



Title	Group-Sequential Three-Arm Noninferiority Clinical Trial Designs in Medical Product Development
Author(s)	落合, 俊充
Citation	大阪大学, 2016, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/55834
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、 〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉 大阪大学の博士論文について 〈/a〉 をご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

氏 名 (落 合 俊 充)	
論文題名	Group-Sequential Three-Arm Noninferiority Clinical Trial Designs in Medical Product Development (医薬品開発における3群非劣性試験の群逐次デザインに関する研究)
論文内容の要旨	
<p>はじめに</p> <p>医薬品による治療は、患者の健康の回復やQOLの改善を図るうえで重要な役割を果たす。医薬品開発の過程では臨床試験が不可欠であり、対照薬との比較による有効性や安全性の評価、用量反応関係の評価といった目的に応じて試験デザインを立案し、統計的評価を行うことが求められる。例えば、試験薬が有効性において既存標準薬に劣らないことを示す非劣性試験では、適切な非劣性マージン（臨床的に許容される既存標準薬との差）を設定することに加え、分析感度（過去に実施された試験と同様に、プラセボに対する標準薬の効果が認められること）を担保したうえで試験薬の効果を評価しなければならない。試験薬と標準薬を用いた2群非劣性試験は、非劣性検証でよく用いられている試験デザインであるが、標準薬の有効性が検証された頃とは許容される併用療法に変化が生じていたり、エンドポイントや患者の組み入れ基準に違いが見られたりする場合には、過去の試験成績に基づいて分析感度の存在を主張することは困難になり得る。この問題に対処する一つの方法として、ICH（日米EU医薬品規制調和国際会議）等のガイドラインでは、試験薬と標準薬と共にプラセボ群を設けた3群非劣性試験が提案されている。3群非劣性試験は、同一試験内で分析感度を評価できる利点があるが、非劣性と分析感度の少なくとも2つの仮説に対して統計的な評価が必要となる。また、評価者の立場によって仮説の設定が異なる可能性があり、評価方法も複数考えられる。本研究では、この3群非劣性試験に着目し、症例数の軽減につながる可能性のある、試験途中での有効性中止を目的とした群逐次デザインの適用を通して、より効率的な試験デザインについて考察した。</p> <p>方法</p> <p>3群非劣性試験での統計的評価は、固定マージン法とFraction法の2つに大きく分類される。固定マージン法は、非劣性マージンを事前に定め、「プラセボの効果」<「非劣性マージンを差し引いた標準薬の効果」<「試験薬の効果」という大小関係の成立を前提とする評価方法である。一方、Fraction法は、試験薬が標準薬の効果の一定割合を保持していることをもって非劣性を検証する方法である。本研究では、これら既存の評価方法を群逐次デザインで利用できるように拡張し、次のルールに基づいて意思決定を行う試験デザインについて検討した。このとき、中間解析時に分析感度や非劣性の成立の判断基準となる棄却限界値には、O'Brien-Fleming型境界と Pocock型境界とを組み合わせることを考えた。これらの条件下で期待症例数や検出力、Type I errorの挙動を数値的に評価し、各試験デザインの特徴を考察した。</p> <p>意思決定ルールA（以下DF-A: Decision-making Framework A） 最終解析までのいずれかの時点で、分析感度と非劣性がそれぞれ検証された場合に試験成功とみなす。 ただし、非劣性の検証は、分析感度が成立した時点以降で行う。</p> <p>意思決定ルールB（以下DF-B: Decision-making Framework B） 同一解析時点で分析感度と非劣性が同時に検証された場合に試験成功とみなす。</p> <p>また、試験デザイン立案時には利用できる情報が限られており、試験薬について十分な事前情報が得られていないことがある。このような場合は、試験途中で必要症例数を見直すことにより、試験の成功確率を高められる可能性がある。そこで、試験途中で1回症例数を見直す選択肢を加え、前述の試験デザインに症例数の再計算を組み入れたときの影響についてもシミュレーション研究を行った。なお、症例数の見直し方として、ここでは症例数を増やすことの</p>	

み考慮する場合、減らすことのみ考慮する場合、増減共に考慮する場合の3通りについて検討した。

結果と考察

解析回数を全4回とし、試験薬群と標準薬群の効果（望大特性）を共に10、プラセボ群の効果を5、各群共通の標準偏差を6.5と仮定したもとで検討した結果、意思決定ルール間で明確な違いはないものの、DF-Aでの期待症例数はDF-Bと比べてわずかながら少なくなる傾向にあった。例えば、各群に等例数を割付け、非劣性マージンを2.5とした固定マージン法を用いた場合、検出力80%を確保できるようにデザインしたもとの期待症例数は、DF-Aで分析感度と非劣性の評価に共にO'Brien-Fleming型境界（以降、OF-OF）を用いたとき388例、分析感度にO'Brien-Fleming型境界、非劣性にPocock型境界（以降、OF-PC）を用いたとき410例、分析感度にPocock型境界、非劣性にO'Brien-Fleming型境界（以降、PC-OF）を用いたとき402例、分析感度と非劣性共にPocock型境界（以降、PC-PC）を用いたとき400例であった。これに対しDF-Bでは、OF-OFで391例、OF-PCで411例、PC-OFで411例、PC-PCで406例であった。なお、中間解析を行わない標準的な固定デザインで検出力80%を確保するために必要な症例数は426例であり、DF-AあるいはDF-Bの試験デザインを用いることで症例数の軽減につながると考えられた。また、DF-Aにおいて固定マージン法を用いた場合、中間解析時に分析感度が成立した際には、その後の非劣性評価でプラセボ群の情報は利用しないことから、その時点でプラセボ群の集積を打ち切ることが可能となる。DF-Aにこのオプションを含めて期待症例数を算出すると、OF-OFで375例、OF-PCで399例、PC-OFで371例、PC-PCで376例となり、固定マージン法でDF-Aを用いる際には、症例数をより少なくできると考えられた。

また、試験途中での症例数の再計算に関するシミュレーション研究より、症例数を減らすことのみ考慮するデザインは、期待症例数は少なくなるものの、試験全体の検出力は低下する可能性があるため、適切な選択肢ではないことが示唆された。例えば、各群の効果を前述と同様に仮定し、計画時の症例数を426例としてDF-A（固定マージン法、棄却限界値の組み合わせとしてOF-OFを利用した場合）を適用したときのシミュレーション結果は、症例数の再計算を行わない場合の期待症例数と検出力はそれぞれ383例、79.1%であったが、計画症例の50%集積時に症例数を減らす方向で見直した場合は367例、75.7%であった。一方、増やす方向で見直した場合は455例、90.7%、増減可として見直した場合は439例、87.7%であった。なお、DF-BあるいはFraction法を用いた場合でも同様の傾向であった。

これらの結果は本研究で仮定した条件下で得られたものであることに注意が必要であるが、本研究により、DF-AあるいはDF-Bのもとの群逐次デザインは、標準的な固定デザインと比べて少ない症例数で早期に試験薬の効果を検証できる可能性を有することが示されたと考える。また、大きな違いはないものの、DF-Bと比べればDF-Aに利点がある。ただし、想定される試験薬の効果や検証したい仮説に応じて評価方法を検討し、意思決定ルールや棄却限界値の組み合わせ、各群の割付比などを設定する必要がある。また、症例数の再計算を行う場合は、計画時より症例数を増やすことを念頭に計画し、再計算の時期も慎重に検討する必要があると考えられる。本研究の成果は、より効率的な臨床試験計画の参考となり、少ない患者負担でよりよい医薬品を医療現場に提供することに貢献が期待される。

論文審査の結果の要旨及び担当者

氏 名 (落 合 俊 充)		
論文審査担 当者	(職)	氏 名
	主 査 教 授	大野 ゆう子
	副 査 招へい教授	濱崎 俊光
	副 査 教 授	大橋 一友

論文審査の結果の要旨

医薬品開発の過程では臨床試験が不可欠であり、試験薬と標準薬を用いた2群非劣性試験は非劣性検証でよく用いられる試験デザインであるが、標準薬の有効性について検証された頃とは許容される併用療法に変化が生じていたり、エンドポイントや患者の組み入れ基準に違いが見られたりする場合は、過去の試験と同様に標準薬の有効性が認められるとは限らないため、試験薬の有効性について述べることは困難になり得る。この問題に対処する一つの方法として、ICH（日米EU医薬品規制調和国際会議）等のガイドラインでは、試験薬と標準薬と共にプラセボ群を設けた3群非劣性試験が提案されている。本研究では、この3群非劣性試験に着目し、症例数の軽減につながる可能性のある、試験途中で有効性中止を目的とした群逐次デザインの適用を通して、より効率的な試験デザインのあり方について検討した。

3群非劣性試験での統計的評価は、固定マージン法とFraction法の2つに大きく分類される。固定マージン法は、非劣性マージンを事前に定め、「プラセボの効果」<「非劣性マージンを差し引いた標準薬の効果」<「試験薬の効果」という大小関係の成立を前提とする評価方法である。一方、Fraction法は、試験薬が標準薬の効果の一定割合を保持していることをもって非劣性を検証する方法である。本研究では、これら既存の評価方法を群逐次デザインで利用できるように拡張し、意思決定ルールA：最終解析までのいずれかの時点で分析感度（プラセボに対して標準薬の有効性が認められること）と非劣性がそれぞれ検証された場合に試験成功とみなす（非劣性の検証は分析感度が成立した時点以降で行う）、意思決定ルールB：同一解析時点で分析感度と非劣性が同時に検証された場合に試験成功とみなす、の2通りについて検討した。なお、中間解析時に分析感度や非劣性の成立の判断基準となる棄却限界値には、O'Brien-Fleming型境界と Pocock型境界とを組み合わせることを考えた。これらの条件下で期待症例数や検出力、Type I errorの挙動を数値的に評価し、各試験デザインの特徴を考察した。

解析回数を全4回とし、試験薬群と標準薬群の効果を共に10、プラセボ群の効果を5、各群共通の標準偏差を6.5と仮定したもとの検討した結果、ルールAでの期待症例数はルールBと比べてわずかながら（最大9例）少なくなる傾向がみられた。各群に等例数を割付け、非劣性マージンを2.5とする固定マージン法の場合、中間解析を行わない標準的な固定標本デザインで検出力80%を確保するために必要な症例数は426例であるが、ルールAあるいはルールBの群逐次デザインを用いることで症例数の軽減（最大38例）につながると考えられた。また、ルールAで固定マージン法を用いた場合、中間解析時に分析感度が成立した際には、その時点でプラセボ群の集積を打ち切ることが可能となる。このオプションを含めて期待症例数を算出すると371~399例と症例数をより少なくできると考えられた。さらに、試験途中で症例数の再計算を許容する群逐次デザインについて検討した結果、症例数を減らすことのみ考慮するデザインは、期待症例数は少なくなるものの、試験全体の検出力は低下する可能性があるため、適切な選択肢ではないこと

が示唆された。

これらの結果は仮定した条件下で得られたものであることに注意が必要であるが、本研究により、ルールA,B下での群逐次デザインは、標準的な固定標本デザインと比べて少ない症例数で早期に試験薬の効果を検証できる可能性を有することが示されたと考える。ただし、想定される試験薬の効果や検証したい仮説に応じて評価方法を検討し、意思決定ルールや棄却限界値の組み合わせ、各群の割付比などを設定する必要があり、症例数の再計算を行う場合は計画時より症例数を増やすことを念頭に計画し、再計算の時期も慎重に検討する必要があると考えられた。これらの成果は、より効率的な臨床試験計画の参考となり、少ない患者負担でよりよい医薬品を医療現場に提供することに多大な貢献が期待される。

以上より、本論文は博士（保健学）の学位授与に値するものとする。