



臨床評価過程における累積情報の統合的活用に向けた 統計基盤の研究

大阪大学臨床医工学融合研究教育センター 特任准教授（常勤）

寒水 孝司

【ポスター 1】

私の専門分野は生物統計学、医学統計学と呼ばれるもので、普段は臨床試験の計画とデータ解析に関する統計学的方法論の研究を行っています。広い目的としては、科学的かつ効率的な臨床試験の実施・実現を通じて、医薬品開発の計画・解析の幅を広げたいと考えています。

今日の開会式でありましたように、近年医薬品の開発が意図した通りに行われているか、そういう結果が出ているかというと、必ずしも充分な状況ではないと言えます。そこで、累積された科学的根拠・情報を適応的に臨床評価の過程に利用するという試みがなされています。

そのような背景のもとで、統計学的方法論を 3 つ取り上げて研究を行いました。今日は時間が限られていますので、複数の主要変数の対処に対する課題について、研究成果を報告させていただきます。

【ポスター 8】

順番が逆になりますが、イメージを持っていただきたいので、ポスター 8 を見ていただければと思います。

ここで示している事例の患者さんは透析中の慢性腎不全患者で、対象とする疾患は高リン血症です。左側の縦の棒がリンの濃度です。後で使うのでカルシウムの濃度も併せて示してあります。この点線が基準範囲の上限です。そうしますと、リンの濃度が基準範囲の上限を超えており、高リン血症の状態を表わしていることになります。

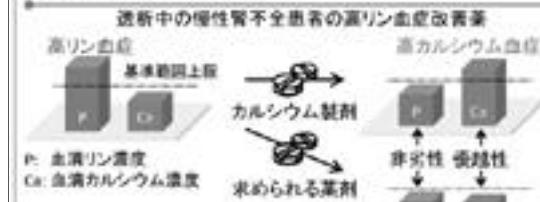
ポスター 1

研究目的

- 科学的かつ効率的な臨床試験の実施・実現を通じて、医薬品開発の計画・解析の幅を広げる
- 累積された科学的根拠・情報を適応的に臨床評価の過程に利用する
- 近年提案されている統計的方法論の特徴・性能・実用化の条件を明確にする
 - ・ 適応的試験計画
 - ・ 複数の主要変数への対処
 - ・ データ適応型生存時間解析

ポスター 8

事例に則した検討



1群あたり必要症例数 ($\Delta_{\alpha}, \Delta_{\beta} = 0.500, 0.455$)

検出力	相関係数							
	0.2	0.5	0.8	0.2	0	0.3	0.4	0.6
80%	54	54	91	94	93	91	90	87
90%	118	118	118	117	117	116	115	113

このような患者さんに対して、従来はカルシウム製剤が一般的に用いられていたわけですが、製剤の特性上、高カルシウム血症を引き起こしてしまうという問題がありました。そこで、新しい薬剤を開発するにあたっては、リンの濃度については従来の薬剤の効果に劣らないことを示し、かつ、カルシウムの濃度については優越性を示すという、2つの項目について統計学的な有意差を示すことが一つの開発戦略になります。

ところが、このような目的に対して、臨床試験の必要症例数を計算する方法が確立していませんでした。そこで、私はその方法について研究を行い、必要症例数を計算する方法を開発しました。

【ポスター 2】

このような状況に対して、「臨床試験のための統計的原則」では、「主要変数は試験の主要な目的に直結した臨床的に最も適切で説得力のある証拠を与える変数であるべきである」という記載がなされています。こちらをちょっと注目していただきたいのですが、このガイドラインでは、「主要変数は通常ただ一つにすべきである」という記載がなされています。とは言いましても、実際に一つに限定するのは難しい状況もありますので、「治療の及ぼす効果の範囲を一つ一つの変数が包含するような、複数の主要変数を用いることが望ましい場合がある」ということで、例外を認めています。

【ポスター 3】

近年、この影響を受けまして、米国の製薬協に相当する PhRMA は、FDA が複数の主要変数の有効性を示すことを要求する 20 の疾病を例示しています。例えば片頭痛では、頭痛の消失、吐き気、光過敏、音過敏の 4 つの項目について、有効性を示すことが求められています。

このような状況に対して必要症例数の計算式を考えるわけですが、データの型に応じて用いる検定方法が異なりますので、それに応じた方法を考える必要があります。

研究助成をいただく前に、正規分布に従うデータについては既に研究を終えていましたので、研究助成をいただいて、二値変数、つまり結果が陽性・陰性、効果があったか・なかったかというような二値で観測される変数について方法論を拡張しました。

ポスター 2

複数の主要変数への対処

臨床試験のための統計的原則 (1999)

○ 2.2.2 主要変数と副次変数

- ・主要変数は、試験の主要な目的に直結した臨床的に最も適切で説得力のある証拠を与える変数であるべきである
- ・主要変数は通常ただ一つにすべきである

○ 2.2.5 様数の主要変数

- ・治療の及ぼす効果の範囲を一つ一つの変数(又は変数の組)が包含するような、複数の主要変数を用いることが望ましい場合がある
- ・第二種の過誤及び必要な被験者数への影響は慎重に考慮すべきである

ポスター 3

複数の主要変数が必要な疾病

○ PhRMA: Multiple Endpoint Expert Team (2007)

- ・規制当局が複数の主要変数の有効性を示すことを要求する 20 の疾病を例示

例: 片頭痛

(1) 痛風消失 (2) 吐き気 (3) 先頭痛 (4) 側頭痛

アルツハイマー

(1) Alzheimer's Disease Assessment Scale Cognitive

(2) Clinician Interview-Based Impression of Change



取り組む課題

主要変数の相関を考慮した症例数設計法の構築

○ 連続型データ(正規変量)

○ 様数型データ(二値変数)

○ 生存時間データ

ポスター 4

提案した症例数設計法

○正規変量

- 二変量正規分布に従う応答变数についてモンテカルロ・シミュレーションによって検出力を近似する方法
- 検出力の推定値が確実よく短時間で得られる

Sone et al. (2006)

○二値変数

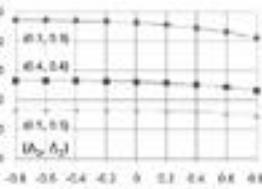
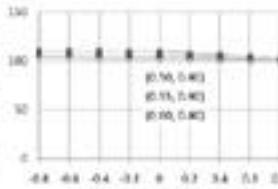
- 二変量ベルヌーイ分布に従う応答变数について検定統計量の漸近正規性に基づいて検出力を近似する方法
- モンテカルロ・シミュレーションを必要としない
- 相関係数に範囲が生じることに注意が必要

○生存時間データ

- 今後の課題

ポスター 5

数値例（正規変量）

有意水準 2.5% 検出力 80%以上
n: 1群あたり必要症例数n: effect size Δ が主要変数間で等しい場合 $\Delta_{ij} = \Delta_0 + \Delta_1 \rho_{ij}$ $\rho_{ij} = \rho$ (相関係数) $\Delta_0 = 0.1$ (主要変数 i の effect size (基準化した効果の差)) $\Delta_1 = 0.1$ (主要変数 i の effect size (ばらつきで基準化した効果の差)) $\rho = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ (相関係数)n: effect size Δ が主要変数間で異なる場合 $\Delta_{ij} = \Delta_0 + \Delta_1 \rho_{ij}$ $\rho_{ij} = \rho$ (相関係数) $\Delta_0 = 0.1$ (主要変数 i の effect size (基準化した効果の差)) $\Delta_1 = 0.1$ (主要変数 i の effect size (ばらつきで基準化した効果の差)) $\rho = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ (相関係数)

既存の多変量正規分布を前提にした方法は必要症例数を4例/群程度過小評価することがある

【ポスター 4】

細かいところは統計の話になりますが、結果の概略はポスター 4 に示す通りです。

【ポスター 5】

こちらはデータが正規分布に従う場合の結果です。このグラフは、縦軸に 1 群あたりの必要症例数、横軸に主要変数間の相関係数を表わしています。グラフから分かるように、2 つの主要変数間で効果がだいたい等しい状況においては、相関係数が高くなるに連れて必要症例数は小さくなる、つまり検出力が高くなるということが確認できました。定性的にはこうなるだろうということは事前に分かっていましたが、それに対する定量的な結果を示すことができました。

【ポスター 6】

二値変数についても同じような傾向が確認できました。ただし、相関係数と聞くとマイナス 1 から 1 の範囲を取ると思いますが、二値変数については状況によって限られた範囲になるという特殊性があります。その点を踏まえて症例数を設計することが必要です。

【ポスター 7】

数値例の要約です。相関係数が負の場合は、ポスター 5、6 のグラフの真ん中の目盛りが 0 ですので、これより左側の結果になりますが、必要症例数はほとんど変化しません。ただし、2 つの主要変数について相関が負ということは考えにくいので、実際には正の結果を見ることになります。

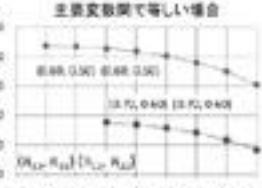
主要変数間の effect size (ばらつきで基準化した効果の差) が

ポスター 6

数値例（二値変数）

有意水準 2.5% 検出力 80%以上
n: 1群あたり必要症例数

n: 成功確率の目盛りが主要変数間で等しい場合

 $\Delta_{ij} = \Delta_0 + \Delta_1 \rho_{ij}$ $\rho_{ij} = \rho$ (相関係数) $\Delta_0 = 0.1$ (主要変数 i の成功確率) $\Delta_1 = 0.1$ (主要変数 i の成功確率)

n: 成功確率の目盛りが主要変数間で異なる場合

 $\Delta_{ij} = \Delta_0 + \Delta_1 \rho_{ij}$ $\rho_{ij} = \rho$ (相関係数) $\Delta_0 = 0.1$ (主要変数 i の成功確率) $\Delta_1 = 0.1$ (主要変数 i の成功確率)

相関係数の解釈・見積もりが比較的難しい

大きく異なる場合は、片方の主要変数に対する必要症例数が全体の必要症例数になります。これは、個別に必要症例数を計算したときに検出力が大きく異なるためであり、効果の小さい方の検出力を確保するための必要症例数が全体の必要症例数になるということです。

effect size が同程度の場合は、相関係数が大きいほど必要症例数は小さくなります。

【ポスター 8】(再掲)

最後にもう一度事例に戻ります。この臨床試験では、リンの濃度とカルシウムの濃度のそれぞれについて必要症例数を計算して、その大きい方を試験全体の必要症例数とするというやり方が採用されていたと考えられます。

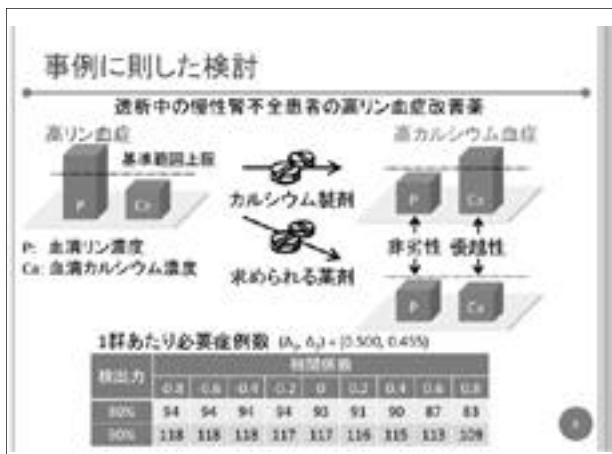
ここで私の方法を使いますと、検出力を 80 % 確保するのであれば、1 群あたりの必要症例数は相関係数の大きさに応じて 83 例から 94 例になります。検出力を 90 % 確保するのであれば、109 例から 118 例になります。したがって、過去の臨床試験の累積情報を用いて、2 つの主要変数の相関係数を見積もって必要症例数を計算することになります。

【ポスター 9】

まとめです。研究助成をいただいて、主要変数間の相関を考慮した症例数設計法を提案いたしました。主要変数間の相関係数が重要な局外母数になりますので、その大きさの評価が重要になります。

近年、創薬に関する情報は加速度的に増加していますが、その情報をいかに選択・縮約して臨床試験の計画に活用するかが今後とも大きな課題になると思います。

ポスター 8 (再掲)



ポスター 7

数値例の要約

- 相関係数が負の場合、必要症例数はほとんど変化しない
- 主要変数間で effect size が大きく異なる場合、effect size が小さい主要変数に対する必要症例数が全体の必要症例数になる
 - 相関係数は必要症例数に影響しない
- effect size が同程度の場合、相関係数が大きいほど必要症例数は小さくなる

ポスター 9

まとめ

- 主要変数間の相関を考慮した症例数設計法を定式化
- 主要変数間の相関係数が重要な(局外)母数になるので、その大きさの評価が重要
- 創薬に関する情報が加速度的に増加する一方で、それらの情報をいかに選択・縮約して臨床試験の計画に活用するかが今後とも大きな課題になる

質疑応答

座長： 症例数の設定の問題は、我々臨床研究をやっている者にとってはとても切実な問題で、できるだけ少ない方が研究費も少なく済むので、大変意味のある発表だったと思います。

会場： すごく短い質問なのですけれど、これは要するに、2つの変数が両方とも有意差が出るのに必要な症例数という意味ですか？

寒水： はい。その通りです。

座長： 大変素晴らしいアイディアなのですが、実際に見てみると 94 が 83 くらいになるということで、あまり劇的な節約効果が見えないのですけれども・・・

寒水： この数値例では、1群あたりの必要症例数を示していますので、全体では 20 例程度の差が生じます。もう少し規模が大きくなつて 300 例とか 500 例ぐらいの研究になると、その分大きな差が生じます。

会場： 実際、2つのアウトカムでその差が出ることが求められるようなものというのは、だいたい関連が強いものですね？

寒水： そうですね。

ただし、ご存知のように関連が強いと言っても、この分野においては相関係数が 0.4 くらいでも高いということがあります。実際にこちらの論文を見ていただくと、20 個の疾病が取り上げられていて、低い・中位・高いという 3 つのカテゴリー分けですが、変数間の相関が記載されてあります。

座長： この場合、たまたまこの例では effect size がある程度ある。もし effect size がすごく小さい場合には、ものすごい効果（効果量）が出る。

寒水： そうですね。

座長： そうすると、節約する絶対的な症例数も多くのなるということですね。

寒水： そうですね。例えば、この例では相関係数が 0 のところが 500 例くらいで、0.8 のところが 450 例くらいですので、全体で 100 例程度の差が生じます。

会場： 最近臨床研究に関する指針とか、サンプルサイズの問題とかが言われていて、一番単純に考えれば、ある 2 つを解析したいという場合に本当にパワーのある解析結果を出すためには、人間の臨床研究の症例数としてはだいたい 300 例と

か 350 例が確保できないと統計学的に厳しいというお話を聞くことがあるのですが、例えば先生の今のこの方法だと、症例数が何%少なくてすむとおっしゃいましたっけ？

寒水： 一概には言えません。

会場： 2 群を t 検定で単純に比較するような場合、そのサイズというのは少し減らすことは充分可能だとお考えですか。

寒水： 相関係数が高い場合には・・・

会場： 高ければ、ですね。その相関係数は、先ほどの話で 0.4 くらいからいいとなるのですけれども、やはり 0.6、0.8 は欲しい場合もあるのです。工業的なものを扱ったりする場合、特にそうなのですが。その辺は人間でやると、人間は 100 人いたら 100 通りですから・・・

寒水： そうです。ばらつきがあります。

会場： そこに対して同じ物を使った場合の比較・・・例えば男女比較でも何でもいいのですが、そういうサンプルサイズというのは 2 群あった場合、2 通りあった場合に、何例くらいずつあったら、先生はまずこれが間違いないだろうと一つの答えを出すのですか？

寒水： それは効果の差とばらつきによって決まることなので、一概にはちょっと言えません。

座長： 言えませんね。

会場： 分かりました。有り難うございました。