い。

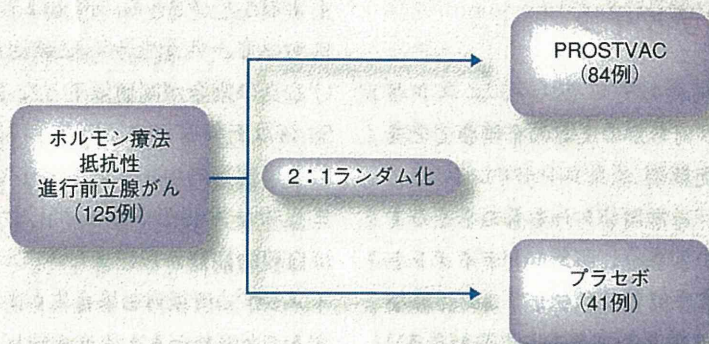


図3 PROSTVAC-VFワクチンの同時対照ランダム化第Ⅱ相試験

第Ⅲ相試験における標準治療との直接対決に値する有望な治療法であるかを主たる目的とし、有効性の代替エンドポイントである無増悪生存期間を主要エンドポイントとして計画された。

主要エンドポイント：progression-free survival HR=0.884、p=0.60
副次エンドポイント：overall survival HR=0.56、p=0.0061

験においても全生存期間を評価せざるをえないような状況では、前述の通り、一般に選択バイアスの問題があることから(C)に該当する。ホルモン療法抵抗性進行前立腺がん症例を対象として実施されたPROSTVAC-VFワクチンの第Ⅱ相試験は、主要エンドポイントをPFSとし、プラセボをコントロール群としたランダム化比較試験デザインを採用している(図3)。本試験では、副次エンドポイントとしていたOSでは比較的大きな差が観察されたものの、残念ながら主要エンドポイントであるPFSでは差が観察されなかった(Kantoffら, 2010⁷⁾)。

(A)~(C)のいずれの状況のデザインであっても、従来の臨床試験でも用いられてきたものと同様であり、エンドポイントと評価したい仮説に応じて、従来の症例数設計法をそのまま用いる

ことができる。群間比較を行うのであれば、効率性及び検出力の観点から、症例数が均等(例えば1:1)になるようにランダム化するのが好ましい。不均等(例えば2:1)となるようなランダム化を行う際にはその必要性について十分な検討を要する。

第Ⅱ相試験でランダム化試験デザインを採用する場合のデメリットとして、第Ⅲ相試験とは異なり検証的試験では決してないにもかかわらず、検証的試験に準ずる結果として過大に解釈されてしまうことがよく指摘される。試験規模が小さく、検証的試験でないということは、本当は治療効果がそれほどないような場合にも、偶然の影響により大きな差が観察されてしまう可能性が高いということを意味する。ランダム化第Ⅲ相試験に先立つスクリーニングとして計画・実施された第

Ⅱ相試験の結果を過信して、検証的な第Ⅲ相試験の計画・実施にブレーキがかかってしまうのであれば、結果的には根根灌枝となる。第Ⅱ相試験で標準治療を同時対照とする2群試験を計画する際には、このような実際上生じうるデメリットに対しても十分に配慮する必要がある。

おわりに

がん患者を対象とした治療法を効率的に開発し、その有用性を検証することを考えると、免疫療法の開発および評価においても、これまでに発展してきたがん臨床試験の方法論の多くが引き続き有用であると考えられる。これらの方法論には、以上では詳細を述べなかったが、近年注目され、徐々に

整理されてきている個別化医療の開発なども含まれる。

免疫療法に対して有用性を期待するならば、適切にデザインされた臨床

試験により効率的にその有用性をできる限り早期に検証することこそが求められる。

文献

- 1) Green S, Benedetti J, Crowley J. *Clinical Trials in Oncology*, 2nd ed. CRC Press; 2002. [福田治彦, 新美三由紀, 石塚直樹, 訳. 米国SWOGに学ぶ がん臨床試験の実践-臨床医と統計家の協調をめざして. 東京: 医学書院; 2004]
- 2) Piantadosi S. *Clinical Trials -A Methodologic Perspective*, 2nd ed. Wiley; 2005.
- 3) Salgaller ML, Marincola F, Cormier, JN, Rosenberg, SA. Immunization against epitopes in the human melanoma antigen gp100 following patient immunization with synthetic peptides. *Cancer Research* 1996; 56: 4749-57.
- 4) Cormier JN, Salgaller ML, Prevetie T, Barracchini KC, Rivoltini L, Restifo NP, et al. Enhancement of cellular immunity in melanoma patients immunized with a peptide form MART-1/Melan A. *The Cancer Journal from Scientific American* 1997; 3: 37-44.
- 5) DiPaola RS, Plante M, Kaufman H, Petrylak DP, Israeli R, Lattime E, et al. A phase I trial of pox PSA vaccines (PROSTVAC-VF) with B7-1, ICAM-1, and LFA-3 co-stimulatory molecules (TRICOM) in patients with prostate cancer. *J Transl Med* 2006; 4:1.
- 6) Arlen PM, Skarupa L, Pazdur M, Seetharam M, Tsang KY, Grosenbach DW, et al. Clinical safety of a viral vector based prostate cancer vaccine strategy. *J Urol* 2007; 178: 1515-20.
- 7) Kantoff PW, Schuetz TJ, Blumenstein BA, Glode LM, Bilhartz DL, Wyand M, et al. Overall survival analysis of a phase II randomized controlled trial of a Poxviral-based PSA-targeted immunotherapy in metastatic castration-resistant prostate cancer. *J Clin Oncol*; 2010 28: 1099-105.
- 8) Gulley JL, Arlen PM, Madan RA, Tsang KY, Pazdur MP, Skarupa L, et al. Immunologic and prognostic factors associated with overall survival employing a poxviral-based PSA vaccine in metastatic castrate-resistant prostate cancer. *Cancer Immunol Immunother* 2010; 59: 663-74.
- 9) Rubinstein LV, Steinberg SM, Kummar S, Kinders R, Parchment RE, Murgu AJ, et al. The statistics of phase 0 trials. *Stat Med* 2010; 29: 1072-6.
- 10) Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *Journal of Clinical Oncology* 1983; 1: 710.
- 11) Simon R. Optimal Two-Stage Designs for Phase II Clinical Trials. *Controlled Clinical Trials* 1989; 10: 1-10.
- 12) Karrison TG, Maitland ML, Stadler WM, Ratain MJ. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non-small-cell lung cancer. *J Natl Cancer Inst* 2007; 99: 1455-61.

①改正された臨床研究に関する倫理指針の基本理念

国立がんセンター中央病院
臨床試験・治療開発部

山下 紀子 Noriko YAMASHITA
藤原 康弘 Yasuhiro FUJIWARA

はじめに

より良い治療法、診断方法、予防方法を開発するため、また、病因を解明したり患者の生活の質を向上するために、日々様々な医学研究が行われている。医学研究に関する本邦の研究倫理指針や治験に関する規制が、近年、相次いで改正された。その1つが平成21年4月1日に施行された「臨床研究に関する倫理指針」(以下、臨床指針)¹⁾である。その改正内容は、研究者のみならず、研究機関の長や倫理審査委員会に対しても及ぼす影響が大きく、各研究機関の長は研究実施体制の大幅見直しに追われた1年間であったのではないだろうか。

体制整備を行うためには、改正された臨床指針の変更点の表面的な記述のみならず、その記述の根拠、本質を理解することが必要となる。よって、本稿では改正された臨床指針の基本理念を概説する。

なお、医学研究は臨床の場で行われる臨床試験や観察研究に限らず、研究所等で行われる基礎研究やフィールドで行われる疫学研究もあり、研究の種類は多岐にわたる。本稿では、世界医師会のヘルシンキ宣言²⁾が適用される「個人を特定できるヒト由来の試料およびデータの研究を含む、人間を対象とする医学研究」を臨床研究と同義語として扱うこととし、以下、臨床研究と呼ぶ。

近年の臨床研究に関する規制・倫理指針の改訂

ここ数年の間に改訂がなされた臨床研究関係の倫理指針、規制は以下の通りである。また、これら倫理規範に基づいているヘルシンキ宣言も平成20年10月に改訂され、日本語訳が同年12月に公開されている。

- ・疫学指針に関する倫理指針(以下、疫学指針)³⁾
(平成19年文部科学省・厚生労働省告示第1号)
(平成19年8月16日告示, 平成19年11月1日施行)
- ・医薬品の臨床試験の実施の基準に関する省令の一部を改正する省令(以下、改正GCP省令)⁴⁾
(平成20年厚生労働省令第24号)

(平成20年2月29日公布, 平成20年4月1日施行(一部、平成21年4月1日施行))

- ・臨床研究に関する倫理指針
(平成20年厚生労働省告示第415号)
(平成20年7月31日告示, 平成21年4月1日施行)
- ・ヒトゲノム・遺伝子解析研究に関する倫理指針(以下、三省指針)⁵⁾
(平成16年文部科学省・厚生労働省・経済産業省告示第1号)
(平成20年12月1日一部改正)

これらのうち、法令の位置づけであるものはGCPのみであり、根拠法として薬事法がある。一方、臨床指針、疫学指針、三省指針は根拠法をもたない「ガイドライン」である。そのため、いずれも「ヘルシンキ宣言」に基づいた倫理規範であり、研究の倫理性と科学性を確保することにより被験者を保護するための文書という位置づけは共通しているにもかかわらず、これまで内容と拘束力が大きく異なっていた。しかし、今回のGCPと臨床指針の改正により、その差は縮まったといえる。根拠法をもつか否かの違いはあるが、倫理規範を遵守しないことにより問われる道義的な責任は同じであろう。どのような研究であっても、人を対象として研究を行う以上、研究者は被験者保護に留意し、適用となる倫理規範を遵守することが求められる。

臨床指針改正のポイント

臨床指針の主な改正内容としては、以下の7項目が挙げられる。

- ① 倫理審査委員会関係
- ② 健康被害に対する補償について
- ③ 研究者等の教育の機会の確保について
- ④ 臨床研究計画の事前登録について
- ⑤ 臨床研究の適切な実施確保について
- ⑥ 観察研究、試料等の保存およびほかの機関等の試料等の利用について
- ⑦ その他、用語等について

このうち、②～④そして⑤に含まれる重篤な有害事象に関する責務の強化、実施状況報告や終了報告の義務化

等、研究者、臨床研究機関の長、倫理審査委員会の責務が強化され、研究者や研究機関の長の負担がGCP並みに増えると思われる事項に注目が集まっている。しかし、あまり注目されていないが、研究の種類によっては迅速審査が可能になったり、一定の条件を満たせば倫理審査委員会への付議を必要としないスキーマが盛り込まれたこと等、審査の迅速化、簡略化につながる内容も盛り込まれており、今回の改正には研究促進の側面もある。

このような改正がなされた背景を次項で説明する。

なお、臨床指針の改正のポイントと運用上の注意点は、局長通知「臨床研究に関する倫理指針の改正等について」(医政発第0731001号、平成20年7月31日)⁶⁾にわかりやすくまとめられているので参考にされたい。

各種倫理指針やGCPが改正された背景と臨床指針改正の基本理念

今回の臨床指針の改正は、当初より予定されていた見直しの時期にあたるため行われたことではあるが、改正の際の局長通知には、以下のように改正理由が記載されている。

近年の生命科学等の科学技術の進展に伴い、その実用化のための応用研究の重要性が一段と増している背景の下で、臨床研究において被験者を保護し、その尊厳及び人権を尊重しつつ臨床研究について、一層の適正な推進を図ることが求められてきたところである。

(中略)

臨床研究をとり巻く環境の変化に対応し、研究倫理や被験者保護の一層の向上を図るため全面的な見直しを行い、(後略)

ここでいう「臨床研究をとり巻く環境の変化」とは、新しいより良い医療を早く受けたいという国民の声が高まり、具体的には患者団体がドラッグラグや未承認薬問題の解消を求める活動を活発に行う等の世論の動きがあり、患者視点でも臨床研究にスピードを求めるようになってきたことと、その一方で、患者を含む研究参加者(以下、被験者)に対する保護の法制化を求める動きがあることを指していると思われる。

新しい治療方法の候補に、本当に期待する効果があるのかどうかを実際に患者で試すことなく調べられる方法があり、患者の情報や採取した組織を用いることなく新しい治療開発につながる新たな知見が見出せるような研究手法があれば、そのような研究においては被験者は存在せず、被験者保護の必要性は発生しない。しかしな

がら、ヘルシンキ宣言の5条に「医学の進歩は、最終的に人間を対象とする研究を要するものである」とあり、また、臨床指針の前文に「医療の進歩は、最終的には臨床研究に依存せざるを得ない」とあるように、実際にはそのような方法はなく、臨床研究には被験者の理解と協力が不可欠である。

であれば、被験者が安心して研究に参加できるように被験者保護を強化しつつ、研究のスピードを上げるしかないが、その実現のためにはどうすればいいのであろう。「はじめに」の項で述べたように臨床研究は多様である。それぞれの研究の属性に応じて求められるレベルの被験者保護に配慮して研究が適正に行えるように、つまり、被験者が身体的リスクを負う臨床試験ではしっかり被験者保護に配慮して研究を行い、その実施に先立ち倫理審査もしっかり行う、それ以外については、被験者が負うリスクの内容や大きさ、取り扱う個人情報や診療情報の利用の範囲、利用方法に応じて研究や倫理審査が行えるようにする等、適切に緩急つけた取り扱いができるように基準を定め、手順を明確にすることが必要である。今回の臨床指針の改正では、その基準を規定し、手順の「方針」を定めたといえるのではないだろうか。

ただし、臨床指針はGCPのように別途運用通知が発出されて基準や手順が明確に示されている訳ではなく、研究機関内での実務に落とし込む際に判断に迷う記述がある。その点について臨床指針を作成した厚生労働省医政局研究開発振興課の担当者は、臨床指針の説明会において質問を受けた際に、「philosophyを理解して各研究機関で判断して決めてください」と回答していた。つまり、各研究機関が臨床指針に基づき研究実施体制を整備する際には、臨床指針のphilosophy、言い換えれば各基準や手順が規定された本質的な理由、基本理念を理解し、自ら判断して決めるしかないのである。

臨床指針のphilosophy/基本理念は、臨床指針の前文にあるように「被験者の人間の尊厳及び人権を守るとともに、研究者等がより円滑に臨床研究を行う」ことであろう。そのために必要となる被験者保護について、次項で解説する。

臨床研究の被験者保護のレベルを規定する判断基準(研究の属性等)

臨床研究に求められる被験者保護のレベルを規定する研究の属性の主なものを以下に挙げ、臨床指針におけるそれぞれの定義を表1に、研究の属性に応じて求められる被験者保護対応を表2(介入研究の場合)、表3(観察