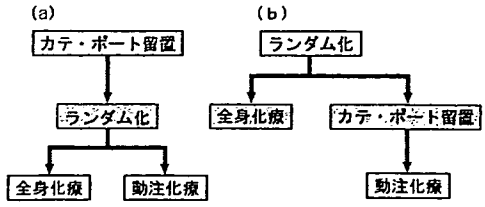

図5：第Ⅲ相試験における登録のタイミング

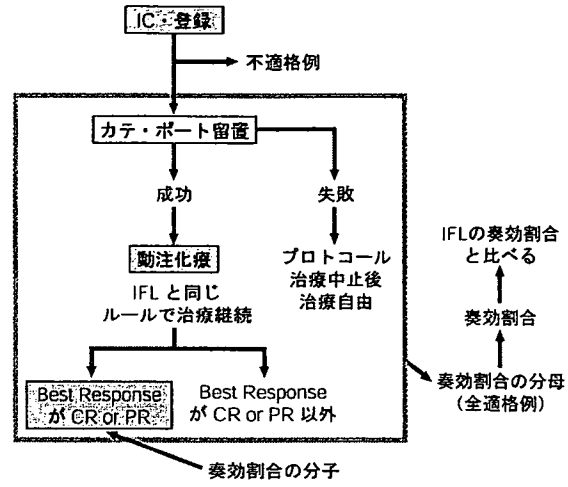

図6：仮想第Ⅱ試験の試験デザイン

おわりに

本文中で敢えて明記しなかったが、本稿のキーワードは「基本は比較」である。

統計家やデータセンターにとって、「何を何とどう比べるのか」を明確にして提案された試験計画は、もう解かれたのも同然の試験問題であり、サンプルサイズの計算は単にプログラムを起動するだけの事務的作業である。逆に「何を何とどうやって比べるのか」が明確でない試験計画は、デザインの完成まで延々と議論を繰り返さなければならないし、場合によっては残念ながら中止せざるを得ないこともある。

難しいと言われる臨床試験デザインのポイントは、実は「比べる」という日常ありふれた思考プロセスにあることを強調しておきたい。

連載「がん臨床試験の実践～JCOGを例に～」第2回

コンセプトの作成と審査 ―第三者に試験の意義を理解してもらう

国立がんセンター がん対策情報センター 臨床試験・診療支援部
福田 治彦

前回は臨床試験（以下、試験）の骨子を形作る過程での、デザインに関する考え方を概説したが、今回は、試験概要を文書として第三者に示すための「プロトコールコンセプト」について解説する。

プロトコールコンセプトの目的

●プロトコール作成は多段階方式が一般的

研究実施計画書（プロトコール）について、JCOGや欧米の共同研究グループ（Cooperative Group）では、最初からいきなり試験実施に必要な事項を網羅した数十ページにも及ぶfull protocol（フルプロトコール）を作成しない。その前段階として試験概要のみで構成される数ページのプロトコールコンセプト※（protocol concept：コンセプト）を作成する2段階方式を採用している。米国のThe Southwest Oncology Group（SWOG）では3段階方式を採用しており、第1段階として原稿1枚で作成される「capsule summary」をSWOGトップ数名で検討し、コンセプト作成を認めるかどうかを決定している。JCOGでもグループによっては、この方式を採用することもある。

2段階（多段階）方式を採用している理由としては、フルプロトコールの作成段階で試験に大きな問題が見つかった場合、修正に多大な時間を要することが多いし、また試験の意義そのものがグループ内外の審査過程で否定されて試験が中止になった場合、フルプロトコール作成にかかった労力と時間が無駄になるといったロスを避けるためである。そのため、「プロトコールコンセプト」の目的は以下の2点に集約される。

① 試験を実施するかどうか（採択/却下）を決定する
② 試験の骨子の段階で試験デザインを適正化する

●試験には“正しい”結論が必要

試験は第Ⅰ相や第Ⅱ相で十数名から数十名、第Ⅲ相となると数百名から数千名の患者さんに参加いただき、年余の時間と多くの人の労力と資金を投入して実施される。したがって、試験から得られる結果・結論はそれに見合う医学的・社会的価値（倫理要件としてはscientific/social valueと呼ぶ）が必要になる。

そして試験は“正しい”方法（scientific validity）によって、“正しい”結論を導かなくてはならない。そのため、米国ではコンセプトをCooperative Group自身による試験実施/非実施

※protocol outlineと呼ぶグループもあるが本稿では「コンセプト」で統一する

の判断材料にすることに加えて、スポンサーである National Cancer Institute（ NCI）の臨床試験管理 / 支援部門である Cancer Therapy Evaluation Program（ CTEP）でコンセプトが審査され、NCI スポンサーの研究としての採否が決められる。つまり、CTEP にコンセプトが承認されて初めて、Cooperative Group は第Ⅲ相試験を実施できるのである。表 1 に、CTEP のコンセプト書式を筆者が和訳したものを示す。ちなみに、第Ⅰ相試験、第Ⅱ相試験、第 I/II 相試験では、もう少し簡略化さ

表 1: NCI-CTEP の第Ⅲ相試験コンセプト書式項目（筆者による和訳、一部省略）

I. 管理情報（ Administrative）
- 研究名、研究番号、研究代表者（ 氏名、住所、電話、FAX、e-mail）、共同研究者

II. 試験概要（ Science section）
1. 研究仮説
2. 目的
   2.1 主目的
   2.2 副次目的（ 複数可）
3. 背景情報
   3.1 試験方法の妥当性とデザインの妥当性
   3.2 試験の重要性
   3.3 すべての関連データ
4. 適格条件（ 選択・除外の根拠を含む）
5. 群 / レジメン（ シェーマを含む）
   5.1 シェーマ
   5.2 群 / レジメン
6. 統計的デザインの詳細
   6.1 エンドポイント
      6.1.1 プライマリー・エンドポイント
      6.1.2 セカンダリー・エンドポイント
   6.2 ランダム化における層別因子・割付調整因子
   6.3 サンプルサイズと検出力
   6.4 解析計画（ 中間解析を含む）
   6.5 登録見込み（ 予定登録患者数 / 月、総予定登録患者数（ 最小～最大））
7. 実施可能性（ 適格患者数見込み、同意取得期待割合、類似の試験における経験を含める）
   7.1 競合する同グループ内の第Ⅲ相試験
   7.2 競合する他グループの第Ⅲ相試験
   7.3 競合する企業の試験

III. 薬剤情報（ Pharmaceutical section）
- 表形式で、薬剤名、CTEP による薬剤配布依頼の有無、治験薬かどうか、治験届者、プラセボ対照かどうかを列記する

IV. 附随研究（ Embedded correlative study section）
1. 附随研究名
   1.1 研究デザイン（ 検体取得方法、記録用紙、画像検査など）
   1.2 研究仮説（ 背景となる研究の情報を含む）
   1.3 統計的デザイン（ エンドポイント、サンプルサイズ、登録見込み / 月）

れた Letter of Intent（ LOI）という書式が用いられている。いずれも CTEP の Web サイトで閲覧できる（ http://ctep.cancer.gov/guidelines/index.html）。

CTEP のコンセプト審査は、各 Cooperative Group 担当の臨床医（ 各がん種の専門家）、使用する薬剤を専門にする臨床医（ 臨床薬理の専門家）、生物統計家、監査担当、薬剤師、薬事担当など異なる職種の専門家集団で行われるため、試験デザインの適正化の観点からも有用と言える。特に、統計的側面については「 Simon の 2-stage design」で有名な統計家の Richard Simon をはじめ、コンセプトやフルプロトコールを毎日のように審査し、経験と理論から新たな統計的デザインを開発している統計家が担当しているため、CTEP の審査を受けることで経験豊富なアドバイスが受けられる。また臨床医による審査では、直近の米国臨床腫瘍学会（ American Society of Clinical Oncology: ASCO）での発表など、最新のエビデンスも踏まえた上で科学的妥当性が検討されているため、既知の事実による無駄な試験の防止、より有用な臨床的仮説や考察の追加が行われている。

以上より、コンセプトを審査に提出する場合、表 1 と同程度の詳細さを有した情報が必要であることを理解されるであろう。臨床研究を計画する際には、外部審査を受けるかどうかは別にして、こうした書式を用いてコンセプトを作成し、研究者同士で検討することで、試験デザインが適正かどうかを確認することは有用と思われる。是非、利用していただきたいと思う。

JCOG における「 コンセプト 」審査

● JCOG におけるコンセプト 審査の変遷

JCOG におけるコンセプトの審査過程の変遷を簡単に紹介する。

1995 年まで、JCOG にはコンセプトのシステムはなく、筆者が JCOG データセンター（ 当時は JCOG 統計センター）に加わった 1996 年に、SWOG の「 capsule summary」を参考にした A4 サイズ 1 枚のコンセプトを各グループから提出していただき、JCOG 運営委員会会議で試験を検討し、実施の採否を決定する仕組みを導入した。しかし運用してみると、A4 サイズ 1 枚のコンセプトに含められる情報では、試験デザインの適正化を判断するという目的が果たせず、結局フルプロトコールを作成する段階になって大きな問題が見つかり大幅なデザインの変更を要した、プロトコールが完成しなかった等の事態が生じた。そこで 1998 年より、SWOG やヨーロッパの European Organization for Research and Treatment of Cancer（ EORTC）のコンセプト書式を参考にプロトタイプを作成し、1999 年より現行に近いコンセプト書式（ 表 2）を用いるようになった。

CTEP の書式では「 記載量に制限はない」と記されているが、JCOG では 8 ページ以内を原則とし、10 ページ程度までを許容範囲としている。

● JCOG におけるコンセプト 作成と審査

現在 JCOG におけるコンセプトの作成と審査は、以下の手順で進められている。

1) グループの研究者によるコンセプトの作成
2) データセンターとの事前相談（ オプション）
3) プロトコール審査委員会でのコンセプト予備審査
4) プロトコール審査委員会会議での検討
5) JCOG 運営委員会での本審査と採決

以下、これらのプロセスを簡単に紹介する。

1) グループの研究者によるコンセプトの作成

前回紹介した試験デザインの骨子にあたる「 対象（ 患者）」「 治療」「 評価」が概ね固まり、各グルー

表 2: JCOG のコンセプト 書式(一部省略)

0. シェーマ
- ランダム化比較試験、複数のレジメンの組み合わせ、複数のモダリティの組み合わせの場合、シェーマを作成する

1. 目的
- 2 ～ 3 行を目安に簡潔に試験目的を記述。対象集団(stage) と評価する治療法を明確に表現する
- primary endpoint(原則 1 つ)、secondary endpoint(s) を記述

2. 背景

以下の内容について、他分野の研究者が理解できる平易な表現にて明確に記述する。極力略称は用いない
- 対象疾患、疫学的事項(疾患の頻度)、疾患の特徴、対象 stage の特定
- 対象集団選択 / 設定の根拠: なぜこの対象としたか。主な適格規準の設定根拠を記述
- 対象患者における標準治療(state-of-the-art treatment) と予後。標準治療が確立されていない場合はその旨明記する (広く用いられている治療は何かを記述)
- 使用する薬剤の説明(単剤における主な毒性、有効性データを含む)
- 治療レジメン選択の根拠、過去の同一薬剤・類似レジメンの臨床試験データ
- 試験デザイン (phase II か phase III か。等) および endpoint の設定根拠
- phase III 以外の場合、次のステップの試験の概略(phase II の場合は、その試験で有望な結果が得られた場合に行うであろう phase III の概要)

3. 患者選択規準

以下の項目について、他分野の研究者が理解できる平易かつ明確な表現を用いて記述する。
- 疾患、組織型: 組織学的・細胞学的確診を要求する場合、生検、過去の手術材料の組織診、細胞診等、許容される方法・検体等を明確に定義する。要求しない場合、臨床診断として許容される範囲を明確にする
- stage: 用いる staging criteria とともに明記する (criteria の名称のみでよい)
- 病変の評価可能性: 測定可能病変のみ／測定可能病変または評価可能病変
- 前治療の規定: 特定の前治療からの休止期間を設ける場合、月でなく「 日」もしくは「 週」で規定する
- PS: ECOG performance status score を用いる
- 併用薬・併用療法に関する制限事項(ある場合)
- 臓器機能(臨床検査値): 主なもののみ

4. 治療計画

特に複数のレジメンや複数のモダリティによる治療レジメンの場合、「 プロトコール治療」の定義を明確に行う。後治療との区別を明確に定義する。

ランダム化試験では群別、集学的治療の試験ではモダリティ別に、小見出しをつけてそれぞれ記述する。

4.x. プロトコール治療中止・終了規準
- プロトコール治療完了とみなすコース数、原病の増悪、治療中止とすべき毒性(有害事象)、コース開始延期の許容範囲もしくはプロトコール治療期間全体の延長許容範囲等の判断規準のうち、主なものを明確に記述する
- 減量規準等、詳細な治療変更規準の記述は不要

5. 効果判定の方法と判定規準
- 効果判定の時期(○コース毎、○コース終了後、プロトコール治療終了後○日…等)
- 評価のモダリティ (頭部 CT、胸部 CT、胸部 X 線等、必須のものと許容範囲が明確になるように記述) と判定規準
- 効果判定を行わない試験では「 本試験では効果判定は行わない」と記述

6. エンドポイントと統計学的考察
- primary endpoint と secondary endpoint(s) を明記
- ランダム割付を行う場合の割付調整因子
- 第Ⅲ相試験: 標準治療群の生存時間分布の推定値(5 年生存率や生存期間中央値) 等とその根拠。臨床的に意味がある群間の差。α、β
- 第Ⅱ相試験: 期待奏効率や閾値奏効率、α、β 等
- 予定症例数と予定登録期間の概算
- 中間解析の概略について記述

7. 集積見込み
- 年間登録数の見込みおよび予定登録期間内の登録見込み
- 原則として同一疾患における過去のJCOG試験での登録実績に基づく。該当するデータがない場合、参加施設の年間適格患者数や手術例数等から推定した結果を記述する

8. その他特記事項

該当する場合、簡潔に記述する。
- 病理中央診断
- 腫瘍縮小効果の中央判定や施設外校閲
- 凍結検体等を用いた付随的研究
- 適応外医薬品を用いる場合その取扱について

9. 研究班

主たる研究班の名称と主任研究者。本試験の研究代表者と研究班の関係を明記。
17指-5「多施設共同研究の質の向上のための研究体制確立に関する研究」班以外で共同研究体制とする研究班があれば何を役割分担するかも含めて明記。

プの会議で研究者間の合意が得られれば、試験の実務責任者となる臨床医である研究事務局が表2の書式に従ってコンセプトを作成する。

また、各グループの代表者は比較的若手である臨床医を2名、医学審査員(medical reviewer)として推薦している。医学審査員は自グループのコンセプト作成を支援する義務を負うと同時に、プロトコール審査委員会(Protocol Review Committee: PRC)にも参加する。医学審査員は、日頃からPRCメーリングリスト(後述)やPRC会議におけるコンセプト検討に参加しているため、JCOGにおけるコンセプト検討の方法に詳しい臨床医と言える。

コンセプトを提出する際には、内容について医学審査員が承知している証左として、「コンセプト事前チェックリスト」を一緒に提出するようになっている(表3)。また、グループ代表者がコンセプトを承認している証左として、カバーレターにはグループ代表者の署名を要求している。官僚主義的と思われるかも知れないが、過去にグループ代表者を初めとする当該グループの研究者が知らないまま、コンセプトが提出され混乱を来したことがあったため、やむを得ず以上の方法で行っている。

2) データセンターとの事前相談

最近開始したシステムとして、PRC提出前にグループがコンセプト内容をJCOGデータセンターに相談することができる。義務ではないためすべてではないが、ほとんどのコンセプトで事前相談が行われている。

相談内容は試験デザインの骨子である「対象(患者)」「治療」「評価」の中でも、特に「評価」に関してエンドポイントや統計的デザインをデータセンターがアドバイスすることによってPRCに提出するコンセプトの質を高めており、PRCでの議論の効率化に貢献していると考える。相談では研究事務局以外の若手臨床医の参加も許可しているため、勉強の場として活用されており、概ね好評を得ている。

表 3: JCOG におけるコンセプト 事前チェックリスト（一部省略）

体制・体裁

1. グループ内での検討
   1) グループ班会議などで十分な（最低 1 回以上）コンセプトの検討が行われているか。
   2) グループの医学審査員 / 委員のレビューを受けているか。
   3) 対象集団、標準治療などについてグループ内のコンセンサスが得られているか。
   4) プロトコール検討会への出席予定者の調整は済んでいるか。（研究代表者、事務局いずれかは必ず出席しなければならない）
2. コンセプトの書式
   1) 章構成は書式に従っているか。
   2) 総ページ数は多くなりすぎていないか。（原則 8 ページ程度）
   3) 治療変更規準を細かく書きすぎていないか。（減量規準などの詳細は不要）
   4) 研究班の項は抜けていないか。

内容

1. 対象集団
   1) 対象疾患の特徴・予後などについて十分に記載されているか。
   2) 対象となるサブグループ（Stage 毎など）の特徴・予後などが十分に記載されているか。
   3) 上記の記載と患者選択規準の整合性が取れているか。
2. 対象に対する標準治療
   1) 標準治療のエビデンスは十分記載されているか。
   2) 海外と日本での標準治療（日常診療）の違いについて十分記載されているか。
3. 試験治療
   1) 試験治療のエビデンスは十分に記載されているか。
   2) 標準治療に比べてのリスク / ベネフィットは十分記載されているか。
   3) 上記のリスクは許容範囲内か。最小化する余地があるか。
   4) その他、開発中の他の薬剤・治療法（欧米で行われている試験など）について記載されているか。
4. 試験デザイン
   1) 臨床的仮説と試験デザイン、エンドポイントについて十分に記載されているか。
   2) 上記の記載と統計記述（α、β、見込まれる差など）の整合性が取れているか。
   3) Primary endpoint、secondary endpoints と試験デザインの整合性は取れているか。
5. 試験の実施可能性
   1) 予定登録数および年間登録数見込みは十分に記載されているか。
   2) 上記の記載の実施可能性は適切と判断されるか。
   3) 保険診療 / 適応外使用 / 医療経済上の問題はあるか。
6. その他
   1) 記載が非専門分野の研究者にも分かる様な平易な表現が用いられているか。
   2) 附随研究が適切に設定・記述されているか。
   3) 病理・効果判定などの中央判定が適切に設定・記述されているか。

3) プロトコール審査委員会でのコンセプト 予備審査

PRC に提出されたコンセプトについて、PRC 事務局は書面による予備審査を行う。予備審査メンバーは、当該グループ以外の医学審査員から medical reviewer（3 名）、および primary reviewer と secondary reviewer（各 1 名）、さらに統計外部委員から統計 reviewer（1 名）、最後に筆者（DC

reviewer）の計７名である。審査項目には

1. 対象集団の選択は適切か
2. 対象に対する標準治療は適切か
3. 試験治療の選択は適切か
4. 試験デザインは適切か
5. 試験の実施可能性はあるか
6. 標準治療への貢献度はどの位か
7. 当該疾患の専門家以外の研究者にも研究の意義が理解できるか

があり、それぞれ自由に意見を述べると共に、各項目に「5（特に優れている）、4（優れている）、3（ふつう）、2（やや劣る）、1（劣る）」の5段階で点数を付ける（満点35点）。この点数はJCOG運営委員会での採否決定の参考資料となる。

審査委員のreview sheetはPRCメンバー全員で構成されるPRCメーリングリストに投稿され、PRCメンバーで共有されるとともに、当該グループの医学審査員を介して、研究事務局を含む当該グループにフィードバックされる。

当該グループはreview sheetの審査意見や質問に対し、PRC会議までに、可能な限りメーリングリストにて回答を行う。メーリングリストではreviewer以外のメンバーでも自由に発言することができる。

4）プロトコール審査委員会会議での検討

予備審査を経て、会議によるコンセプト検討（PRC会議）が行われる。PRC会議は通常18時に開始し、短い場合は2時間程度だが、3時間を超える場合もある。

PRC会議には、コンセプトを提出した研究事務局と当該グループの医学審査員1名以上の出席が義務づけられており、可能な限り当該グループの代表者も出席するよう要請している。さらに当該グループの他の研究者も出席可能となっている。PRCメンバーはreviewer以外であっても可能な限り出席してもらい、JCOGデータセンターの統計部門とJCOG運営事務局の研究支援部門は原則として出席するため、毎回概ね30～40名での会議となる。PRC会議では原則として、研究支援部門の医師が司会を行う。進行は疾患の背景→試験の対象集団→対象集団の標準治療→試験治療の順に議論し、それらを踏まえて試験デザインの検討を行う。

PRC会議には当該グループの医学審査員、統計reviewer、データセンターのグループ担当統計家が出席しているため、たとえ当該グループでは常識となっていることでも他分野の研究者から見て疑問に思われることについては、遠慮なく第三者的な批判が行われるため、他分野の専門家同士のpeer reviewと位置付けられる。当該グループは疑問に対して、他分野の専門家が納得するrationale（論拠）示さなければならないため、この会議は若手の臨床研究者の教育の場としても貴重な機会になっている。

書面による予備審査で未解決となっていた問題の多くは、PRC会議で直接議論することによって解決の道が見つかることが多く、当該グループ側はJCOG運営委員会での採否決定（本審査）までに、コンセプト原案における問題点の解決策を練ることができる。

PRC会議で検討された内容は、研究支援部門の医師が作成する議事録にまとめられ、PRCメーリングリストによって出席者間で内容が確認される。議事録は、JCOG運営委員会への提出資料となる。議事録には、PRC会議で当該グループ側が示せなかったデータやエビデンスを「追加コメント」として記載することができるため、JCOG運営委員会会議での検討をスムーズに進めることに役立っている。

5) JCOG 運営委員会での審査と採択

予備審査、PRC 会議で検討されたコンセプトは、年 4 回開催される JCOG 運営委員会会議の場で行われる本審査にて採否が決定する。

運営委員会の本審査では、まず、PRC に提出されたコンセプトを用いて、当該グループ代表者もしくは研究事務局が試験の概要を説明し、質疑を行う。続いて primary reviewer が PRC 会議の要約と PRC 会議で出されたアドバイス等が議事録を用いて説明される。

運営委員全員での質疑応答の後、運営委員全員が「採択」「却下」「保留」のいずれかに挙手をする。出席している運営委員の過半数が「採択」に挙手した場合、そのコンセプトが「承認」され、JCOG 試験としてのプロトコール作成が許可される。多くの場合は運営委員の大多数が「採択」に挙手することで、承認される。しかし、「採択」挙手がかろうじて過半数になることもあるし「採択」挙手が過半数に届かず、「却下」されることもある。

コンセプトで重視される「背景」

●コンセプトで重要なのは「試験の意義を示す記述－声明文」的側面

前回紹介した S. J. Pocock の「Clinical trials: a practical approach」(1984, John Wiley & Sons Ltd.) には「プロトコール」の意義として、「試験の意義を示す記述－声明文」的側面と「試験の実施手順－マニュアル」的側面の 2 つの側面があるとされている。

JCOG におけるコンセプトの取り扱いを考えると、重要なのは「試験の意義を示す記述－声明文」的側面であることが理解されよう。臨床試験を行うグループは他の分野の専門家に対して、その研究の重要性、計画の適切性、実施の可能性を示す必要があり、そのツールがコンセプトと言える。したがって、コンセプトで最も重要なのは、表 1 における「3. 背景情報」であり、表 2 における「2. 背景」なのである。

試験によって求められる結論が臨床的に価値の高いものであれば、導かれる結果を信頼性の高いものにすることは、ある意味技術的な問題と言える。試験デザイン上の問題点はほとんどの場合、統計家を初めとするデータセンターのメンバーが知恵を絞ることによって、解決が可能であるためである。

●臨床研究者の最も重要な使命はコンセプトの記述にある

試験を計画する臨床医(臨床研究者)の最も重要な使命は、臨床的価値の高い研究仮説(clinical question)を設定すること。そしてその重要性を他分野の専門家に理解してもらうと同時に、協力が得られるようなコンセプトの記述(作文)をすることである。

前回強調した試験デザインの骨子である「患者」「治療」「評価」を「標準治療は何か？」を核として具体化するプロセスは、今回示した JCOG および CTEP の書式に従って、質の高いコンセプトを記述するプロセスそのものと言える。前回と今回の記事を併せて参照しつつ、質の高いコンセプトの作成を試みていただきたい。

連載「がん臨床試験の実践～JCOGを例に～」第3回

プロトコールの作成と審査(前編)

――異なる背景を持つ人が使うコミュニケーションツール

国立がんセンター がん対策情報センター 臨床試験・診療支援部
福田 治彦

第1回では試験デザインを、前回は試験デザインを文書として第三者に示すツールである「プロトコールコンセプト」を取り上げた。今回は臨床試験のメインツールである「プロトコール」を取り上げる。プロトコールの内容は大部となるため、2回に分けて掲載する。前編の今回は、プロトコールとは何か、プロトコールは誰が読むか等を確認し、後編ではプロトコールの標準化および各章で注意すべき点を述べる。

1. プロトコールとは何か？(What is protocol?)

●プロトコールの定義

一般的には「研究実施計画書」、治験では「治験実施計画書」のことを「プロトコールまたはプロトコル(protocol)」と呼ぶ。そもそも、プロトコールとはどういう意味なのか、広辞苑には「①(条約の)原案。議定書。②外交儀礼。③コンピュータ・システムで、データ通信を行うために定められた規約。～以下省略～」と記載されている。一方、Oxford現代英英辞典では、以下のように記載されている(カッコ注、下線は筆者による)。

「1. a system of fixed rules and formal behaviour used at official meetings, usually between governments, 2. the first or original version of an agreement, especially a treaty(協定、条約) between countries, etc., 3. a set of rules that control the way data is sent between computers, 4. a plan for carrying out a scientific experiment or medical treatment.」

つまり研究実施計画書をプロトコールと呼ぶのは原義からの転用であって、古くは、条約原案や外交儀礼書がプロトコールと呼ばれていたことがわかる。さらにフリー百科事典『ウィキペディア(Wikipedia)』では、以下のように記載されている(下線は筆者による)。

『プロトコル(protocol、プロトコールとも)とは、複数の者が対象となる事項を確実に実行するための手順等について定めたもの。日本語では、場合に応じて規定、議定書、儀典などと訳される。

国際儀礼上のプロトコルとは、国際的な実務、交流の場における公式なルールや慣習などを指すものである。主に、列席者の序列、国旗の取扱についての原則、パーティーなどの場における進行や服装・マナーについての一般的な運用方法を示すものだが、これといった厳格な定義はない。日本では、外務省大臣官房儀典官室(英語名"Protocol Office")が「外交官及び領事官の接受・差遣、外国人に対する栄典の授与に関する推薦、外交上の儀礼」などといった「プロトコール」を

所掌している。なお、コンピュータ用語などでは"Protocol"は「プロトコル」と表記・発音されることが多いが「国際儀礼」の意味で使う場合はプロトコール」とされることが一般的である。これは、前者が英語由来の外来語であるのに対して、後者がフランス語に由来する外交用語であることが理由であると思われる。』

以上の引用には、プロトコールの本質を表すいくつかのキーワードが含まれている。1つは「複数の者、国際、国家間、政府間」というユーザー（または読者）に関する特性であり、もう1つは、「formal、official、公式な」という位置づけに関する特性である。すなわちプロトコールとは「異なる背景や文化を有する複数の（異なる職種の）人間が共同で作業を（臨床試験を）行う上でのルールや手順を記述した公式文書」と定義することができる。ちなみに、S. J. Pocockの『Clinical trials: a practical approach』（1984, John Wiley & Sons Ltd.）では「formal document specifying how the trial is to be conducted: どのように試験が行われるかが記述されている公式文書」と「公式」という言葉がきちんと入っている。これに対して、類義語であるレジメン（regimen）は、「a set of rules about food and exercise or medical treatment that you follow in order to stay healthy or to improve your health」（Oxford現代英英辞典）とある。「なにか身体に良いことをするための一連の手順」がレジメンであるから、筋トレプログラムもレジメンの1つと言えよう。ただし、ダイエットプログラムを「レジメン」と呼ぶかどうかは定かではない。

●プロトコールの位置づけ

本題に戻ろう。以上のことから、「プロトコール」は治療レジメンの上位概念であり、治療レジメンを含めた、研究全体の手順を記した公式文書と位置付けることができる。少なくとも臨床試験においては、研究実施計画書全体をプロトコールと呼び、治療内容の一連の規定を治療レジメンと呼ぶと混乱しないであろう。ちなみに「プロトコールをたくさん走らせる」のように、研究・試験そのものを指す場合もあるが、混乱を避けるためにも、臨床試験そのものは臨床試験と呼び、その手順を記した文書はプロトコールと呼ぶことを推奨する。

2. 誰が読むのか？（Who reads protocol ?）

●メモや覚書はプロトコールではない

例えば、実験方法を先輩から教えてもらう場合、多くの基礎研究者は試薬の調製方法、実験機器の使い方、反応させる時間等を聞いて記録し、それを基に自分なりのマニュアルを作成するであろう。このマニュアルは「作業を確実に行うため、一連の手順を書いた文書」という意味ではプロトコールに似ているが、自分で読む文書、すなわち公式文書ではないため、プロトコールとは言えない。

同様に同じ実験室に在籍する、同じ程度の知識と技術を持った、同じ専門用語で会話する基礎研究者間で共通の作業を行うための文書やメモ書き等も、おそらくプロトコールとは言えないだろう。しかし異なる知識・技術を有する実験助手、または機器や試薬のメンテナンスを担当する助手と手順を共有する目的で作成された「○○研究室××遺伝子解析作業手順」といった文書であれば、プロトコールと呼べると思われる。

●プロトコールにおける3つのユーザー

ところで、プロトコールは、誰が読み・使うのであろうか？筆者はローカルユーザー、セントラルユーザー、レビューアーの3つに分けて考えている。

1) ローカルユーザー(local users)

ローカルユーザーとは患者・被験者の診療に当たる医療現場でプロトコールを使う臨床医、CRC(臨床研究コーディネーター)、看護師、薬剤師、臨床検査技師などを指す。多くの試験では、職種や対象疾患に関係なく、ユーザーになることが考えられる。したがって、プロトコールは当該疾患を専門とする臨床医には常識であることについても、非専門家のユーザーにも理解できるように詳しく書かれていなければ、その機能を果たし得ないことになる。

こうした認識は、既に20年以上の歴史を有するJCOGであってもまだ浸透しているとは言い難く、未だに「施設の担当医にとっては自明であるから、プロトコールに記載する必要なし」と考える臨床医もまだ数多く存在する。

2) セントラルユーザー(central users)

多施設共同臨床試験を行う際、施設のリソースを除いた中央機構のことを「セントラル」と呼ぶ。セントラルユーザーについて、JCOGを例にすると、JCOGデータセンターのデータマネージャー、統計家、JCOG運営事務局のプロトコールコーディネーターや試験進捗情報管理担当者、各種委員会の事務局担当者が該当する。

その他にも企業治験におけるモニターや、施設訪問監査の担当者もセントラルユーザーと言えるだろう。

3) レビューアー(reviewers)

上記1と2は試験実施者側のスタッフであるため、「ユーザー」と位置付けたが、ユーザーとは異なる用途でプロトコールを読むレビューアーがいる。

レビューアーはローカルレビューアーとセントラルレビューアーに分けて考えることができる。ローカルレビューアーには、施設の倫理審査委員会の委員や事務局があげられる。一方、セントラルレビューアーは、プロトコール審査委員会(Protocol Review Committee: PRC) や効果・安全性評価委員会(Data and Safety Monitoring Committee: DSMC) の委員や事務局、規制当局で許認可のために試験結果を審査する審査官があげられる。

これら1～3を通じて言えることは、実際にプロトコールを読み、使用する人を考えると、当該疾患の専門家であっても、プロトコールを作る側の責任者である臨床医(JCOGでは研究事務局と呼ぶ) は、人数の上ではごくわずかに過ぎないということである。したがって、その人にしか理解できない書き方でプロトコールを作成し、「事足れり」と考えることは如何に浅薄であるかがわかると思う。

3. 何が書かれるべきか？ (What should protocol include?)

●プロトコールとプロトコールコンセプトの違い

前回、プロトコールの意義として「試験の意義を示す記述－声明文」的側面と「試験の実施手順－マニュアル」的側面があり、プロトコールコン

セプトでは声明文的側面がより重要であると述べた。しかし、プロトコールコンセプトはあくまでも研究組織が試験実施の是非を検討するための内部文書であり、たとえ質の高いコンセプトが作られたとしても、プロトコールで声明文的側面を省略できるわけではない。

●プロトコール記述における2つの側面

「プロトコールに何が書かれるべきか？」については、Pocockの『Clinical trials』に則って、「試験の意義を示す記述－声明文」的側面と「試験の実施手順－マニュアル」的側面から概説する。

1）試験の意義を示す記述－声明文的側面

声明文的側面としてプロトコールに書かれるべき内容は、以下の通りである。

①なぜ、今、この試験を行う価値があるのか（rationale）
②研究者相互の認識の共通化に必要な情報
③試験実施機構の意思決定（その試験を行うかどうか）の判断材料となる情報
④施設が受け入れるかどうか（施設IRB審査）の判断材料となる情報
⑤患者さんへの説明に必要な情報

上記のうち、プロトコールコンセプトのシステムを有している研究組織では、①～③をプロトコールコンセプト、およびプロトコールでカバーしているが、④と⑤はプロトコールのみでカバーしなければならない。また、④と⑤の内容はプロトコールの「背景」に記載されるべき内容である。

2）試験の実施手順－マニュアル的側面

マニュアル的側面としてプロトコールに書かれるべき内容は、以下の通りである。

①対象患者の特定、登録の手順
②治療の内容、評価の方法
③データ収集と管理の方法、解析の方法
④責任体制、連絡先

前章で述べたように、これら試験の意義や実施手順は、非専門家の臨床医および臨床医以外で試験に携わる人が理解できる文章で書かなければならない。当然、日本語として正しく、かつ読み易いことが必要であることは言うまでもない。

4. どういう構造であるべきか？（What should protocol be structured?）

●プロトコールの“お作法”とは

プロトコールの章構成について、誰がいつどうやって決めたのか定かではないが、『Clinical trials』には、表1のような例が載っている。プロトコールの章構成はそれぞれの組織が各々決めてきたが、いくつかを比べてみると一定の“お作法”があることに気付く。表2に米国The Southwest Oncology Groupの例を示したので、まずは表1と表2をじっくりと見比べてみていただきたい。

章構成についてはヨーロッパのEuropean Organization for Research and Treatment of Cancer（EORTC）や日米欧三極合同でのICH-GCPも概ね同様である。こうした共通の“お作法”というべき慣習を整理してみたのが表3である。“お作法”の例を1つ述べると「試験の意義を示す記述－声明文」的側面は試験の意義（背景、目的）に書かれており、それ以外の章は基本的には「試験の実施手順－マニュアル」的側面が書かれていることに気づくだろう。

● JCOGにおけるプロトコールの書式

最後にこれらの考察を踏まえて標準化したJCOGにおけるプロトコールの章構成を表4に示した。JCOGのすべてのプロトコールは、この章構成で作成される。特殊な研究で、章ごと該当する内容がない場合も、章をなくして番号を繰り上げるのではなく、「該当せず」として章番号も標準化している。

次回はJCOGでのプロトコールの標準化と各章毎の注意点について触れたいと思う。

表1. プロトコール章構成の例1(Pocock: Clinical Trials)

1. 背景と総論的な目的	Background and general aims
2. 特定の目的	Specific objectives
3. 患者選択規準	Patient selection criteria
4. 治療計画	Treatment schedules
5. 患者評価の方法	Methods of patient evaluation
6. 試験デザイン	Trial design
7. 登録とランダム割付	Registration and randomization
8. 患者の同意	Patient consent
9. 必要症例数	Required size of study
10. 試験進捗のモニタリング	Monitoring of trial progress
11. 記録用紙とデータの扱い	Forms and data handling
12. プロトコール逸脱	Protocol deviations
13. 統計解析計画	Plans for statistical analysis
14. 責任体制	Administrative responsibilities

表2. プロトコール章構成の例2(The Southwest Oncology Group)

0. シェーマ	10. 効果判定
1. 目的	11. 統計的考察
2. 背景	12. 専門的検討
3. 薬剤情報	13. 登録方法
4. 病期分類規準	14. データ提出
5. 適格規準	15. 特記事項
6. 層別因子	16. 倫理と規制要件
7. 治療計画	17. 参考文献
8. 評価する毒性と用量変更規準	18. CRF一式
9. スタディカレンダー	19. 付表

表３．章構成の標準仕様

●試験の意義 ・背景、目的	通常、初めにある
●患者選択と登録 ・適格条件、登録手順 ●治療と毒性 ・治療計画、予期される毒性、治療変更規準 ●評価 ・臨床評価項目、臨床検査、効果判定の方法	臨床現場／データセンターで何度も参照する部分 実際の時系列に添う事が望ましい
●解析とデータ管理 ・記録用紙、データ収集、エンドポイントの定義、解析計画、統計学的考察 ●品質管理と品質保証 ・モニタリング、中央判定（施設外判定）、監査 ●倫理や規制要件 ・倫理、インフォームドコンセント、有害事象の報告 ●管理責任体制 ・研究組織、結果公表のポリシー	1度は読んでおく

表４．プロトコール章構成の例３（JCOG）

0. 概要	10. 有害事象の報告
1. 目的	11. 効果判定とエンドポイントの定義
2. 背景と試験計画の根拠	12. 統計的事項
3. 本試験で用いる規準・定義	13. 倫理的事項
4. 患者選択規準	14. モニタリングと監査
5. 登録・割付	15. 特記事項
6. 治療計画と治療変更規準	16. 研究組織
7. 薬剤情報と予期される有害反応	17. 研究結果の発表
8. 評価項目・臨床検査・評価スケジュール	18. 参考文献
9. データ収集	19. 付表 説明／同意文書、CRF一式など

腫瘍内科，1(2)：144-148，2007

特集 臨床試験

中間解析*

吉村健一**

Key Words：clinical trials, interim analysis, group sequential analysis, alpha-spending function, multiple comparison

はじめに

現在実施中の臨床試験の途中結果をみたところ，新治療に割り付けられた対象者の成績が標準治療に割り付けられた対象者のそれに比べて明らかに優越しているようにみえる状況に直面した場合，研究者としてどういうアクションをとるべきであろうか．あるいはこれとは反対に，試験計画時に抱いた期待に反して新治療群の成績が標準治療群に比べて明らかに劣っているように見える状況であった場合，研究者としてどういうアクションをとるべきであろうか．どちらの状況においても実際に観察された群間差の大きさに応じて適切なアクションをとるべきであることは当然であるが，試験を早期中止すべきかどうかについて臨床的観点，統計的観点，倫理的観点を総合して悩ましい決断を迫られる状況が現実には意外に多く存在する．近年，試験の実施中にその時点までに累積した結果を統計的に適切な方法を用いて評価すること(これを中間解析という)がしだいに一般的となっている．中間解析の目的は，その結果に基づいて試験継続か早期中止かを判断することである．臨床試験における統計的原則を述べた国際的ガイドライン〔International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) E9〕[1]においても以下のように定義されている．

中間解析　Interim Analysis

試験の正式な完了以前に，有効性または安全性に関して試験治療群間を比較することを意図して行われるあらゆる解析．

このガイドラインに従えば，臨床試験の途中で群間比較を行うことを中間解析といい，中間解析で行うことは生存期間などに代表される有効性に関する群間比較のみに限られず，毒性などの安全性に関する群間比較も含まれる．本稿では試験の正式な完了以前に行うこの種の群間比較のことを中間解析，試験の正式な完了以後に中間解析による早期中止がなかった場合に当該試験の主たる結論を下すために行う群間比較のことを最終解析と呼ぶこととする．

ここでは，この中間解析についてその概要を解説することを目的とする．

倫理的観点からみた中間解析

臨床試験ではヒトを対象者としており，統計的側面からのみではなく倫理的側面からも適切な研究デザインを採用することが要請される．

これから実施しようとする研究において中間

* Interim analysis.

** Kenichi YOSHIMURA, Ph.D.: 国立がんセンター がん対策情報センター 臨床試験・診療支援部/医学統計室長〔〒104-0045 東京都中央区築地5-1-1〕; Biostatistics and Epidemiology Section/Clinical Trials and Practice Support Division, Center for Cancer Control & Information Services, National Cancer Center, Tokyo 104-0045, JAPAN

解析を予め計画しておかないことは,研究者が試験を実施するより前にもっていた有効性や安全性に関する予想が誤りであった場合に対する倫理的な担保が欠落していることに対応する．先行する研究や生物学的な根拠にいくら精緻に基づいて当該試験を計画しようとも，実際には事前の期待を超えるほど大きく新治療群が標準治療群に比べて優れることや，これとは反対に事前の期待に反して新治療群が標準治療群よりも著しく劣ることなどは往々にしてありうる．前者の事前の期待を超えて新治療群が優る場合，最終解析時点よりも早期の時点であっても十分に検証的な結果を得ることが可能である．中間解析を実施することによって，実施しない場合に比べてより早期の段階で十分に検証的な結論を得ることができるのであれば倫理的な観点より好ましいと言える．また，試験途中で新治療群が著しく劣っていることが明らかになった場合に試験を早期中止できるように中間解析を適切に計画しておくことも倫理的な観点より好ましいと言える．世界医師会によるヘルシンキ宣言(ヒトを対象とする医学研究の倫理的原則)[2]においても以下のようにされている．

> *医師は，内在する危険が十分に評価され，しかもその危険を適切に管理できることが確信できない場合には，ヒトを対象とする医学研究に従事することを控えるべきである．医師は，利益よりも潜在する危険が高いと判断される場合，または有効かつ利益のある結果の決定的証拠が得られた場合には，すべての実験を中止しなければならない．*

臨床試験において中間解析を適切に計画する必要性が倫理的な観点からも支持されていると言える．

一方で，実際には早期中止にたる結果が得られていないのにもかかわらず，不適切な中間解析による結果に基づいて早期中止することも倫理的な問題となりうる．ここで不適切な中間解析というのは，統計的観点から不適切なもの，実施運営上不適切なものなど，不適切な手順で行われたもの全てを指す．たとえば，中間解析時の結果において見かけ上は臨床的に大きな差がある場合でも，後述する統計的に適切な調整を伴うと十分には検証されているとは言えない状況は大いにありうる．一般には，試験早期に行った中間解析であるほど臨床的印象と統計的調整を伴った結果の間での食い違いが生じやすい．統計的調整を伴うと早期中止に至る結果が得られないということは現在得られているものが偶然の範囲内であるということを一般に意味する．つまり，そのまま試験を継続した場合には中間解析時の印象とまったく異なるような結果が将来の解析時に得られる可能性が決して小さくないということである．このような状況では臨床的印象だけにとらわれることなく，統計的観点，倫理的観点をも総合した上で当該試験に対する適切な判断を行うべきである．

統計的観点からみた中間解析

中間解析を計画した試験デザインを採用する場合，最終解析に加えて中間解析を実施するという意味で計2回以上の群間比較を同一試験内で行うことになる．これより，統計的観点からは検定の多重性(multiplicity)と呼ばれる問題が生じる．

検定の多重性とは検定を2回以上行うことによって生じる．これが生じた場合には後述する適切な方法を用いない限り，試験全体でのαエラーの確率を事前に定めたもの以下に保つことができない．αエラーとは本当は差がないのに差があると誤って判断しまう誤りのことである．p値が5％以下となった場合に統計的有意と判断する(これを有意水準5％という)検定を1回のみ行う場合にαエラーを犯す確率は5％に等しくなる．これは仮説検定の定義から導かれるものであり，実際に得られたp値にはよらない．一方で，2つの検定をそれぞれの有意水準5％で行うとαエラーを犯す確率は5％より大きくなることが知られている．簡単に，例として男女どちらにおいてもそれぞれ新治療と標準治療を比較する状況，つまり男性のみに限って1回比較し，さらに女性のみに限って1回比較するという計2回の検定を行う状況考える．当然ながら1人の対象者が重複して男女どちらの属性ももつということはないため，統計的にはこの2回の検定は互いに独立であるという．独立なk

個の検定をそれぞれの有意水準を5％として行った場合，本当はどちらにおいても差がないのにも関わらず，少なくとも1つ以上を誤って統計的有意であると判断してしまう確率(これを試験全体でαエラーを犯す確率という)は$1-(1-\alpha)^k$により求められる．ここでαは個々の検定で用いる有意水準を表す．前述の$k=2$の状況においては試験全体でαエラーを犯す確率が$1-(1-0.05)^2=9.8\%$となる．つまり，本当は男女ともに新治療と標準治療の間に差がないのにもかかわらず，少なくとも男女のどちらか一方以上で統計的有意差があると誤って判断してしまう確率が，それぞれの検定で設定した有意水準5％の約2倍にあたる約10％であることを意味する．この試験全体でαエラーを犯す確率は，式から求められるとおり，検定の回数kが増えれば増えるほど上昇する．

中間解析における検定の多重性とは，前述した通り，中間解析と最終解析あわせて2回以上の群間比較を同一試験内で行うことから生じる．例として，全生存期間をプライマリーエンドポイントとして新治療と標準治療を比較するランダム化試験を考える．この試験では最終解析時点までに両群あわせて200イベントを観察するように計画されているとする．ここでは全生存期間に興味があるため，サンプルサイズではなくイベント数を単位とする．ログランク検定やCox回帰に代表される生存時間解析ではサンプルサイズにかかわらずイベント数が大きいほど一般により高い検出力が得られる．両群あわせて200イベントが観察された場合，標準治療に比べた新治療のハザード比0.67に対応する群間差に対して片側有意水準を5％とするログランク検定の検出力は約90％となる．ハザード比は群間差を表す指標の1つであり，生存時間解析において一般によく用いられるものである．さて，この例において100イベントが観察された時点で中間解析を1回行うこととする．この中間解析時点は最終解析時点に得られるイベントの100/200＝1/2が得られた時点である．これより，この中間解析は情報時間(information time)0.5の時点で行う解析であると統計学的に表現されることが一般的である．同様にして，最終解析は情報時間1の時点で行う解析であると表現できる．これらの設定の下，中間解析および最終解析どちらにおいても有意水準5％の検定により群間比較を行った場合，本当は差がないのにもかかわらず，少なくとも一方の解析時点で統計的有意と誤って判断してしまう確率(同様に試験全体でαエラーを犯す確率という)を数値計算により求めると約8％となる．独立な検定に対して求めたものに比べ，試験全体でのαエラーを犯す確率の上昇分がやや小さくなっているのは，2つの検定が独立ではなく相関をもっているからである．中間解析の際に解析対象となったイベントは当然ながら最終解析の際にも重複して解析対象となるという点において両者は独立でない．参考として，1試験内で行う中間解析回数ごとの試験全体でのαエラーを犯す確率をコンピュータ・シミュレーションにより求めたものを表1に示す．ここでは中間解析はすべて等間隔で行い，中間解析も最終解析もすべて有意水準5％の検定を用いて群間比較を行うという設定の下，それぞれ10,000,000個の疑似臨床試験データをコンピュータ上で発生させることによって数値的に求めた．

表1　中間解析を等間隔に行った場合の実施回数と試験全体でαエラーを犯す確率

中間解析実施回数	αエラーを犯す確率
0	5.0
1	8.0
2	10.1
3	11.7
4	13.0
5	14.1
6	15.0
7	15.8
8	16.5
9	17.2
10	17.8

検定の多重性の問題を調整するために用いる統計手法を一般に多重性調整法と呼ぶ．この多重性調整法では一般に試験全体でαエラーを犯す確率が名目水準以下となるように，個々の検定で用いる有意水準としてこの名目水準より小さめの値を用いる．よく知られる多重性調整法としてボンフェローニ法がある．検定が独立でない状況では，検定の間の相関が強くなるほど

このボンフェローニ法は過度に調整する傾向があるというデメリットをもつ．ただしその調整手順の単純さから得られるメリットとして，多くの状況に対してユニバーサルに用いることができる．ボンフェローニ法では個々の検定で用いる有意水準として，名目上定めた試験全体でαエラーを犯す確率をこれから行おうとする検定の数で割り算したものを用いる．つまり，名目上定めた確率を等分割してそれを個々の検定の有意水準として用いる．たとえば，試験全体でαエラーを犯す確率を0.05に定めた下でこれから10個の検定を行おうとする場合，ボンフェローニ法によって調整された個々の検定の有意水準は0.05/10＝0.005となる．これは個々の検定のp値が0.005以下となった場合に統計的有意，そうでない場合に統計的有意でないと判断することに対応する．検定が独立の場合には前述した式から試験全体でαエラーを犯す確率を求めると$1-(1-0.005)^{10}=0.0489$となり，確かにこれが名目上定めた確率0.05を超えないことが確認できる．行おうとする独立な検定が100個あろうとも，同様に個々の検定の有意水準0.05/100＝0.0005とすれば$1-(1-0.0005)^{100}=0.0488$となり，確かに名目上定めた確率0.05を超えない．ただし，検定が独立でない状況ではボンフェローニ法は必ずしも適切な多重性調整法とならない．一般に検定間の相関が強くなるほど，実際の試験全体でαエラーを犯す確率は過度に小さくなる．この確率を過小になるに従ってβエラーを犯す確率は過大となり，これにより個々の検定における検出力が低下し，さらには群間差の推定における精度が低下してしまう．

前述した通り，中間解析と最終解析の間には相関があるため，ボンフェローニ法などの簡易な多重性調整法は適切でない．当該試験において中間解析を複数回実施しようとする場合には，それら中間解析の間にも相関がある．中間解析を行うにあたっては中間解析に特化した多重性調整法である群逐次解析法(group sequential analysis)を用いることが一般的である．群逐次解析法としてさまざまな方法がこれまでに提案されているが，その中でもよく用いられる方法はα消費関数(alpha-spending function)と呼ばれる関数を利用する方法(これをα消費関数法という)である．ここではα消費関数法の数理的説明は省かせていただくものの，多重性調整の概要はボンフェローニ法の際に述べたものに通ずるものがあり，名目上定めた試験全体でαエラーを犯す確率を中間解析や最終解析などを行う度に分割して用いる方法である．時間が経過するにつれて累積した結果に基づいて中間解析の実施回数は増えていくとともに，中間解析ごとに試験全体でαエラーを犯す確率を小出しに使っていき，あたかも中間解析ごとにαを消費しているようにみえる．これより他の多重性調整法と同様に，個々の検定で用いる有意水準は試験全体でαエラーを犯す確率より小さめの値を用いることになる．ボンフェローニ法と大きく異なる点は，試験全体でαエラーを犯す確率の分割の仕方である．群逐次解析法では，検定ごとに等分割することは必ずしも一般的ではなく，また前述の通り検定が独立ではないことから検定の間にある相関を適切に考慮した上で個々の検定でも用いる有意水準を定める．α消費関数の関数型としては一定の範囲内であれば任意のものを用いることが実際上は可能であるが，がん領域で頻繁に用いられるものはO'Brien-Flemingタイプのα消費関数である．このO'Brien-Flemingタイプのα消費関数は試験早期には早期中止する可能性が非常に低い一方で，試験が進むにつれて早期中止する可能性が徐々に高くなる傾向をもつ．試験早期ほどサンプルサイズも小さく，これより必然的に情報量も少なくなる．そのような試験早期の状況で早期中止する可能性を低くする一方で，情報量がより多くなる最終解析に近い時点であるほど早期中止されやすいことは一般的に臨床的にも受け入れやすい性質であるといえる．もちろん中間解析においても統計的に偶然誤差を加味した上で判断を行うわけであるが，結果の一般化可能性の観点から考えるに情報量の多い結果の方が系統的誤差(偏り，バイアス bias)が少ないと一般的に考えることができるためである．

ほかにもO'Brien-Flemingタイプに並列してテキストに紹介されることが多いα消費関数の関数型としてはPocockタイプがある．しかしながら，これが用いられる事例はがん領域に限らずともきわめて稀である．参考までに，Pocockタ

イプのα消費関数を用いた場合，それまでに累積した情報量によらず等しくαを消費する．つまり，たとえ試験早期であろうとも最終解析であろうとも等しくαを消費するため，O'Brien-Flemingタイプの関数に比べて試験早期であっても早期中止する可能性が高い．

通常用いられる群逐次解析法では，中間解析時点や中間解析回数が結果に依存しないことを前提にして多重性調整を行っていることには注意が必要である．たとえば，1回目の中間解析の結果，見かけ上の群間差が臨床的には十分過ぎるほどに存在していたものの，統計的調整を伴うとp値が有意水準よりもわずかに大きかったため統計的有意とは判断できないという結果を得たとしよう．この場合，研究者としてはこれほどの群間差が存在するならば可能な限り早く試験を中止して結果を公表すべきであるため，次回の第二回中間解析時期の前倒し，あるいは事前に計画されていなかった中間解析の追加を考えたくなるかもしれない．しかしながら，このように実際の結果に依存した中間解析時期の決定，あるいは中間解析の追加を行うとαエラーを犯す確率が名目上定めたものを超えてしまう．つまり，検定の多重性が存在する状況と同様に，本当は差がないのにもかかわらず誤って群間差があると判断してしまう確率を一定範囲内に制御することができなくなってしまう．たしかに本当にこれほどの群間差があるのであれば倫理的にも試験を早期中止すべきであると考えられるかもしれないものの，良かれと思って行ったそのような対応によって結果的にαエラーを犯す確率が高まってしまう危険性を有するのである．後述する独立データモニタリング委員会によって当該試験に関わる研究者とは独立にこのような判断を行おうともこの危険性は避けられない．つまり，このような状況においても決して安易な対応をするべきでなく，その必要性を臨床的観点，統計的観点，倫理的観点を総合して吟味した上で適切な判断を行うべきである．

実施運営の観点からみた中間解析

中間解析の結果として統計的有意ではなく，これにより試験継続となった場合にも，その中間解析の結果を当該試験に関わる研究者が見てしまうと，有形無形を問わず，試験の適切な運営に影響を与えてしまう可能性がある．たとえば，中間解析時に統計的有意でないものの，新治療が標準治療に比べて見かけ上優れているような結果が得られたとする．もしもその結果を当該試験に関わる研究者が見てしまった場合には，中間解析前後で試験に登録される対象者の属性などが大きく異なってしまいうる．このような場合，最終解析の結果が一般性に乏しい結果となってしまうかもしれない．あるいは新治療が一般的にも用いることが可能な治療なのであれば，その試験はいまだ検証的結果を得てはいないのにもかかわらず，中間解析を境に登録なされにくくなってしまうことで試験自体が検証的結果を得られないものになってしまうかもしれない．中間解析の実施運営上，その結果を評価する目的で当該試験に関わる研究者とは独立な委員会(これを独立データモニタリング委員会という，効果安全性評価委員会等と表現されることもある)を設置することが一般的である．通常，中間解析結果は独立データモニタリング委員会の構成メンバーによって詳細に評価，検討され，試験継続に関する結論のみが当該試験に関わる研究者に伝わるような形式をとる．構成メンバーには当該領域に精通した臨床家や統計家らを含むことが望ましいとされる．統計家も当該試験に関わっておらず独立であることが好ましいという意見もあるものの，必ずしも世界的にコンセンサスが得ているわけではない．統計家の独立性に関しては現在でもその良し悪しに関する議論が存在しており，世界的にみると必ずしも独立となっていないのが現状である．

文　献

1) 厚生省医薬安全局審査管理課．臨床試験のための統計的原則 (平成10年11月30日付医薬審第1047号)．1998.
2) World Medical Association. Declaration of Helsinki. Human Participant Protections Education for Research Teams.(In：日本医師会・訳．ヘルシンキ宣言．ヒトを対象とする医学研究の倫理的原則)．2004.