



効率的な医薬品開発のための統計的方法の研究

濱崎 俊光 (はまさき としみつ)

大阪大学大学院医学系研究科・医学統計学 准教授

スライド-1

第19回ヘルスリサーチフォーラム
千代田放送会館 2012年11月10日(土)

複雑性を増す医薬品開発における 効率的な試験デザインの選択

効率的な医薬品開発のための統計的方法の研究

濱崎俊光
大阪大学大学院医学系研究科・医学統計学

共同研究者
Scott Evans, PhD
Harvard School of Public Health
Guosheng Yin, PhD
University of Hong Kong/MD Anderson Cancer Center

スライド-2

1. Introduction Background and Objectives

『現在の医薬品開発は複雑化・多様化している』

- 医薬品の効果は多面的であり、これまでの主要なエンドポイントの一つに限定する仕方では、医薬品の効果を十分に評価することはできない。
- 複数のエンドポイントを主要変数(Co-primary Endpoints)として設定し、医薬品の評価を行う試みがある。

- アルツハイマー病(CHMP Guideline, 2008)
Cognitive Endpoint: ADAS-cog
Functional Endpoint: ADL, CIBIC-plus, ADCS-CGIS
- 過敏性腸症候群(IRS) (FDA Draft Guidance, 2010)
IBS with constipation: Pain intensity and Stool frequency
IBS with diarrhea: Pain intensity and Stool Consistency
- その他の疾患領域: Oncology, Prevention of Mother-to-Child HIV/Hepatitis B Transmission, Cardiovascular Disease Therapy

**Guidance for Industry
Irritable Bowel Syndrome —
Clinical Evaluation of
Products for Treatment**

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only. Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the Federal Register of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (2FF-015), Food and Drug Administration, 1015 Wilson Lane, 5th, Silver Spring, MD 20910. All comments should be identified with the draft number listed in the notice of availability for publication in the Federal Register.

For questions regarding this draft document contact Rop/76 at 301-796-0912 or Ann Marie Tomaszewski at 301-796-0914.

The 19th Health Research Forum, November 10, 2012

【スライド-1】

本研究はHarvard UniversityのScott Evans先生とUniversity of Hong Kong及びMD Anderson Cancer CenterのGuosheng Yin先生と行いました。

【スライド-2】

皆さんご存知のように、現在、医薬品の開発の効率は大変低下しています。そのため、医薬品開発の成功率を向上させるために諸種のとり組みが行われ、医薬品の開発は大変複雑化・多様化しています。現在一つの取り組みとして行われているのが、「医薬品の効果というのは多面的であり、これを今までのように一つのエンドポイントに基づく評価は十分でない、したがって、複数のエンドポイントを設けて医薬品を評価すべきだ」という議論です。例えばアルツハイマー病において評価は2つのエンドポイント(例えば、認知機能とQoL)で行うことがガイドライン等で議論されていますし、過敏性腸症候群では、米国FDA (Food and Drug Administration)が最近ガイダンス案を作成しており、その中では痛みと便の数(便数)を評価項目として吟味することが議論されています。

【スライド-3】

最近の医薬品の承認申請の流れでは、このように2つ(以上)のエンドポイントを置いたときに、いずれか1つのエンドポイントのみで有意性を評価すればよいわけではなく、エンドポイントのすべてで有効性を示すことが求められています。

このような複数のエンドポイントを取り扱うときの症例数設計を考えた場合、例えば1つのエンドポイントに基づいて症例数を決定した場合、その症例数のもとで2つのエンドポイントでともに有効性が示せるかどうかという、差があるのに差を見いだせない誤りの確率は大きくなります。エンドポイントの数が増えるに伴いその確率は大きくなり、ときには80%の確率で誤って評価しています。

こういった誤りの確率を制御するためには症例数を大きく必要がありますが、例えば1つのエンドポイントから2つのエンドポイントにした場合には、症例数は1.2倍から1.3倍、5つのエンドポイントを置いた場合には1.5倍から1.7倍というように、症例数は大変に大きくなってしまいます。これでは実施可能性が大変低くなり、実施できたとしても試験期間が延長し費用が増大するといった問題があります。

【スライド-4】

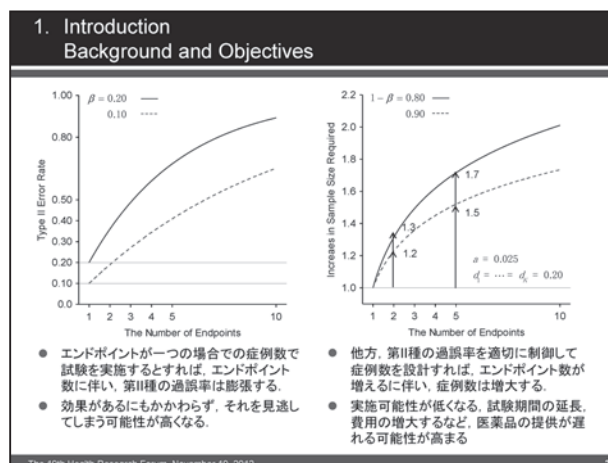
そこで今回の研究では、症例数をいかに縮小させるかを研究しましたが、単に症例数を縮小させるのではなく、情報量を確実に担保しながら症例数を縮小させることに取り組みました。

一つは、エンドポイントを2つ以上設定した場合には、当然それらのエンドポイント間には必ず相関が存在します。したがって、その相関をうまく症例数の計算に活かして症例数を縮小しようという試みです。もう一つは、最近FDA等で議論されているのが、試験の途中で評価を行って、そこで計画時の仮定が適切かどうかを一旦見直して、必要に応じて計画を変更し継続するという試みです。

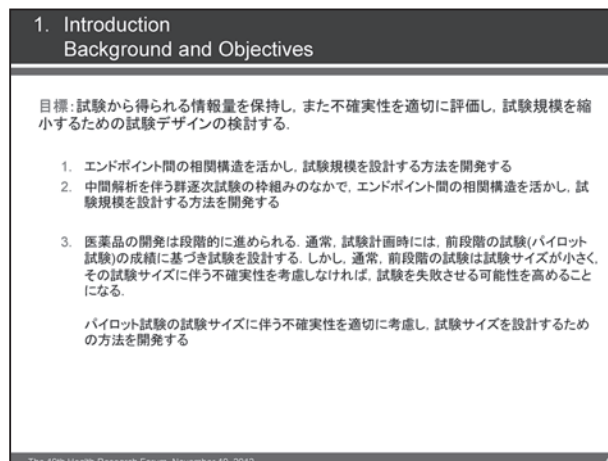
今回は、その2つの観点から、研究を行いました。

さらには、試験を計画する際には、外部の公表データを使ったり、あるいは該当する医薬品のPOC (Proof of Concept) 試験と呼ばれるような先行するデータを使うことがあります。しかし、その外部試験やPOCの試験規模は一般的にそれほど小さくなく、そういった比較的、規模の小さい試験からのデータに基づいて症例数を設計すると、失敗してしまう可能性

スライド-3



スライド-4

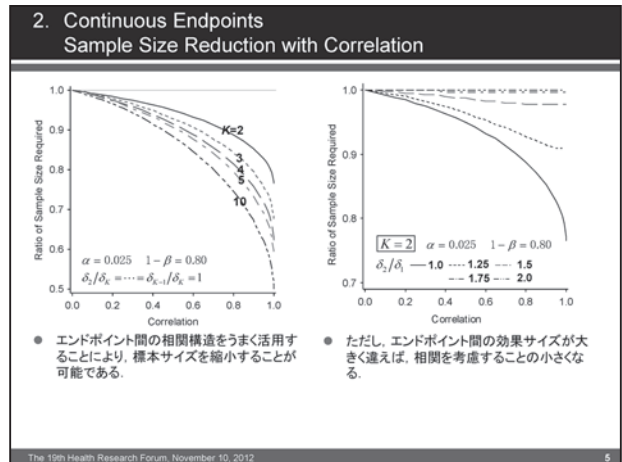


が高まることが知られています。そこで、それを改善するための方法を研究しました。

【スライド-5】

実際に症例数の設計に相関を入れて考慮した場合にどうなるかということですが、適切に相関を考慮することによって、例えばエンドポイント間に相関が0.5ぐらいあると10%程度要求されたものよりも小さくなることが分かりました。あるいは極端に相関が強ければ、例えばエンドポイントを10個置いていた場合には、ほぼ半分ぐらいに小さくすることが分かりました。ただしここで問題があるのが、2つのエンドポイント間のいわゆる効果サイズというのは、片方が大きくて片方が小さい場合には、ほとんど相関を考慮しても、実は標本サイズがほとんど下がらないという問題です。

スライド-5



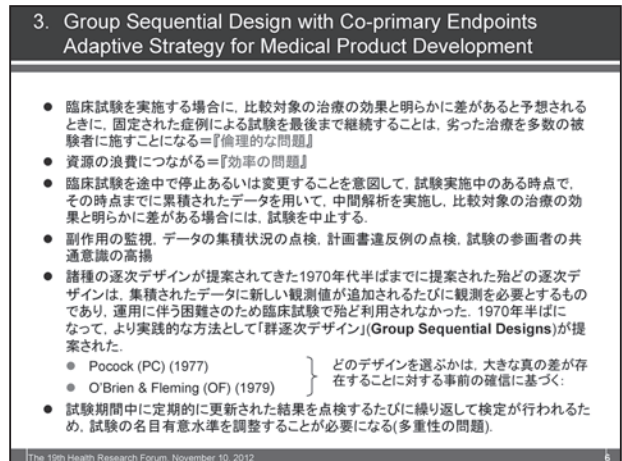
【スライド-6】

さらに症例数の縮小するために何を考えたかということ、試験の途中でいわゆる停止あるいは変更を意図して、試験途中で治療割り付けを明らかにし解析を行い、計画を見直して次の段階に移るといったことです。いわゆる中間解析を試験計画に盛り込んで、さらにエンドポイント間の相関を考慮することによって症例数を縮小させようと考えました。

当然、途中で見直すことによって症例数を小さくできるということは、いわゆる倫理的な問題あるいは効率的な問題を改善することにつながります

これまでの研究としては Pocock 型デザインと O'Brien & Fleming 型デザインのなかで、1つのエンドポイントの場合について議論されてきました。Pocock 型デザインというのは、「かなりの高確率で有効性を示すことができる」ということが試験計画時に確信を持っている、早く試験を止めたいという場合に用いられる方法で、O'Brien & Fleming 型デザインの場合は、「あまり確信度がないので、できれば後半で試験をストップしたい」ということを考えて行われているデザインです。

スライド-6



【スライド-7】

こういったデザインの推進に関してはFDAにおいても議論されて、最近ガイダンス案も公表されています。

【スライド-8】

実際に相関を考慮することによってどの程度まで症例数を縮小できるかを、このグラフは示しています。これは実際に解析を10回まで行った場合に、症例数がどう変化していくかを見たものです。

10回を計画していた場合に10回まで、つまり途中で実際に止めずに試験を行ったということを前提としています。O'Brien & Fleming型デザインでは、例えば1回であれば516例ですが、10回行った場合には相関を全く考慮しない場合には540例まで上がっていきます。ところが相関を考慮することによって、例えば10回の場合では540例から530例、510例、480例あるいは極端な場合430例まで縮小することができます。Pocock型デザインにおいても640例からだいたい510例と、100例以上の縮小を見込むことが可能であることが、今回の研究から分かりました。

【スライド-9】

今の場合10回を計画していれば、全て最終段階の10回までいくことを前提としているのですが、中間解析を設けているということは、当然早期に止めることが前提となっています。ですから、それを早期で止めると考えた場合には、例えば5回を計画していればだいたい4回ぐらいのところまで60%の確率で試験を中止することが可能ですし、あるいは10回まで計画していれば、だいたい8回ぐらいのところまで、60%の確率、つ

スライド-7

3. Group Sequential Design with Co-primary Endpoints Adaptive Strategy for Medical Product Development

Guidance for Industry Adaptive Design Clinical Trials for Drugs and Biologics

DRAFT GUIDANCE

このこれまでの因習的なデザイン・開発と比較し

- 同一の情報をより効率的に提供できる可能性
- 試験目的の達成の確実性を向上させる可能性
- 後続の試験をより適切に計画・立案することに役立つような、治療効果に関する理解を深めることができる可能性: 例えば、用量反応関係や部分集団における良好な推定値を得る
- 開発にかかる人的・資金的資源、時間を削減できる可能性-再配分
- その薬剤を必要とする患者に迅速に提供できる可能性

http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf

The 19th Health Research Forum, November 10, 2012

スライド-8

3. Group Sequential Design with Co-primary Endpoints Maximum Number of Subjects

● 相関を適切に考慮することで、最大症例数を小さくすることができる

$\alpha = 0.025$ $1 - \beta = 0.8$
 $\Delta_1 = \Delta_2 = 0.20$
 $\rho = 0.0$ — 0.3 — — 0.5 — — 0.8 — — 0.99 — —

The 19th Health Research Forum, November 10, 2012

スライド-9

3. Group Sequential Design with Co-primary Endpoints Probability of Early Stopping

● 早期に試験を有効中止する可能性を設けることで、試験参加者の数の縮小、試験費用の削減および試験期間の短縮につながる。

$\alpha = 0.025$ $1 - \beta = 0.8$

The 19th Health Research Forum, November 10, 2012

まり3回に2回は試験をストップすることができるということになります。

【スライド-10】

これを盛り込んで症例数を計算し直すとどうなるかといえば、O'Brien & Fleming型デザイン例えば、2回の場合だと502例という症例数だったが、相関を考慮することによって484例、466例、429例、極端な場合は377例というように症例数を縮小できることが分かりました。

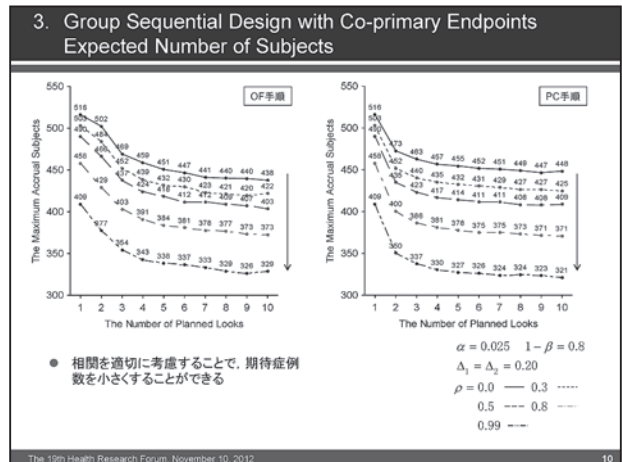
【スライド-11】

最後にこういった検証試験を計画するにあたっては、当然私たちは何らかのデータに基づいて計画することになります。医薬品の開発の場合には、POCと言われるパイロットスタディの結果を用いて試験を計画することになります。ところがパイロット試験は、元々試験規模はそれほど大きくありません。そのデータの平均値に基づいて次の試験を計画すると、試験計画では私たちは検出力を80%つまり10回行えば8回は適切に見いだせるという確率のもとで症例数を設定していたにもかかわらず、実際にはその症例数では目標の検出力には到達していません。例えば、元々100の症例数が必要だったところ30例ぐらいで行った場合には、だいたい検出力は50%ぐらいで程度しかなく、80%の検出力を達成できないことになっています。

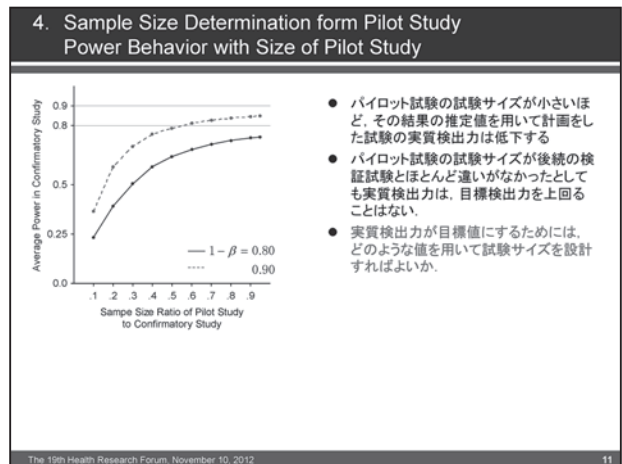
【スライド-12】

通常、医薬品の開発においては症例数の設計において、平均値でなく、保守的に平均値の95%信頼区間の下限をとることがあるのですが、それでは逆に症例数を大きくしてしま

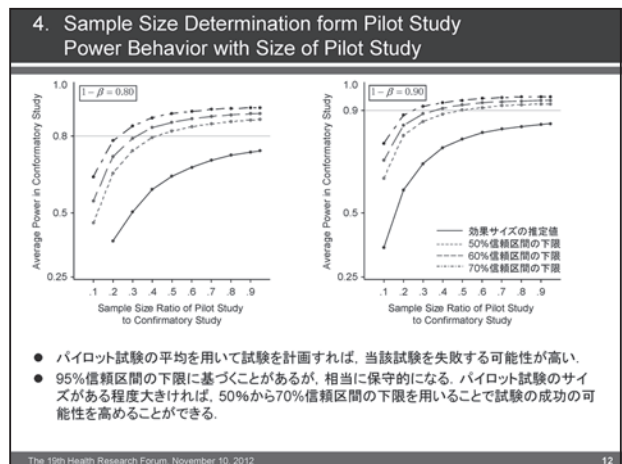
スライド-10



スライド-11



スライド-12



うことが知られています。

そこで今回は、50%信頼区間、60%信頼区間、70%信頼区間を用いることによって、実際に目標としている検出力を達成できるような症例数設計の方法を考えました。

以上が私たちの研究です。

【スライド-13】

まとめです。

私たちは、医薬品が多様化している中でいかに症例数が大きくなることを小さくするかという観点から、標本サイズの検討を行ってきました。

一つの枠組みとしては、エンドポイント間の相関を生かして試験規模を設計するという方法を開発しました。

もう一つは、中間解析を伴う群逐次試験の枠組みの中で症例数をかなり大幅に縮小するという方法を提案しました。

また、パイロットスタディからいかにうまく成功する症例数を設計するかということについて、吟味させていただきました。

このようなことから、私たちの提案法では、合理的かつ実施可能な症例数により、試験参加者の負担軽減になる、あるいは試験費用の削減及び試験期間の短縮につながるような成果を得ることができました。

【スライド-14】

今回ご支援をいただいたおかげで、かなり研究が進捗しました。公表論文は、英文誌5件、発表12件を行うことができました。また、今年の8月にはこの結果をまとめるワークショップを開催して、FDAの統計家にもご参加をいただいて、大変盛大な研究のまとめとすることができました。

これもひとえに選考いただいた選考委員の先生方、またファイザーヘルスリサーチ振興財団の関係者の皆様のおかげだと思っております。この場をお借りし、お礼を申し上げます。どうもありがとうございました。

スライド-13

5. Summary Sample Size Determination for Complex Clinical Trials

- 医薬品の効果は多面的であり、これまでの主要なエンドポイントの一つに限定する仕方では、医薬品の効果を十分に評価することはできない。
 - 複数のエンドポイントを主要変数(Co-primary Endpoints)として設定し、医薬品の評価を行う試みがある
 - しかし、設定したすべてのエンドポイントについて、統計的有意性を示すためには、大規模な試験の実施が必要になる。
- より合理的かつ実施可能な標本サイズ試験から得られる情報量を保持し、また不確実性を適切に評価し、試験規模を縮小するための試験デザインの検討する。
 1. エンドポイント間の相関構造を活かし、試験規模を設計する方法を開発した。
 2. 中間解析を伴う群逐次試験の枠組みのなかで、エンドポイント間の相関構造を活かし、試験規模を設計する方法を開発した。
 3. パイロット試験の試験サイズに伴う不確実性を適切に考慮し、試験サイズを設計するための方法を開発した。
- 合理的かつ実施可能な症例数により、試験参加者の負担軽減、試験費用の削減および試験期間の短縮につながる。

The 15th Health Research Forum, November 10, 2012

13

スライド-14

List of Published Papers/Presentations

- Published Papers (他1論文)**
Hamasaki T, Sugimoto T, Evans SR, Sozu T: Sample size determination for clinical trials with co-primary outcomes: Exponential event-times. *Pharmaceutical Statistics* (First published online: 19 October 2012, DOI: 10.1002/pst.1545).
Sozu T, Sugimoto T, Hamasaki T: Sample size determination in clinical trials with multiple co-primary endpoints including mixed continuous and binary variables. *Biometrical Journal* 54, 716-729, 2012.
Sugimoto T, Sozu T, Hamasaki T: A convenient formula for sample size calculations in clinical trials with multiple co-primary continuous endpoints. *Pharmaceutical Statistics* 11, 118-128, 2012.
Ando Y, Hamasaki T: Practical issues and lessons learned from multi-regional clinical trials via case examples: A Japanese perspective. *Pharmaceutical Statistics* 9, 190-200, 2010.
- Presentations (他6発表)**
Hamasaki T, Sozu T, Sugimoto T, Evans S: Sample size determination in clinical trials with multiple correlated endpoints. *Joint Statistical Meetings 2012*, San Diego, United States, July 28 - August 2, 2012 (Topic contributed paper).
Hamasaki T, Evans S, Sugimoto T, Sozu T: Sample size determination in clinical trials with multiple binary relative risks. 2012 ENAR Spring Meeting, Washington DC, United States, April 1-4, 2012 (Topic contributed paper).
Evans S, Hamasaki T, Hayashi K: Design and data monitoring of clinical trials with co-primary benefit/risk endpoints using prediction. *Joint Meeting of the 2011 Taipei International Statistical Symposium and the 7th Conference of Asian Regional Section of the IASC*, Taipei, Taiwan, December 16-19, 2011.
Hamasaki T, Evans S, Sugimoto T, Sozu T: Sample size determination for clinical trials with two correlated time-to-event co-primary endpoints. *Joint Meeting of the 2011 Taipei International Statistical Symposium and the 7th Conference of Asian Regional Section of the IASC*, Taipei, Taiwan, December 16-19, 2011.
Sugimoto T, Hamasaki T, Sozu T: Sample size determination in clinical trials with two correlated co-primary time-to-event endpoints. *The 7th International Conference on Multiple Comparison Procedures*, Washington DC, USA, August 29-September 1, 2011.
Hamasaki T, Evans S: Power and sample size determination for clinical trials with two correlated binary relative risks. The 2011 International Conference on Applied Statistics, National Taipei University, Taipei, Taiwan, May 26-27, 2011 (Invited Presentation).

The 15th Health Research Forum, November 10, 2012

14

質疑応答

会場： 一つ質問ですが、コ・プライマリにする場合の多重性の調整をどうしているかということです。単純に有意水準を分割しているのか、閉手順で順番をつけるという手法をやっているのか、というところを教えてください。

濱崎： 今回の場合、コ・プライマリの問題なので、両方とも有効性を示すという問題になっておりますので、多重性の調整は必要ありません。ただし、Type 2 error エラーの方が膨張しますので、Type 2 error を制御するために、エンドポイント間の相関を無視すれば検出力を通常例えば80%としていたものを、各エンドポイントについて90%に上げる必要があります。例えば90%と90%をかけ算すると80%になるので、そういった調整が必要になります。ただそのように上げてしまうと、症例数が1.2倍から1.4倍と上がってしまうので、今回は、それを相関を考慮することによって下げるという対処をとっています。

会場： どちらのエンドポイントもポジティブでないと結果を良いと言わないので、多重性の調整はしないということですか。

濱崎： 必要ないということです。

会場： 我々のがんの領域では、最近、二次治療以降の治療効果が非常に強くなってきて、オーバーオールサバイバル (OS) ではかなりサンプルサイズが増大してしまうので、プログレッションフリーサバイバル (PFS) とOSをコ・プライマリにおいてやるという手法がよく出てきています。先生のご発表だと、PFSとOSは当然相関はあると思うのですが、その相関を考慮することで、がんの領域でもサンプルサイズを減らせるという内容でよろしいでしょうか。

濱崎： はい。実は最近そういった結果について学術雑誌に投稿して、公表させていただいております。

座長： 複数のエンドポイントをおかれた場合の相関という話なのですが、元々エンドポイント自身のウエイトあるいは重要度が臨床的に違う場合は、すごく差が出るケースもあり得るのでは。そこはどのように組み込むことができるのでしょうか。

濱崎： 今回の場合、研究においてはエンドポイントについてはウエイトを設けておりませんが、解析のところでウエイトを設けるという手法があります。そういったものを逆に解析で使うということであれば、ウエイトをつけてサンプルサイズを計算することも可能になります。

座長： そういうこともできるわけですね。

濱崎： ただし、解析については順序をつけて解析を行うということが必要になってきます。

座長： お聞きしていると、サンプルサイズとコストの問題のバランスのちょうど良いところを見つけ出すのが結構難しいようにも思います。今日具体的な例はお話にならなかったのですが、具体的に何か比較された例はあるのですか。既に開発されたものでいいのですが。

濱崎： 現在、ハーバード大学の研究者がエイズの試験に携わっており、そちらではやはりこのようなコ・プライマリ・エンドポイントで評価を行っております。実際に今回のような方法を適用を検討している最中です。

会場： 教えていただきたいのは、試験デザインするにあたり相関性の情報を予め求める、あるいは中間解析のときに相関性について補正されると思いますが、そのへんのファクターというのはどのように組み込まれて研究されたのでしょうか。つまり予め相関性の情報を持ってくる…例えばOSとPFSですと、8週目あたりだとその関連性はより濃くなっているとか、相関性自体も自体も情報が変わってくると思うのです。

濱崎： そうですね、大変に技術的な話になりますが、例えば生存時間の場合には、後半部分で相関が強くなる場合、むしろ前半部分で相関が強くなる場合、中間部分で相関が強くなる場合と、いくつか相関を定義するモデルがいくつかあります。私たちはそれらの3パターンについて、どのように症例数設計を行うかということの評価させていただきました。本日は時間の関係で詳細は省かせていただきますが、既に論文を公表させていただいています。後ほど論文を紹介させていただきます。

会場： どうもありがとうございました。