

機関番号：82602

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500267

研究課題名（和文）リスクとベネフィットを定量的に統合評価する臨床試験のデザインに関する統計学的研究

研究課題名（英文）Statistical research on the design of clinical trials where the quantitative balance of benefits and risks is taken into account.

研究代表者

西川 正子（ NISHIKAWA MASAKO ）

国立保健医療科学院・技術評価部・統計高度利用室長

研究者番号：50373395

研究成果の概要（和文）：本研究は、様々な疾患を持つ患者に対して、最も有効で安全な治療法を迫るための適切な臨床試験デザインの方法論を提示することを目標とし、治療方法の有効性および安全性を正確かつ精度良く検定し得る評価方法を確立した。具体的には、従来ほど多くの被験者を必要としないで、従来よりも短期間で、有効性（ベネフィット）と安全性（毒性・リスク）のバランスを定量的に反映させて、それらを同時に正確に評価する検定方法を確立した。

研究成果の概要（英文）： This research aimed at proposal of an appropriate design of clinical trials to establish most effective and safe treatment for patients suffering from various diseases. We propose the methods which take the quantitative balance between efficacy and safety into account. The methods need less number of patients than conventional methods. As a result, clinical trial with the proposed methods will be completed in shorter period.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,000,000	0	1,000,000
2009年度	1,300,000	0	1,300,000
2010年度	1,100,000	0	1,100,000
年度			
年度			
総計	3,400,000	0	3,400,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：医療・福祉、医薬生物統計、臨床試験

1. 研究開始当初の背景

根拠に基づく医療の大切さが認識され、特に、承認申請のための治験は臨床試験の実施の基準（GCP、厚生労働省令）という法的な規制があり、臨床試験の計画・実施・管理・評価も厳格になされつつある。科学面では、統計ガイドラインに臨床試験のための統計的原則が示されている。そこでは、偏りを極力抑えること、および第1種の過誤を増大さ

せないことを求めている。第1種の過誤を増大させる原因の1つとして検定の多重性があり、多重性の調整が必要となる。そのため、統計ガイドラインでは「臨床試験の主要な評価項目は1つに絞る」ことを薦めている。そのため、治療の有効性に関する評価項目が一義的に採用され、安全性は相対的に重視されない傾向にある。その結果、例えば、新型インフルエンザの治療薬タミフルのように承

認後に重大な安全性の問題が出現する場合も見られる。また、関節リウマチでは、近年、生物学的製剤と呼ばれる非常に有効な治療薬が出現しているが、それらは間質性肺炎などの重篤な有害反応を頻繁に引き起こすという問題がある。関節リウマチでは臨床症状の改善に関する有効率で有効性を評価し、有害事象の発現率で安全性を評価することが多いが、本来は、有効性と安全性のバランスを医学的に勘案し、これらを定量的に統合評価する必要があると指摘されている。しかし、これまではそれに答える方法がなく、既存の方法として、有効性と安全性それぞれについて1つずつ逐次的に多重検定を行う方法や、有効性と安全性から「合成エンドポイント（評価指標）」を評価する方法があるが、検出力が低い上に有効性と安全性の定量的なバランスを勘案できないという欠点がある。

これを改良するため、研究代表者らは、複数の計量値から合成されるエンドポイント、および複数のイベント発現（死亡、心筋梗塞、脳卒中、心疾患の理由による入院など）までの時間から合成されるエンドポイントについてそれぞれの項目の医学的重大さを反映した検定方法を考案した。これに対し、Prof. Senn など著名な研究者たちから、このような医学的観点を取り入れた評価方法が必要であるというコメントがあり、医学的観点を勘案した評価方法の重要性を改めて認識した。本研究では、この成果を応用・発展させ、リスクとベネフィットの相対的なバランスを勘案し、これらを統合して1回の検定で結果が出せるような検定方法を確立する。

2. 研究の目的

様々な疾患を持つ患者に対して、最も有効で安全な治療・予防を施すことは医学の使命である。最適な治療方法を確立するためには、その有効性・安全性を確認する信頼性の高い臨床試験の開発が必須である。本研究の最終目標は、様々な疾患に対する適切な臨床試験デザインについての方法論を追求・提示し、治療方法の有効性および安全性を正確かつ精度良く検定し得る評価方法を確立することである。この観点から従来の臨床試験デザインを見直すと、検定の多重性という問題のために多くの被験者や長い期間を必要としたことが検討事項として挙げられる。本研究の具体的な目的は、この問題を改善し従来ほど多くの被験者を必要としないで、しかも従来よりも短時間で、正確に有効性（ベネフィット）と安全性（毒性・リスク）の両方をそれらのバランスに基づいて統合し同時に評価する検定方法を確立することである。その際、バランスの重みを定量化する方法、一つの検定結果を出すための成分の統合方法が重要な課題となる。

3. 研究の方法

(1) 治療法の有効性を有効率で、安全性を副作用の発現率で評価して、リスクとベネフィットの定量的なバランスを勘案し、医学的重大さを加味する新しい評価方法を確立するために、研究代表者らが行った「医学的な重要性を加味した臨床試験のデザインと評価方法に関する統計学的研究」の研究成果が適用できるように、sphericityの性質を本研究に応用することを検討する。

(2) 有効性と安全性で変数のタイプ（有効率および発現率・計量値・イベント発現までの時間）が異なる組合せの場合、計量値とイベント発現までの時間については前研究の成果が応用できると考えられる。これに対し、有効率や発現率のような割合の組合せについては簡単に応用できないので、sphericityの条件の緩和を検討する。これまでの結果を国際学会で発表し、海外の研究者と意見交換および情報収集を行う。

(3) リスクとベネフィットの定量的なバランスについては、従来ほとんど考えられていないので、医学分野での見解を調査し、その見解を検定統計量の定式化に反映できるように工夫する。調査の方法としては、関連する治療やQOLの分野の文献調査、医学専門家（医師など）との会合、および医学系の関連する学会を聴講することにより情報を収集する。さらに、医療経済学的観点からの文献調査も行う。その分析結果を基に、考案する検定統計量を定式化する。考案した方法の理論に基づき解析プログラムを開発し、simulationにより開発手法の性能の評価を行い、開発の基礎情報を得る。

(4) 研究代表者らの先行研究の成果を別な側面から応用することを検討する。すなわち、イベント発現までの時間の評価をイベントを発現した被験者の割合という見方をし、合成エンドポイントの新しい構成方法を検討し、考案する検定統計量に反映する方法を検討する。これまでの結果を国内学会や国際学会で発表し、本研究について国内や海外の研究者と意見交換する。

(5) さきに作成した解析プログラムを改訂してsimulationを行い、リスクとベネフィットの定量的なバランスの与え方を変化させて、第1種の過誤や検出力の改善について考察し、検定統計量の挙動を分析する。その結果に基づいて、リスクとベネフィットの定量的なバランスの与え方の方針を検討し、それを検定統計量の挙動に反映させるような調整方法を統計的な観点から考案する。検定統計量の漸近的な性質の検討を行うために、多変量対称分布の性質を応用した証明方法を発展させ、本方法の漸近的な性質の証明を行う。公表された実データを用いて、医学的見

解を反映させたリスクとベネフィットの定量的なバランスに関する情報を与え、本方法の挙動のテストを行う。

(6) リスクとベネフィットの定量的なバランスについて、引き続き、医学分野の学会に参加し見解を調査したり、文献調査などを行ったりして情報収集し、検討を続ける。

(7) ブートストラップ法を応用することについて検討を行う。帰無仮説下の状況を作ってブートストラップ法を用いて検定統計量の分布を見る際に、本研究における帰無仮説下の状況としては無限個数が想定されるので、それらからどのような状況を選択すれば帰無仮説全体を代表することになるのか理論的な検討を行い、解析プログラムを作成してsimulationにより検定統計量の挙動を分析する。一方、帰無仮説の検定に信頼区間を応用することも検討する。1変量の割合の差の場合についてのブートストラップ法を応用した信頼区間の構成に関する先行研究を、本研究における2変量の割合の差の場合に拡張することを検討する。理論展開を行い、解析プログラムを作成してsimulationにより第1種の過誤や検出力について考察し、2変量の信頼領域を利用した帰無仮説の検定の挙動を分析する。

(8) リスクとベネフィットの定量的なバランスの与え方に柔軟性を与えることを検討する。これまでは主に定量的なバランスには比例関係が存在することを仮定したが、その条件を緩くした条件で表現し、検定統計量の挙動に反映させるような調整方法を統計的な観点から考案する。公表された実データを用いて、医学的見解を反映させたようなリスクとベネフィットの定量的なバランスに関する情報を与え、本方法の挙動のテストを行う。得られた知見を国内学会や国際学会で発表し、本研究について国内や海外の研究者と意見交換する。

4. 研究成果

(1) 主な成果

従来、臨床試験デザインでは、治療法の有効性（ベネフィット）および安全性（毒性・リスク）を正確かつ精度良く検定しようとすると、複数のエンドポイントを設定することが必要となる。それにより、検定の多重性という問題が発生し、そのため多くの被験者や長い試験期間を必要とした。

本研究の目的は、この問題を改善し、従来ほど多くの被験者を必要としないで、しかも従来よりも短期間で、正確に有効性と安全性の両方を、それらのバランスに基づいて統合し同時に評価する検定方法を確立することである。そのために、治療法の有効性と安全性のバランスを定量化する方法、および一つの検定結果を出すための成分の統合方法が

重要な課題となり、この課題の解決に主眼を置いた。

① 治療法の有効性を有効率（割合）で、安全性を副作用の発現率（割合）で評価する場合、リスクとベネフィットの定量的なバランスを勘案し、医学的重大さを加味する新しい評価方法を確立するために、研究代表者らの、平成18-19年度の「医学的な重要性を加味した臨床試験のデザインと評価方法に関する統計学的研究」の研究成果が適用できるように、sphericityの性質を応用することを検討した。有効性と安全性で変数のタイプ（有効率および発現率・計量値・イベント発現までの時間）が異なる組合せの場合、計量値とイベント発現までの時間については前研究の成果が応用でき、新しい検定統計量を考案してその解析プログラムを作成した。一方、有効率や発現率のような割合の組合せについては簡単に応用できないので、sphericityの条件の緩和を検討した。複数のエンドポイントが割合である場合の検定の多重性の調整に関する文献調査を行った。これらの成果を中間発表として国際学会で発表し[学会発表⑦]、本研究について海外の研究者と意見交換および情報収集を行った。

② リスクとベネフィットの定量的なバランスについては従来まったく考えられていないので、医学分野での見解を調査するために、関連する治療やQOLの分野の文献調査、医学専門家（医師など）との会合、医学系関連学会において情報収集し、検定統計量の構築・分析と解析プログラムの開発を行った。

①で考案した検定統計量についての解析プログラムを改訂してsimulationを行い、リスクとベネフィットの定量的なバランスの与え方を変化させて、第1種の過誤や検出力の改善について考察し、検定統計量の挙動を分析した。その結果、リスクとベネフィットの重みに比例して重大なエンドポイントの結果が反映されるとは限らないことがわかった。意味がある定量的な相対的重みには上限があると考えられた。検定統計量の改良のために、研究代表者らの、平成18-19年の医学的な重要性を加味した臨床試験のデザインと評価方法に関する研究の応用として、イベントを発現した被験者の割合という見方から、オッズ比として取り扱うことを検討した。合成エンドポイントとしては有効率と毒性発現率についての治療群の間でのオッズ比を、ベネフィットとリスクのバランスを反映して合成することを検討した。また、割合のまま取り扱う場合も検討した。割合の推定値の分散は割合の大きさに依存するので、医学的バランスを取り入れた統計量の帰無仮説下の分布の分位点を定めるのは難し

いことが分かった。そこで、分散を式により計算する代わりにブートストラップ法を応用することについて検討を開始した。

これまでの知見を国内学会や国際学会で発表し[雑誌論文①、学会発表⑤、⑥]、本研究について国内や海外の研究者と意見交換した。そこで、リスクとベネフィットの定量的なバランスを考えるための方法としては、基本情報としてその治療領域や治療法などに関する文献のシステムティックレビューやメタアナリシスを行うことが有用であると考えられた。

③ ②に引き続いてブートストラップ法を応用することについて検討を行った。帰無仮説下の状況を作って、ブートストラップ法を用いて検定統計量の分布を見ることが通常多くなされているが、本研究における帰無仮説下の状況としては無限個数が想定され、代表的な帰無仮説について検討を行った。解析プログラムを作成して小規模の simulation により検定統計量の挙動を分析した。一方、帰無仮説の検定に信頼区間を応用することも検討した。1変量の割合の差の場合についてのブートストラップ法を応用した信頼区間の構成に関する先行研究を、本研究における2変量の割合の差の場合に拡張することを検討した。理論展開を行い、解析プログラムを作成して小規模の simulation により第1種の過誤や検出力について考察し、2変量の信頼領域を利用した帰無仮説の検定の挙動を分析した。

提案する方法により、リスクとベネフィットの定量的なバランスには単調性の自然な仮定のみを必要とし、従来法よりもバランスの与え方に柔軟性を持たせることが可能になった。対立仮説の領域がより柔軟に設定できるようになったので、リスクに比べてベネフィットが非常に大きな場合や、大きなベネフィットがあるけれど若干のリスクもあるような場合は、臨床試験における被験者数は従来よりも少なくできる。小規模の仮想データを用いて、医学的見解を反映させたリスクとベネフィットの定量的なバランスに関する情報を与え、本方法の挙動テストを行い、挙動の妥当性を確認した。

これまでの知見を国内学会や国際学会で発表し[学会発表①、②]、本研究について国内や海外の研究者と意見交換した。

(2) 国内外における位置づけとインパクト

従来の有効性と安全性の評価方法では、有効性を優先して安全性の評価がおろそかになったり、第1種の過誤を調整することによって、十分な検出力が保証できなかつたり、また、多くの被験者、さらに多くの時間とコストを必要とした。

研究代表者らは、これらの問題を解決するために先行研究において、医学的的重大さを反

映する検定方法として、複数の計量値から合成されるエンドポイント、および複数のイベント発現までの時間から合成されるエンドポイントについての新しい検定方法を開発し、国際学会で発表して高い評価を得た。本研究ではこの点を発展させ、割合・計量値・イベント発現までの時間からなる複数のエンドポイントで構成される合成エンドポイントの各成分のリスクとベネフィットの定量的なバランスを勘案し、医学的的重大さに基づく重みを付けてこれらを統合し、1回の検定で結果が出せるような検定方法を考案した。

この新しい評価方法では、1. 検定の多重性に起因する第1種の過誤の増大が防止され、検出力の高い検定が可能になる。2. リスクに見合わないベネフィットしかない治療や予防方法の早期の無効中止決定も本方法の中間解析への応用により可能になる。例えば、WHI 試験では試験開始から約 16600 人に治療を割付け、9年目の中間解析でリスクがベネフィットより大きいとして無効中止を決めたが、本研究の成果が応用されればもっと早期に無効中止の決定ができると考えられる。3. リスクに比べてベネフィットが非常に大きな場合は臨床試験における被験者数は従来よりも少なくすむ。よって、この新しい評価方法に基づくリスクとベネフィットのバランスを勘案する臨床試験のデザインが可能となる意義は極めて大きい。また、この結果を世界に情報発信することにより、国際的にも大いに貢献できる。

(3) 今後の展望

医学分野で本方法の紹介を行ったり、臨床医との共同研究を行ったりして、本方法が実際の臨床試験において利用され特長が発揮できるように医学系での共同発表など広報を行う。本方法の中間解析への応用のための理論開発を行う。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

- ① Nishikawa M., Tango T., Ohtaki M., Statistical Tests Based on New Composite Hypotheses in Clinical Trials Reflecting the Relative Clinical Importance of Multiple Endpoints Quantitatively, Biometrical Journal, 査読有, Vol. 51, 2009, pp.749-762.
- ② 西川正子、生存時間解析における競合リスクモデル、計量生物学会、査読有、29 巻、2008.12、p. 141-170

[学会発表] (計 7 件)

- ① Nishikawa M., Tango T., Ohtaki M. and Hida E., One sided-tests based on efficacy-toxicity trade-off in randomized controlled trials, XXVth International Biometric Conference, 2010.12.10, Florianopolis, Brasil
- ② 西川 正子、丹後 俊郎、大瀧 慈、飛田英祐、臨床比較試験における有効性と安全性のトレードオフを加味した仮説検定、2010 年度 統計関連学会連合大会、2010.9.6、新宿区
- ③ 中川智文、西川正子、丹後俊郎、抗癌剤第Ⅱ相臨床試験の多段階デザインにおける早期無効中止基準に関する検討、2010 年度 統計関連学会連合大会、2010.9.6、新宿区
- ④ 川口修、飛田英祐、西川正子、丹後俊郎、部分的区間打ち切りデータにおける二標本検定法の比較と評価について、2010 年度日本計量生物学会年会、2010.5.21、東京都立川市
- ⑤ Nishikawa M., Tango T., Ohtaki M., Hida E., New Composite Hypotheses and Statistical Tests in Randomized Clinical Trial where two time to event data are primary Endpoints, East Asia Regional Biometric Conference, 2010.2.12, Manipal, India
- ⑥ 西川正子、丹後俊郎、大瀧 慈、飛田英祐、臨床比較試験における 2 つのイベント発現までの時間の相対的な医学的重さを考慮した複合仮説と検定、2009 年度統計関連学会連合大会、2009.9.7、京田辺市
- ⑦ Nishikawa M., Tango T., Ohtaki M., Statistical tests based on new composite hypotheses for time to event data in clinical trials reflecting the relative clinical importance. XXIVth International Biometric Conference, 2008,7.15, Dublin, Ireland

[図書] (計 4 件)

- ① 西川正子、朝倉書店、区間打ち切りデータの解析. In: 医学統計学の事典(丹後俊郎, 小西貞則, 編集)、2010、pp.256-257
- ② 西川正子、朝倉書店、競合リスクモデル. In: 医学統計学の事典(丹後俊郎, 小西貞則, 編集)、2010、pp.252-255
- ③ 西川正子、朝倉書店、複数のエンドポイントの包括的検定. In: 医学統計学の事典(丹後俊郎, 小西貞則, 編集)、2010、pp.74-75
- ④ 西川正子、朝倉書店、有害事象の解析. In: 医学統計学の事典(丹後俊郎, 小西貞則, 編集)、2010、pp.64-65

6. 研究組織

(1) 研究代表者

西川 正子 (NISHIKAWA MASAKO)
 国立保健医療科学院・技術評価部・統計高度利用室長
 研究者番号：50373395

(2) 研究分担者

該当せず ()
 研究者番号：

(3) 連携研究者

丹後 俊郎 (TANGO TOSHIRO)
 昭和女子大学大学院・生活機構研究科・客員教授
 研究者番号：70124477
 飛田 英祐 (HIDA EISUKE) 国立保健医療科学院・技術評価部・主任研究官
 研究者番号：30469952