臨床評価における統計学上の論点：医薬品の承認申請および 審査における多重性の取り扱いについて

Regulatory Points to Consider on Multiplicity Issues in Clinical Trials of New Drug Applications

柴田大朗

Taro Shibata

国立がんセンター がん予防・検診研究センター 情報研究部

Statistics and Cancer Control Division, Research Center for Cancer
Prevention and Screening, National Cancer Center

e-mail: tarshiba@gan2.res.ncc.go.jp

In the development and regulatory review of pharmaceuticals, multiplicity is one of the important issues, and an adequate treatment of multiplicity is required by regulatory guidelines. Though the importance of statistically valid adjustment of multiplicity in drug development is nowadays well understood by practitioners, the more is required in the view of government agencies responsible for drug evaluation. To use the drug effectively and safely in clinical practice, not only the validity of the method itself, but also the validity of the formulation of clinical questions and the adequacy of the interpretation of results are required. In this article, two examples of new drug application review results, one by the US FDA and the other by the Japanese regulatory agency, are reviewed and discussed.

Key words: Multiplicity, Clinical Trial, Drug Approval.

1. はじめに

臨床試験に関する多重性については、Lewis et al. (2001) の指摘にもあるように、副次変数が複数ある試験の解釈や、部分集団（サブグループ）解析結果の解釈に関して、薬事関連規制上、明確なルールやコンセンサスがあるわけではなく、ICH E9 ガイドライン「臨床試験のための統計的原則」（厚生省医薬安全局審査管理課（1998））においてもこのような観点からの言及はない。欧州規制当局の Committee for Proprietary Medicinal Products (CPMP; 現 Committee for Medicinal Products for Human Use (CHMP)) の Points to Consider (CPMP (2002)) として多重性の問題が取りあげられている理由もこの辺にあると思われる。

多重性を統計学的に適切に取り扱う方法論については、改めて言及するまでもなく多くの検討がなされてきており、実務レベルにおいても広く応用されている。妥当な方法論が採用され、かつ、結果が期待通りのものであったならば、医薬品の評価は大きな山を越えたものと考えられる方々も少なくないと思われる。しかしながら実務においては、通常、方法論そのものの評価ならび

に結果が統計学的に有意であったか否かの評価に留まらず、得られた臨床試験結果を踏まえ、詳細な議論が行われる。

この問題の取り扱いが難しいのは(かつ、様々な立場から議論することが必要であり、またその意義があると思われるのは)、解釈やそこから派生する薬事関連規制上の取り扱いが、疾患ならびに当該医薬品・医療機器の特徴、医学・薬学に関する科学的な水準、医療・社会環境等の context に大きく依存するためでもある。本稿では、(1) 多重性を調整し統計学的に有意な結果を得つつもその結果が採用されなかった事例 (erlotinib の事例)、(2) 多重性の調整を行っていない結果の解釈に関する議論が公表されている事例 (インジセトロン事例)、各々を取りあげ、公表論文や本邦ならびに米国における医薬品承認審査に関する公表資料をもとに関与された当事者の解釈を推察しつつ紹介する。

なお、本稿では承認審査に関する公表資料を引用し多重性の取り扱いに関する論点を取りあげて議論を進めることとするが、当該申請資料作成に関与された多くの方々のご尽力を否定する意図は無いこと、また、医薬品そのものの総合的な評価を行うものではないことについてご理解頂きたい。

2. Erlotinib の事例

Erlotinib (Tarceva) は 2004 年 11 月に米国で非小細胞肺癌に対して承認された HER1/EGFR 阻害薬である (Cohen et al. (2005))。Shepherd et al. (2005) によると、前化学療法が無効であった患者に対し、本剤はプラセボと比較して生存期間を延長している (生存期間中央値 erlotinib 群 6.7 か月、プラセボ群 4.7 か月群、ハザード比およびその 95% 信頼区間 0.70 [0.58-0.85])。

本剤は、前年に本邦における承認後に間質性肺炎の副作用が社会問題化した動向を十分に知りながらも第 II 相試験の結果に基づき 2003 年に米国 Food and Drug Administration (FDA) が迅速承認したゲフィチニブ (イレッサ) に類した作用機序が期待されるものであり、Shepherd らは生存期間の延長だけでなく咳、呼吸困難、疼痛に関連する QOL の向上についても報告している。ゲフィチニブの QOL や症状緩和に関わる効果については、2002 年 9 月 24 日に開催された Oncologic Drug Advisory Committee (ODAC) において FDA 側が臨床試験における評価に関する方法論上の問題から否定的な見解を提示しているが (<http://www.fda.gov/ohrms/dockets/ac/cder02.htm>)、探索的な検討の範疇を出ないとはいえ、Shepherd らの報告の背景にはこういった副次的な効果に対する期待がある状況であった。結果として、Shepherd らは事前に QOL の項目を 3 つに絞り Hochberg の方法 (Hochberg (1988)) により多重性を調整した解析を行ったところ、状態が悪化するまでの期間に統計学的に有意な差が見られたと報告しているが (Shepherd et al. (2005))、これに対して FDA は本剤投与に伴う QOL に関して、申請者から提示された結果は添付文書に反映させるに足るものではないとの見解を示している。

Shepherd らの報告を見る限り、生存期間と QOL という複数のエンドポイントに関する評価に関わる多重性は調整していないものの、QOL の解析としては、数多くある項目のうち咳、呼吸困難、疼痛の 3 項目を予め選択し、かつこれらの多項目間の多重性については調整されていることから、薬剤の特徴を明らかにするという観点からの主張であれば、一見大きな問題はなさそうに

思われる。

この件について、前述の FDA の報告 (Cohen et al. (2005)) では根拠が詳細に示されていないが、Web 上で公表されている review report (FDA (2004)) では詳細な説明が加えられている。Review report において FDA の審査官は、申請者側の解析に一部ミスがあるもののそれ自体は解析結果を大きく変えるものではないことを指摘した上で、引き続き次のような評価を下している。

- The primary endpoints and the method of QoL analysis was specified differently in the statistical analysis plan which was finalized on June 11, 2003, after all patients had been entered on the study. Given that this was a secondary endpoint and the analysis plan was finalized after the completion of accrual to the study (last patient entered on January 31, 2003) analysis of QoL data can only be considered as exploratory. [p11]
- [T]his reviewer performed an exploratory analyses using the ANCOVA model with baseline score as covariate for the change scores from baseline to Week 4 and Week 8. [...] [T]hese results are not as significant as the time to deterioration analyses results for these 3 symptoms. [p32]
- This reviewer further conducted time to deterioration analysis for other [...] symptoms/domains that were measured in QoL questionnaire. [...] It is to be noted that [...] ‘physical functional domain’, and ‘global QoL scale’ appear to be worse in the erlotinib group compared to placebo (HR > 1). [...] Furthermore, the dyspnea symptom domain was not significant. This raises doubts about the robustness of the significant findings in the time to deterioration of cough, dyspnea and pain. [p32]

すなわち、FDA は本件に関して①データ解析そのものの整合性、②手続き上の問題点の有無を確認した上で、③異なる方法で解析を行った場合に同様の結果が得られるか否か、④取りあげられている変数を“取り巻く”他の変数の動きが如何なるものであるか、さらに、⑤取りあげられている変数を異なる角度から眺めたときに同様の結果が得られているか否か、等を評価している。多重性の統計学的な取り扱いそのものが一見妥当であったとしても、また、本事例の場合はその結果が New England Journal of Medicine という peer reviewed journal に掲載されたものであるが、医薬品の承認審査の過程でその結果の頑健性や一貫性が保たれているか否かが問題として取りあげられ、議論されること自体は至極当然である。一方、(プロトコルに規定されていなかったことも同時に考慮されているとはいえ) 副次変数であることをもって、探索的な位置づけとして取り扱うというこの件に対する FDA の割り切りについては、実務的な観点はさておき、科学的な観点から疑問視される向きもあるかもしれない。この点については後述する。

3. インジセトロンの事例

インジセロン (シンセロン) は 2004 年 1 月に本邦で承認された、抗悪性腫瘍剤 (シスプラチン等) 投与に伴う消化器症状 (悪心、嘔吐) を効能・効果とする 5-HT₃ 受容体拮抗型制吐剤である。本剤の添付文書、申請者によって作成された申請資料概要、国立医薬品食品衛生研究所 医薬品医療機器審査センター (以下、審査センター; 現 独立行政法人医薬品医療機器総合機構) によ

り作成された審査報告書（審査センター（2003））はいずれも医薬品医療機器総合機構の Web で公開されている（<http://www.info.pmda.go.jp>）。審査の過程では、抗がん剤投与後の連日投与に関わる意義、ならびに作用機序（5-HT₄ 受容体拮抗作用の意義）等について議論されている。

連日投与に関わる意義については、申請者によって作成される申請資料概要によると、抗がん剤投与日以降の連日投与に関する有用性を検証する試験として 157 例を対象にプラセボ対照二重盲検並行群間比較試験が行われた結果、主要評価項目で群間差が見られなかったものの、①シスプラチン投与量 75 mg/m² 未満の患者集団における解析〔部分集団解析〕、②QOL 調査票のスコアに関する検定を悪心・嘔吐の有無に関する二値変数に対する検定に変えた解析〔変数の切り分けの変更、検定方法の変更〕、③悪心あるいは悪心・嘔吐ありの症例の割合を 1 日ごとに群間比較した解析〔経時データの（解析対象集団が時点毎に異なる）多時点・多項目にわたる解析〕で統計学的に有意な群間差があることを提示し、連日投与に関する有用性が主張されている。

これに対し、審査センターの審査報告書（p42）では、

- 本解析の試験計画上の位置付けを説明するよう求めたところ、これらの解析は開鍵後に実施された事後的な解析であること、探索的な要素が強いことから多重性の調整は行っていないことが説明された。審査センターは、これらの解析結果が、あくまで開鍵後に探索的な意味合いで行われた解析であること、また、欠測のために評価日毎に集計対象となっている患者数が異なっており、欠測という現象と QOL 評価との関連性を無視することが困難であることにも鑑み、評価日毎の群間比較結果をもって本剤の有用性を主張することは適当ではないと判断している。
- また、他の 5-HT₃ 受容体拮抗薬においても化学療法開始後 24 時間以内の急性期の悪心・嘔吐に対する有効性は確立しているものの、24 時間以降の遅発性の悪心・嘔吐に対する 5-HT₃ 受容体拮抗薬単剤の有効性に関する報告は様々であり、併用療法の有効性についても議論が分かれるところである（N Engl J Med 332; 1, 1995）（Clin Oncol 15; 124, 1997）。
- 以上のことから、審査センターは抗悪性腫瘍剤投与終了後、翌日以降に見られる悪心・嘔吐に対する本薬の有効性は明らかとは言えず、抗悪性腫瘍剤投与日以外に本薬を投与する臨床的意義は不明であると考えている。

とされている。

一方、5-HT₃ 受容体拮抗型制吐剤であるオンダンセトロンを対照とした実薬対照比較試験の結果から本剤の作用機序に関わる主張（5-HT₄ 受容体拮抗作用を併せ持つ意義）を行う申請者に対する審査センターの見解は、審査報告書（p35）に以下のように記されている。

- 審査センターは、実薬対照比較試験 [...] において、投与後 6 時間後、投与後 18 時間後の評価では差が見られていないものの、投与後 12 時間後の有効度評価で本薬群と対照群との間に有意差が見られている点を本薬が 5-HT₄ 受容体拮抗作用を併せ持つことによる有用性の一つとして主張していることに関して、この差に合理的背景があるか否かについて申請者に説明を求めた。 [...]
- 審査センターは、実薬対照比較試験 [...] の結果より本薬は、急性期（シスプラチン投与 24 時間後）の悪心・嘔吐に対する有効性に関しては、オンダンセトロンとほぼ同等（非劣性）であると考えているが、12 時間後までの悪心・嘔吐に対する有効性に関しては、実薬対照比較試験

[...] が 12 時間後までの悪心・嘔吐に対する有効性を主要評価項目として行われた試験でないこと、患者における薬物動態の観点からは投与後 9 時間にわたって 5-HT₃ 受容体及び 5-HT₄ 受容体への作用が期待される血漿中濃度が維持されると想定されながら 6 時間後までの評価で差が見られない理由が明確ではないこと、非臨床において本薬の 0~12 時間の嘔吐抑制効果はオンダンセトロンと同様であること、シスプラチンを高用量投与された患者に限るとどの時点でも本剤群での有効率が高いとの主張がなされている一方で、シスプラチンを低用量投与された患者においては対照薬において高い有効性が見られている時点もあること、更に、6 時間後の評価で差が見られなかった理由として悪心又は嘔吐の発現がシスプラチン投与後直ちに発現するわけではないとの説明については、第 II 相試験においては 6 時間後の時点で既に最低用量群と 8 mg 群との間に有意な差を検出し得ていることから、本薬が 5-HT₄ 受容体拮抗作用を有していることによりオンダンセトロンよりも優れているとする申請者の主張の根拠は不明確であると考えている。

すなわち、オンダンセトロンとの非劣性が証明されていることは認められるものの、本剤の新たな特徴であるとされている 5-HT₄ 受容体拮抗作用については、多時点間の比較のうち一部分だけが抜き出されて評価されているが、①これが予め計画されたものではないこと、②5-HT₄ 受容体拮抗作用による効果発現が期待される時期に関する薬物動態学的観点からの考察と整合性がとれていないこと、③非臨床試験結果との整合性がとれていないこと、④部分集団間での傾向に乖離が見られること、⑤先行試験結果との比較において解釈の一貫性が損なわれていること、などから主張の根拠が不明確であり、更に、⑥24 時間より短い時点での評価に関する臨床的な意義 (p38, 引用略)、⑦類薬も含めた 5-HT₄ 受容体に関する薬理学的な考察 (p38, 引用略) 等も踏まえ、申請者の主張が受け入れられない旨の結論が導かれている。

ここでは、新薬の新規性や画期性をどのように示すか、仮に適切な試験デザインなり開発戦略が決められていたとして、現実の試験結果が主張を裏付けるに十分であるか・妥当なものであるかが問題となっている。指摘しておきたいことは、個別の論点を見る際に、多重性の調整がなされていないことは問題ではあるが、単に多重性の調整をしていないことのみをもって医薬品の承認申請を行う側（企業側）の主張が否定されているわけではないということである。

4. 考 察

多重性の調整のなされている事例、そうでない事例を 1 つずつ紹介した。ここで取りあげた事例は、単に形式的な多重性の調整の有無だけが議論の対象になっているわけではなく、その評価にあたっては、薬理学的、医学的検討に加え、様々な角度からの統計学的検討も行われている。このような検討には、大きく分けて二つの論点がある。一つは実際の臨床上の問題を適切に定式化し得ているかという点、もう一つは得られた結果を臨床現場に提供するという観点から適切に解釈し得ているかという点である。統計学の役割を狭義にとらえると、定式化された問題に対して解析を実施し、結果を得るというプロセスがその根幹をなすことは間違いがないが、実務レベルでは、その前段階（定式化）および後段階（解釈）の部分を、他の領域の専門家と共に吟味する必要がある。本稿で取りあげた 2 つの事例は、いずれも多重性という問題に対して、狭義の統計学

的問題として議論が終わっているわけではないことを指摘しておきたい。

なお、結果の解釈については context に大きく依存するものであり、ケースバイケースであると言ってしまうとそれまでであるが、事例を通して考えておきたい事項は、副次変数に対する多重性の調整がどの程度まで必要で、どこまで厳密に行わなければならないかという議論は、解析結果がどのような目的で使われるかに影響されるということである。これについては、効能・効果として示したい場合、類薬との差別化の根拠としたい場合、プロモーションに使いたい場合、探索的な解析として役立てたい場合等、状況に応じてどこまで厳密さを求めるかが変わり得る。例えば、erlotinib の例で、仮に生命予後の改善あるいは生命予後の改善に対するエビデンスの蓄積状況が先行薬と全く変わりが無いという状況にあれば、企業としてはより積極的に QOL 等の副次変数に基づく評価を主張したいという思惑が働くかも知れない。その場合は、FDA の評価も前述のようなあっさりしたものではなく、申請者との間でより深い議論が行われ、場合によっては ODAC の席上で論点として取りあげられることと想像される。

もちろん、開発の過程・審査の過程でこのような議論を突き詰めていくと、臨床試験の対象や対照群の設定等の試験デザイン、或いは開発戦略そのものの見直しが必要となる場合もあり得ることから、結果を見た上でどう解釈するかではなく、計画段階に徹底的な議論を行うべき問題であろう。生物統計学の立場からは、既存薬や類薬の開発状況および医療現場の unmet need に関する現存の状況ならびに近い将来の動向を把握し、他の分野の専門家と検証すべき臨床的仮説、開発の各段階で得ておくべきデータの質と量を慎重に吟味し、これらを踏まえた開発の方向性を十分くみ取った上で問題を定式化し、試験デザインに反映させるという一連のプロセス、さらに開発の過程において規制当局と企業との間で行われる議論のプロセスでの積極的な関与が可能であると思われる。

一方、このような問題は程度の差こそあれ、ある一面では企業と規制当局との間の交渉の問題とも考えられる。このように問題を捉える場合の問題解決の方向性としては、例えば 2005 年 4 月に発出された FDA の draft guidance (FDA (2005)) のようなものを対外的に提示していくアプローチと、開発の過程で実施される規制当局との議論の中で見解をすりあわせていくアプローチがあり得る。ただし、前者だけでは医薬品の開発にとって本質的な問題とも言える個別性・特殊事情に対応しにくく、一方、後者だけでは判断の一般化が図りにくい上に、ともすれば“規制当局と企業との癒着”が取り沙汰され得る社会的状況において、手続き上の不透明性が医薬品の評価あるいは医薬品そのものに対する社会からの価値判断にバイアスをかけてしまう危険性がある。両者のバランスをとりつつ、既存事例の吟味を通じたノウハウの蓄積、近い将来問題になり得る事項について公開の場で議論することによるコンセンサスの形成、更に（人的リソースの不足している現状では生物統計学に限らずほぼ全ての領域で実現不可能かもしれないが）規制当局に蓄積した経験から得られる知見の公表・社会へのフィードバックを行うことによる情報の共有等を積極的に進めていく必要がある。

最近では、審査報告書・審議結果通知書等がインターネット上で公表されるようになり、また、FDA の Advisory Committee 等の資料については最終的に承認されなかった医薬品についても情報を入手することもできる。更に、FDA の review report は分量が多く全てに目を通すことは容

易でないが、中には peer reviewed journal に公表された医学論文の内容を補うものもあり、例えば、一流医学雑誌に有意差を示した結果が掲載されていながら承認されていない医薬品がある状況を取りあげて、これを即規制当局ないし企業の怠慢とすることが、場合によっては危険な議論であると想像することも出来る。また、FDA は特にがんの領域においては審査の内容を論文としてまとめて公表することを積極的に行っており、そのような報告の中で、医薬品の有効性および安全性に影響を与える可能性の要因でありながら既存の公表論文においては取りあげられていない探索的な解析結果が提示されることもある (Johnson et al. (2005))。このような公開情報をもとに、生物統計学的問題に留まらず、医薬品の評価全般について広く社会の理解を得ることは、甚だ迂遠ではあるが、画期的な新薬を安全かつ効率的に世に出していくために必要なのではないかと思われる。

本稿では詳細に取りあげなかったが、部分集団解析結果の扱いについて触れておきたい。Lewis et al. (2001) においても言及されているように、全体で有意な差が証明出来た場合であっても医薬品の承認範囲が部分集団に限定されることはあり得る。このような部分集団の扱いは、今後、Pharmacogenomics (PGx) の応用や、国際共同試験の事例が増えるにつれ、より議論が必要となることは間違いない。

PGx の応用としては、いわゆる enrichment による臨床試験の効率の向上が注目されることが多いようだが、試験結果の一般化可能性や特定の因子を持たない患者集団における有効性・安全性データの蓄積を医薬品評価のどの段階で行うべきかといった観点から、十分な検討を行った上で試験デザインを採用することが必要となる。開発の段階では、特定の因子の有無で対象を絞り込むのではなく、いずれの集団も試験対象として含め効果を確認する enrich しないアプローチを採用することの意義は大きく (宇山 (2005))、そのようなアプローチを行う場合には、部分集団に対する多重性の調整について必要性を慎重に検討することになる。その際、生物学的な仮説に基づいた・生物学的に意味のある部分集団における治療効果の検証 (大橋 (2005)) をどう定式化するか、そのために必要となる臨床開発早期における仮説の絞り込みに役に立つ試験計画・データ収集をどう行うか、といった点が今後の課題となるのだろう。多重性を調整した上での生物学的なマーカーを用いた投与対象の絞り込みを伴う仮説検証のアプローチについては、Simon (2005) や Freidlin and Simon (2005) の研究などがある。

なお、事例として紹介した erlotinib やゲフィチニブについては、EGFR 遺伝子変異・発現の状況が予測因子 (predictive factor) であろうと推測されているが、FDA は現時点では erlotinib の投与対象を当該部分集団に絞っていない。また FDA は、2 つの第 III 相試験 (無増悪生存期間をエンドポイントとした erlotinib とプラセボとの比較試験、ならびに生存期間をエンドポイントとした erlotinib と pemetrexed あるいはドセタキセルとの比較試験) を、いずれも全例で EGFR 発現の有無を測定した上で enrich せずに試験を実施し、EGFR 発現の有無それぞれの部分集団での解析についても検討する計画の臨床試験の実施を企業に求めている (Cohen et al. (2005))。この結果は FDA の Web (Post Marketing Commitments, <http://www.accessdata.fda.gov/scripts/cder/pmc/index.cfm>) によると 2008 年末に得られる予定とされていることから、そう遠くない時期に多重性の調整の要否も含め、どのような仮説を置くべきか、結果をどう解釈すべきか、また、実際にどのように結果を解

積したかが明らかになるが、本邦においても並行して、同様の問題に関する生物統計学的な観点ならびに行政的な観点からの検討を進めておくことが重要であろう。

次に、藤越氏の問題提起（藤越（2005））に対する私見を記しておきたい。藤越氏は α の調整が不要な事例を、(a) 探索的な解析と見なせる場合、(b) すべての仮説に対して検証が必要な場合、(c) 仮説間に階層構造が規定されている場合に分類し、6 つの問題提起を行っている。

- ①(b) や (c) の時にサンプルサイズ設定をどのように行うべきか：仮説間の相関が特定出来ない場合であっても、少なくとも独立であると見なした場合の検出力は予め確認し、開発のリスクを把握するためにも見込みを明らかにしておくべきと考える。また、パターン (c) の場合には、医薬品として存在意義を示すために最低限証明出来なければならない仮説と付加的な仮説との区別をし、少なくとも前者については十分な検出力を担保することが重要と思われる。そのためには、既存の医療現場での unmet need を把握し、それと仮説との関係を理解する必要がある。
- ②(c) の場合の階層構造に整合した信頼区間の要否について：厳密な計算が可能な場合にそれを提示することには一定の意義があると考えますが、実務の現場では得られた信頼区間の解釈が容易であるケースは少ないようにも思われる。厳密なものを提示する場合には、生物統計の専門家以外にもその意味が伝わるような配慮が期待される。
- ③三群比較試験の 3 つの目的（プラセボに対する被験薬の優越性、プラセボに対する標準薬の優越性、標準薬に対する被験薬の非劣性）が必須となる / 必須とならないのはどのような場合か：医薬品に有効性があることを示し、かつその有効性の程度が標準薬との相対的な関係において十分なものであること（非劣性であること）を示すという目的で三群比較試験が実施されるのであれば、3 つの目的が必須になると考える。もちろんその際の優越性・非劣性を示し得たと主張する根拠となる仮説の設定（非劣性マージンの設定等を含む）に関してバリエーションはあるだろう。一方、三群比較試験とは言っても、有効性の証明のみを行えば良いという割り切りの元で実施される試験があるのも事実である。日本においてはほとんど見受けられないが、標準薬との相対的な関係に関して検証的なレベルの証拠の提示を目指していないケースである。いずれにせよ、どのようなデザインを採用するべきかという議論と、どのようなものを医薬品として社会が許容するかという議論とは、独立に議論することは可能であるが、実質的には切り離して取りあげることは困難であるように思われる。
- ④Global test strategies の指す内容について：個別薬剤の用量を複数用いた併用療法の要因試験を実施する際には目的が二つある。すなわち、単剤使用に対する併用使用の有効性を示す目的と、用量の組み合わせを選択する目的とである。前者においては ICH E4 に「症例数は、対比較において単一のセルを相互に区別するに十分なほどには大きい必要はない。なぜなら、それぞれ単剤で用いた場合および併用した場合の用量-反応関係、すなわち用量-反応曲面を導くために全てのデータが使用できるからである。」と記されているように、個別の用量の組み合わせ間での比較は必ずしも必要ない。Points to Consider の記述はこれを指しているものと思われる。
- ⑤複雑な用量反応試験において検証すべき仮説が複数存在する場合に仮説構造をどのように設定して多重性を制御すべきか：仮説構造の設定そのものが高度に個別性の高い問題であり全てのケースに適用出来る方針を提示することは難しいと思われる。これまでに述べてきたことの繰

り返しになるが，unmet medical need と開発している医薬品に期待される性質，さらにそのような医薬品が満たすべき条件との関係を整理した上で問題を統計学的な仮説に翻訳するという方針をとることになるのではないだろうか．

- ⑥群間の比較が可能な情報を用いた盲検化レビューとはどのようなもので，それを当局は肯定しているのか：Points to Consider が想定している方法がどのようなものかは不明であるが，群間の比較が可能な情報を用いて解析方法を選択することは，基本的には許容されないものと思われる．仮に手続き上盲検下でレビューを行った場合であっても，特定の治療と関連することが既知の事象（副作用等）の状況から解析担当者の意図の有無はさておき，結果として“都合の良い”解析方法を選択してしまう可能性はあるかもしれない．解析方法の選択を盲検下でレビューに基づき行う場合であっても，注意は必要と思われる．

医薬品の評価にあたって多重性の問題が厳密に議論されるのは，第一義的には医薬品に関する規制があるためである．しかしながら，その規制が目指すところは，医薬品が有効かつ安全に使用される社会的な仕組みを用意することであるとも考えられる．規制が設けられるに至った社会的・歴史的背景を振り返ると，多重性の問題は，調整の要否という単純な二値判断の問題であることに留まらず，患者や医療関係者が生命に関わる意思決定を行う際の根拠となる臨床試験の解析結果を如何に提示することが望ましいかという問題でもある．このような視点に立った議論が生物統計学の専門家と他の領域の専門家との間で共有されることが，新規性の高い医薬品の開発・適正使用の更なる促進につながることを期待したい．

謝 辞

本稿の内容について示唆に富むご意見を頂きました独立行政法人医薬品医療機器総合機構新薬審査第一部部長 森和彦氏に深く感謝いたします．

参 考 文 献

- Cohen, M. H., Johnson, J. R., Chen, Y.-F., Sridhara, R., and Pazdur, R. (2005). FDA Drug Approval Summary: Erlotinib (Tarceva(R)) Tablets. *The Oncologist*, **10**: 461-466.
- Committee for Proprietary Medicinal Products (CPMP). (2002). Points to consider on multiplicity issues in clinical trials, CPMP/EWP/908/99, September **19**: 2002, [<http://www.emea.eu.int/pdfs/human/ewp/090899en.pdf>]
- Freidlin, B. and Simon, R. (2005). Adaptive signature design: An adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients, *Clinical Cancer Research*, **11**(21): 7872-7878.
- Food and Drug Administration (2004). Statistical Review and Evaluation, NDA/Serial Number: 21-743/N000, October 1, 2004, [http://www.fda.gov/cder/foi/nda/2004/21-743_Tarceva_StatR.PDF]
- Food and Drug Administration (2005). Draft guidance for industry: Clinical trial endpoints for the approval of cancer drugs and biologics, April 2005,

[<http://www.fda.gov/cber/gdlns/clintrialend.htm>]

藤越慎治 (2005) Points to Consider on Multiplicity Issues in Clinical Trials, 2005 年計量生物セミナー「臨床評価における統計学上の論点」抄録集, 41-49, 2005 年 9 月 21-22 日.

Hochberg, Y. (1988) A sharper Bonferroni's procedure for multiple tests of significance, *Biometrika*, **75**(4): 800-3.

Johnson, J. R., Cohen, M., Sridhara, R., Chen, Y.-F., Williams, G. M., Duan, J., Gobburu, J., Booth, B., Benson, K., Leighton, J., Hsieh, L. S., Chidambaram, N., Zimmerman, P., Pazdur, R. (2005). Approval Summary for Erlotinib for Treatment of Patients with Locally Advanced or Metastatic Non-Small Cell Lung Cancer after Failure of at Least One Prior Chemotherapy Regimen, *Clinical Cancer Research*, **11**(18): 6414-6421.

国立医薬品食品衛生研究所 医薬品医療機器審査センター (2003). インジセトロン審査報告書, 衛研発 3723 号, 平成 15 年 11 月 6 日,

[<http://www.info.pmda.go.jp/shinyaku/g0401.html>]

厚生省医薬安全局審査管理課 (1998). 「臨床試験のための統計的原則」について, 医薬審 1047 号, 平成 10 年 11 月 30 日,

[<http://www.nihs.go.jp/dig/ich/eindex.html>]

Lewis, J., Louv, W., Rockhold, F. and Sato, T. (2001). The Impact of the International Guideline Entitled Statistical Principles for Clinical Trials (ICH E9). *Statistics in Medicine*, **20**(17-18): 2549-2560.

大橋靖雄 (2005). 個別化医療: 統計的視点. *Molecular Medicine* **42**(8): 912-917.

Shepherd, F. A., Rodrigues Pereira, J., Ciuleanu, T., Tan, E. H., Hirsh, V., Thongprasert, S., Campos, D., Maoleekoonpiroj, S., Smylie, M., Martins, R., van Kooten, M., Dediu, M., Findlay, B., Tu, D., Johnston, D., Bezjak, A., Clark, G., Santabarbara, P., Seymour, L. (2005). Erlotinib in previously treated non-small-cell lung cancer. *New England Journal of Medicine* **353**(2): 123-32.

Simon, R. (2005). Roadmap for Developing and Validating Therapeutically Relevant Genomic Classifiers, *Journal of Clinical Oncology*, **23**(29): 7332-7341.

宇山佳明 (2005). ゲノム薬理学と医薬品承認審査. *Cancer Frontier*, **7**: 195-205.