

氏 名 小 山 暢 之

学位（専攻分野） 博士(学術)

学 位 記 番 号 総研大甲第438号

学位授与の日付 平成12年3月24日

学位授与の要件 数物科学研究科 統計科学専攻

学位規則第4条第1項該当

学 位 論 文 題 目 臨床試験における頻度論的中间解析手法

α -spending function 法と Indifference-zone アプローチ

論 文 審 査 委 員 主 査 教 授 柳本 武美

教 授 平野 勝臣

助 教 授 栗木 哲

助 教 授 佐藤 俊哉

Anthony Hayter (統計数理研究所)

論文内容の要旨

臨床試験を実施する際には、試験に参加する被験者の安全性と利益を可能な限り確保することが重要であり、そのために試験の途中でデータを評価する試験デザインを採用することもしばしばある。このような試験途中でのデータの評価を中間解析といい、これまで様々な統計学的手法が提案されている。本稿では頻度論的な中間解析手法として、仮説検定のための α -spending function 法と選択問題としての手法である indifference-zone アプローチを取り上げ、これらの方法を用いた際の効率的な試験デザイン法について検討することを目的とする。

・ α -spending function 法

試験治療間の比較などを目的に中間解析時点ごとに仮説検定を繰り返し行う場合、検定の多重性の問題が生じ、帰無仮説の下で少なくとも 1 回の検定で有意差ありとなる確率は、各検定時の名義的有意水準よりも大きくなる。このため、中間解析時に仮説検定を行うためにはこの多重性を調整する方法が必要となる。 α -spending function 法は Lan and DeMets(1983)により提案された中間解析に伴う検定の多重性を調整した棄却限界値の算出法である。この方法はあらかじめ有意水準 α を消費する割合を定めた関数を一つ固定するだけで、各中間解析時点ごとにその関数とそれまでに得られた情報量からその時点での棄却限界値を算出できる。このため、解析回数や解析時点に依存せず試験全体の有意水準を事前に定めた値に保つことができることから、非常に柔軟な方法として現在でも実際の臨床試験でよく用いられている手法である。しかしながら、その適用上の問題点として乱用と解析時点の最適性が上げられる。 α -spending function 法の乱用とは、得られたデータに依存してその後の解析時点や解析回数を変えることを意味する。データに依存して解析時点が選択された場合には α -spending function 法を用いて棄却限界値を算出しても試験全体の有意水準が事前に定めた値に保たれないことが指摘されている。しかし、乱用に対応するための明確な統計学的方法はこれまで与えられていない。本稿では、このような「乱用」への対処法として、ある一時点の中間解析結果により、その後の中間解析回数を一回だけ増やす可能性があるという単純な状況に対する棄却限界値を α -spending function から算出する方法を提案した(本稿 3.1 節)。また、この方法を用いた棄却限界値の算出例として、1 回目の中間解析で検定統計量の値が有意ではないが、棄却限界値の 8 割の値を越えた場合に一度だけ中間解析を追加で行うとしたときの棄却限界値を算出した。その結果、中間解析を増やした場合の棄却限界値は大きな値となり、試験結果を保守的に解釈しなければならず、中間解析結果に依存して解析時点を増やすことにはあまり大きな利点がないことが示唆された。

一方、 α -spending function 法では試験全体の有意水準は解析時点に関係なく維持できるが、検出力や対立仮説が正しい場合の試験の期待停止時間あるいは期待被験者数(情報量)は解析時点や解析回数に依存して変化する。このため、検出力や試験の期待停止時間などを指標として解析時点の最適性について考察することが可能である。本稿では解析時点の最適性について、代表的な 3 つの α -spending function(O' Brien-Fleming 型、 Pocock 型、直

線型)を用いた場合に以下の条件の下で期待被験者数(情報量)が最小となる解析時点を数値計算により検討した。

- ・ 最大被験者数(情報量)を固定した場合
- ・ 検出力を固定した場合

その結果いずれの条件下でも、O' Brien-Fleming 型を用いた場合にはデータが全体の 1/2 を越えた段階で information fraction(興味あるパラメータに対する中間解析時点での Fisher 情報量とすべてのデータが得られたと仮定したときの Fisher 情報量の比)に対してほぼ等間隔で解析を行うと期待被験者数が最小となることが示された。同様に、Pocock 型や直線型では試験全体で information fraction に対してほぼ等間隔に解析を行うと期待被験者数が最小となる。この結果は臨床試験の計画時に中間解析時点を検討するための一つの指針となり、また、これらの関数を α -spending function として用いる限り、極端に解析間隔を狭めたり、広げたりする解析計画は被験者数の観点からは効率が悪いということが示されたことになる。

・ Indifference-zone アプローチ

複数の群の中からある基準を満たす群を選択したい場合、あらかじめ単純な選択ルールを定めてデータを収集し、その選択ルールを満たした治療群のみを選択するという方法がある。このような試験方法を選択問題(selection problem)や選択確率(selection probability)に基づく方法という。選択問題では、正しく治療群を選択できる確率(選択確率)がある一定レベル以上になるように選択ルールを定めるが、この確率は興味あるパラメータに依存する。このため、選択すべき群と選択すべきでない群の間のパラメータ値にはある一定の値以上の差があるという仮定の下で選択確率の最小値を評価する方法が indifference-zone アプローチである。

本稿では、臨床試験において試験治療の有効性が「有効」、「無効」のような二値データで表される場合に複数の試験治療の中から対照治療に劣らないものを過不足なく選択するための試験デザインを indifference-zone アプローチにより検討した。選択問題としての試験デザインの多くは、複数の試験治療の中から最も良いものを一つだけ選択することを前提としており、臨床試験への応用についてもこれまで Simon や Thall らによりいくつか提案されている(本稿 2.6 節)。一方で、対照治療を基準にそれより劣らないものをすべて選択するための試験デザイン法はほとんど提案されていない。しかし、新治療の開発過程においては複数の治療法の中から有望なものを見出すためのスクリーニングを目的とした臨床試験もしばしば実施され、このような場合は必ずしも一つの治療法のみを選択すればいいというわけではない。このため、本稿第 4 章においてこのような状況における one-stage および two-stage での試験デザイン法について提案した。One-stage デザインは選択ルールを一つ設定し、すべてのデータが収集された段階でその選択ルールを適用して群を選択する。Two-stage デザインでは、試験の途中で一度中間解析を行い、その時点で対照治療よりも劣ると判断された試験治療群は削除し、残った群のみで試験を最後まで行う。このため、中間解析時と最終時点それぞれに対し選択ルールを設定することになるが、この方法により劣った治療を施される被験者の数を少なくすることが期待でき、one-stage デザインよりも平均的に少ない被験者数で試験を実施できる可能性がある。本稿では、いずれの場

合もあらかじめ定めた選択確率を達成するために必要な被験者数が最小となるように選択ルールを設定することを提案し、実際に試験治療群の数が 2,3,4 の場合に選択確率が 0.8 以上となる被験者数と選択ルールを数値計算により与えた。その結果 two-stage デザインでは、one-stage デザインに比べ期待被験者数が数%～ 10%程度少なくなることが示された。また、対照治療群へ試験治療群よりも多くの被験者数を割り付けた場合についても同様に one-stage および two-stage デザインを検討した結果、対照治療の有効率が 0.8 以下であれば、より少ない被験者数で試験が実施可能であることが示された。

参考資料である 'Complete classes in comparison of sequential binomial experiments' (早稲田大学理工学部 草間 時武 教授との共著、'Statistics & Decisions, Volume 18 (2000)' 掲載予定) は、逐次二項試行実験(sequential binomial experiments)を取り上げ、2つの実験間に充分性を定義して半順序を導入し、その半順序での完備類について理論的に検討した論文である。本論文では完備類に含まれる実験の停止則はパス(その点に至るまでの過程)に依存しないことや、ある実験に対し十分な実験が存在した場合に期待標本数は十分な実験の方が大きいことなどを示し、完備類に含まれる実験に関するいくつかの性質を明らかにした。また、完備類に含まれる実験の中から確率的な停止則をもつ実験を除くことができない、つまり、完備類を非確率的な停止則をもつ実験のみで構成することができないことも示している。本学位論文の関連として、対象としている確率モデルが二項分布であり、逐次的実験の比較を目的としていることが 4 章における indifference-zone アプローチ方での試験デザイン法と共通し、実際、完備類の性質からパスに依存しない選択ルールを用いて試験をデザインすることの一つの妥当性が導かれる可能性が考えられる。しかしながら、この点についての検討は十分になされておらず、また、本学位論文では group sequential 法を扱った論文であることから、論文 'Complete classes in comparison of sequential binomial experiments' は参考資料とした。

論文の審査結果の要旨

同君は早稲田大学で主として数学、修士課程で草間教授の指導のもとに、数理統計学、特に逐次選択問題の研究に取り組んだ。就職してからは実務に携わる一方で、臨床試験に係る統計的研究を行うことを志望して本学に入学した。従来の研究と実務での経験から、入学時より中間解析に関する研究を選んだ。臨床試験における中間解析は、倫理的側面及び被験者数を減らすことの資源・経済的側面の両面から研究が待たれている。理論面では確率過程の理論の援用が期待されている。

学位請求論文では中間解析手法として、主として2つのアプローチが採用されている。その一つは、頻度論的な立場から仮説検定のための α -spending function法であり、他の一つは、選択問題としての手法である Indifference-zone を設定した研究である。その結果は効率的な試験デザインが設定されるための基礎研究である。また同時に理論的な研究をも継続して行い、学位論文に関連した結果をまとめている。

論文は4章からなる。1章は問題の提示である。2章では関連した研究が手際よくまとめられている。また臨床試験の現場で生じる問題と統計理論の関連についてもレビューされている。同君の主要な貢献は3章及び4章で論じられている。

3章では α -spending function法と呼ばれる方法が研究されている。これは中間解析で生じやすい多重性の問題を避けるための一つの工夫である。この方法は Lan and DeMets (1983)により提案されて以降、多くの研究がなされてきた。同君は、中間解析回数を1回のみ増やすことがある状況で棄却限界値を算出する方法を検討している。一つの結果として、中間解析を増やした場合の棄却限界値は大きな値となることを指摘して、中間解析結果に依存して解析時点を増やすことにはあまり大きな利点がないことを示唆した。次いで、解析時点の最適性が検出力の観点から調べられている。 α -spending functionの関数型を変えた場合その影響が問題となる。その結果、O'Brien-Fleming (1979)が提案した関数型では、最適な解析時点としてデータが全体の1/2を越えた段階でほぼ等間隔に行うことになること、別の関数型ではほぼ等間隔になるとの興味ある結果を得ている。これらの結果は数値計算によるものである。

4章では Indifference-zone を設定したもとの選択問題が研究されている。この設定は、ある一定の値以上の差を、選択すべき群と選択すべきでない群の間に設ける。既存の研究としては Simon ら(1985)や Thall (1989)らがある。選択問題を臨床試験に適用する場合には、複数の試験治療の中から対照治療に劣らないものを的確に選択することを要請する。この要請を実現するためには、一定の幅を仮定することになる。この論文では One-stage デザインと Two-stage デザインを比較して、中間解析の効果を検討している。One-stage デザインではすべてのデータが収集された段階でその選択ルールを適用して群を選択し、Two-stage デザインでは試験の途中で一度の中間解析を行う。中間解析の回数としては前節よりも単純であるが、複数の試験治療群と対照群を考察している意味ではより複雑である。試験治療群の数を変えて Two-stage デザインの効果を調べ、期待被験者数を少なくすることを示している。その上でこの結果が臨床試験の実施上で実用的であることを主張している。この結果の一部は、現在投稿中である。

参考資料として、関連した理論研究であり国際誌に受理されている “Complete classes in

comparison of sequential binomial experiments”（早稲田大学理工学部 草間 時武 教授との共著）が挙げられている。この論文では、逐次二項試行実験において、2つの実験間に十分性を定義して、その半順序での完備類が検討されている。学位請求論文との関連が強くないと判断し参考論文としている。しかし審査委員の中には、この論文は数学的な力を裏付けるものであり、重く評価できるとする意見があった。また逐次的統計手法の観点からみれば本論文とも関連がある。この他に、既に受理された論文として和文論文が2編挙げられている。

論文を全体的に評価すると、4章で得た成果は十分に論文としてのレベルにあると判断された。3章の結果も高く評価する意見があった。この学位請求論文の長所は実際の臨床試験への適用が意図的に想定されていることである。以上の結果この論文は、統計科学専攻において同君に学術博士を授与させるにふさわしいと判断された。