

Reporting of Multi-Arm Parallel-Group Randomized Trials Extension of the CONSORT 2010 Statement の和訳

著者 Edmund Juszczak, MSc¹; Douglas G. Altman, DSc²; Sally Hopewell, DPhil²; Kenneth Schulz, PhD^{3,4}

翻訳 馬場 亜沙美 (BABA Asami)^{1*}, 鈴木 直子 (SUZUKI Naoko)¹, 野田 和彦 (NODA Kazuhiko)¹,
波多野 絵梨 (HATANO Eri)¹, 高橋 徳行 (TAKAHASHI Noriyuki)¹, 中村 駿一 (NAKAMURA Shunichi)¹,
LIU XUN¹, 柿沼 俊光 (KAKINUMA Toshihiro)¹, 山本 和雄 (YAMAMOTO Kazuo)¹

Key Words: ヒト臨床試験, ヒト試験, 特定保健用食品, 機能性表示食品, 研究デザイン, CONSORT, ガイドライン, アウトカム, SPIRIT

Japanese translation of “Reporting of Multi-Arm Parallel-Group Randomized Trials Extension of the CONSORT 2010 Statement”

Authors: Edmund Juszczak, MSc¹; Douglas G. Altman, DSc²; Sally Hopewell, DPhil²; Kenneth Schulz, PhD^{3,4}

Translator: Asami Baba^{1*}, Naoko Suzuki¹, Kazuhiko Noda¹, Eri Hatano¹, Noriyuki Takahashi¹,
Shunichi Nakamura¹, Xun Liu¹, Toshihiro Kakinuma¹, Kazuo Yamamoto¹

Keywords: human clinical trial, Human trials, Food for specified health uses, Food with function claims, Research Design,
CONSORT, guideline, outcome, SPIRIT

*Correspondence author: Asami Baba

Affiliation (Author)

¹ NPEU Clinical Trials Unit, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

² Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

³ FHI 360, Durham, North Carolina

⁴ The University of North Carolina at Chapel Hill, School of Medicine

Affiliated institution (Translator)

¹ ORTHOMEDICO Inc. [2F Sumitomo Fudosan Korakuen Bldg., 1-4-1 Koishikawa, Bunkyo-ku, Tokyo, 112-0002, Japan.]

所属 (著者) :

¹ NPEU Clinical Trials Unit, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

² Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

³ FHI 360, Durham, North Carolina

⁴ The University of North Carolina at Chapel Hill, School of Medicine

* 責任著者 : 馬場 亜沙美

所属 (翻訳) : ¹ 株式会社オルトメディコ (〒112-0002 東京都文京区小石川 1-4-1 住友不動産後楽園ビル 2 階)

本項について

本稿は、EQUATOR Network が提供するガイドラインの一つである「Reporting of Multi-Arm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement」の本文を翻訳したものである。また、表は「Reporting of Multi-Arm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement」で拡張された項のみを抽出した。

キーポイント [Key Points]

問題 [Question]

並行群間比較デザインを用いている3群以上の多群間ランダム化比較試験を報告する場合、どのような追加情報を提供すべきか？

結果 [Findings]

この報告ガイドラインは、Consolidated Standards of Reporting Trials (CONSORT) 2010 声明を拡張したものである。CONSORT の10項目が修正され、各拡張項目について、適切な報告例と付随する説明が示されている。

意義 [Meaning]

ガイドラインのチェックリストは、多群間ランダム化比較試験の透明性のある報告を促進し、厳密性と再現性の評価を助け、方法論の理解を深め、臨床医、ジャーナル編集者、査読者、ガイドライン作成者、資金提供者における結果の有用性を高めるのに役立つ可能性がある。

抄録 [Abstract]

重要性 [Importance]

ランダム化比較臨床試験における報告の質は最適とは言えない。研究の透明性を高めることが最も重要な時代において、不十分な報告は試験結果の信頼性と妥当性の評価を妨げている。Consolidated Standards of Reporting Trials (CONSORT) 2010 声明は、ランダム化比較臨床試験の報告を改善するために作成されたが、その主な焦点は2群の並行群間比較試験であった。3群以上の並行群間比較デザイン（通常は等確率で、参加者をいずれかの治療群に同時にランダムに割り付け、治療を比較する）を用いる多群間比較試験は比較的良好である。多群間比較試験における報告の質は大きく異なるため、判断や解釈を困難にしている。CONSORT 2010 声明の要素の大部分は、多群間比較試験にも同様に適用されるが、いくつかの要素には多群間比較への適応が必要であり、場合によっては、さらなる課題を明確にする必要がある。

目的 [Objective]

多群間比較試験の報告を容易にするため、CONSORT 2010 声明の拡張版を示すこと。

デザイン [Design]

2014年のCONSORTグループ会議の後、全著者を含む「ガイドライン作成グループ」が結成された。著者らは2014年から2018年までの間、隔月で直接もしくはテレビ会議によって会合を行い、チェックリストと付随する文章を作成・改訂し、さらに電子メールで追加の議論を行った。草稿は、より広範なCONSORTグループのメンバー36名と、他に選ばれた臨床試験の専門家として知られる5名に回覧され、レビューを受けた。14名から幅広いフィードバックが寄せられ、それらのコメントを詳細に検討した後、拡張版の最終改訂版が作成された。

結果 [Findings]

この多群間比較試験用の CONSORT 拡張版は、CONSORT 2010 チェックリストの 10 項目を拡張し、優れた報告の例と各拡張項目の重要性の根拠を示している。主な推奨事項として、多群間比較試験は、多群間比較であることを特定されるべきであり、すべての治療群に関連する明確な目的と仮説が必要である。主要な治療の比較は特定されるべきであり、著者は多群から生じる計画および計画外の比較を完全かつ透明に報告すべきである。多重性の統計的調整が適用される場合は、その根拠と方法が記述されるべきである。

結論と関連性 [Conclusions and Relevance]

この CONSORT 2010 声明拡張版は、多群並行群間ランダム化比較臨床試験の報告に関する具体的なガイダンスを示し、このような試験の報告における透明性と正確性の向上に役立つはずである。

1. はじめに [Introduction]

多群間ランダム化比較臨床試験にはいくつかの形式があるが、一般的には、複数のアクティブな介入、アクティブな介入の組み合わせ、ある介入の異なる用量（またはレジメン）、プラセボ、非アクティブな介入、通常治療などの要素を組み合わせたものである。これらの要素は様々な方法で組み合わせることができ、その結果、可能な試験構造が多く考えられる。例えば、3つの治療群を有する試験において、「A1 対 A2 対 A3」は同じアクティブな介入の異なる用量の評価を表す。また、「A1 対 B1 対 C1」の試験は、2つの異なるアクティブな介入とプラセボの評価を表す。さらに、「A1 対 A2 対 B1」で比較する試験は、2つの異なる用量のアクティブな介入と別のアクティブな介入の評価を表す。

複数の新しい介入を同時に評価することで、効果的な介入を見出す可能性が高まる¹⁾。「多群間」（すなわち、複数の群）比較のデザインを使用することで、「2 群間」（すなわち、2つの群）比較試験を連続して実施する場合に比べて効率が当然高まるため、リソースの有効利用によりコストが低くなるはずである。ほとんどの場合、対照群を共有することで、別々に「2 群間」比較試験を実施するよりもサンプルサイズが小さくなる²⁾。参加者が新しい介入に割り付けられる確率を上げることで、適格者を組み入れる割合が高くなる可能性がある。がん領域における一部の多群間比較試験は、同等の 2 群間比較試験よりも募集が迅速に行われている¹⁾。多群間比較試験に対する反論は、主に統計的検出力に関わるものであり、これは出版された試験のサンプルサイズが不十分な可能性があるためである³⁾。潜在的な

参加者数が限られている場合、治療群を追加することは、すでに不十分な検出力をさらに低下させる可能性がある。

多群間比較試験は比較的一般的である。2012 年の 1 ヶ月間に発表され、PubMed に収録されたすべてのランダム化比較試験を詳細にレビューしたところ、1351 件中 1062 件（79%）が並行群間比較試験であった⁴⁾；この 1062 件のうち、149 件（14%）が 3 群間比較試験で、76 件（7%）が 4 群間以上の比較試験であった。

本稿では、Consolidated Standards of Reporting Trials (CONSORT) 2010 声明^{5,6)}に基づき、多群間比較試験報告用の CONSORT チェックリストの拡張版が示されている。CONSORT チェックリストの本体と異なる項目については、例示と説明が含まれている。多群間比較試験とは、並行群間比較デザインを用いたランダム化比較臨床試験のうち、3 群以上を有するものと定義される。臨床試験における介入群を説明するには、「群 (arms)」または「群 (groups)」という用語を同じ意味で使用可能だが、本報告ガイドラインでは「多群間 (multi-arm)」という用語を使用する。要因 (factorial) 試験や多群多段階 (multi-arm multi-stage) 試験、アダプティブ (adaptive) 試験など、その他の多群間比較デザインは、別の問題を引き起こすため、考慮しない。

2. ガイダンス作成方法 [Guidance Development Methods]

2-1. 作成グループ [Writing Group]

ガイドライン作成グループ (E.J., D.G.A., S.H., K.S. を含む) は、2014 年の CONSORT グループ会議

後に結成された。2014年から2018年にかけて、オックスフォードを拠点とする著者たちは、米国を拠点とする著者たちと2ヵ月に1回、直接もしくはテレビ会議で会い、その他の機会でも複数回会った。その際にチェックリストと付随する例や文章を作成・改訂し、さらに電子メールで追加の議論を行った。

2-2. 検索戦略 [Search Strategy]

多群間ランダム化比較試験の方法論に関連する論文を特定するため、PubMedで「multiarm」, 「multi-arm」, 「multiple arm」, 「multiple treatment」, 「multiplicity」の用語と、Publication Typeにおける「randomized controlled trial」の用語をトピックとして組み合わせて検索を行い、247報の候補論文を特定した。著者1名(S.H.)がタイトルと抄録を評価し、このCONSORT拡張版との関連性または関連する可能性を検討した。また、著者が個人的に収集した関連論文や臨床試験の方法論に関連する書籍の目次を検索して得られた多群間比較試験の実施と報告に関する情報を使用し、検索を補足した。

2-3. レビューと改良 [Review and Refinement]

このCONSORT拡張版チェックリストを作成するには、正式なデルファイ調査は行われなかった。草稿は2017年4月により広範なCONSORTグループに回覧され、そこに36名のメンバーと他に選ばれた5名の臨床試験の専門家が参加している。14名からフィードバックが寄せられ、そのコメントを詳細に検討した後、拡張版チェックリストの最終改訂版と付随する説明が作成された。

3. 結果 [Results]

3-1. チェックリストの項目と説明 [Checklist Items and Explanation]

表(本稿では、拡張された項のみを抽出しているため、変更のない項に関しては、「Reporting of Multi-Arm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement」の表を参照する)は、多群並行群間ランダム化比較試験を報告するために修正されたチェックリストを示している。多群間比較デザインによって複雑さが増すことを考慮し、そのデザインに関連する必要な報告事項に合わせていくつかの項目が拡張されている。CONSORT 2010

声明からの拡張が必要な項目については、優れた報告の例とともに説明されている。言及されていない項目については、アドバイスが2群並行ランダム化比較試験と同じである。

すべての例が出版された論文から引用されているため、優れた報告に望ましい要素をすべて示していないものが散見されるのは避けられない。その場合、あるいは判定が曖昧な場合には、優れた報告の具体的な側面が示されている。一部の例では、文脈を説明するために括弧内に文章を追加した。また、ランダム化比較試験の抄録を報告するためのCONSORT 2010チェックリストがレビューされた。別の抄録用チェックリストは提案されていない。ただし、著者がすべての目的を明確に報告し、治療群の数を明記するがその条件である。

4. 多群間比較試験のためのCONSORTチェックリストの拡張版 [CONSORT Checklist Extension for Multi-Arm Trials]

4-1. タイトルと抄録 [Title and Abstract]

4-1-1. 項目 1a. CONSORT 2010: タイトルにランダム化比較試験であることを記載する。 [Item 1a. CONSORT 2010: Identification as a randomized trial in the title.]

多群間比較試験の拡張版: 多群間ランダム化比較試験であることをタイトルに明記する、または参加者がランダムに割り付けられた治療群の数を示す。

「HARMONY 3: プラセボ、シタグリプチン、グリメピリドと比較し、メトホルミン服用中の2型糖尿病患者におけるアルビグルチドの有効性と安全性を評価する104週間のランダム化二重盲検プラセボおよび実薬対照比較試験⁹⁾。」

「緩和ケア患者のせん妄症状に対する経口リスペリドン、ハロペリドールまたはプラセボの有効性検討: ランダム化比較臨床試験¹⁰⁾」

「2次進行型多発性硬化症多群間ランダム化比較試験 (MS-SMART): 2次進行型多発性硬化症に対する3種類の神経保護薬の有効性を比較する多群間第IIb相ランダム化二重盲検プラセボ対照比較臨床試験¹¹⁾。」

表 1 多群並行群間ランダム化試験報告用のチェックリスト：CONSORT 2010 声明の拡張版^{a, b}

章 (Section/Topic)	項目番号 (Item No.)	CONSORT 2010 声明	多群間比較試験の拡張版
タイトルと抄録 [Title and abstract]			
	1a	タイトルにランダム化比較試験であることを記載する	多群間ランダム化比較試験であることをタイトルに明記する、または参加者がランダムに割り付けられた治療群の数を示す
	1b	試験デザイン、方法、結果、結論の構造化抄録（詳細はは、「雑誌および会議録でのランダム化試験報告の抄録に対する CONSORT 声明」を参照） ⁷⁾	治療群数の特定；治療群の追加・中止の詳細
はじめに [Introduction]			
背景と目的 [Background and objectives]	2a	科学的背景と論拠の説明	多群間デザインを用いる根拠
	2b	特定の目的または仮説	すべての治療群の研究課題を明記 検証されるすべての仮説と主要な比較を明確に記述
方法 [Methods]			
試験デザイン [Trial design]	3a	試験デザインの記述（並行群間、要因分析など）、割付比を含む	治療群数を明記
	3b	試験開始後の方法上の重要な変更（適格基準など）とその理由	追加または中止された治療群の詳細（該当する場合）とその理由、および/または割付比の変更
症例数 [Sample size]	7a	どのように目標症例数が決められたか	各主要な比較について、計画されたサンプルサイズとその決定方法の詳細
ランダム化 [Randomization]			
統計学的手法 [Statistical methods]	12a	主要・副次的アウトカムの群間比較に用いられた統計学的手法	多重性の調整が適用されなかった場合はその旨を明示し、調整が行われた場合はその方法を明記する
結果 [Results]			
募集 [Recruitment]	14a	参加者の募集期間と追跡期間を特定する日付	募集期間と追跡期間が治療群間で異なる場合（例えば、群の追加や中止）、募集期間と追跡期間、その差異の理由、統計学的関連事項を説明する
アウトカムと推定 [Outcomes and estimation]	17a	主要・副次的アウトカムのそれぞれについて、各群の結果、介入のエフェクト・サイズの推定とその精度（95% 信頼区間など）	事前に決めた各治療群の比較結果

略語：CONSORT = Consolidated Standards of Reporting Trials

a: このチェックリストは、各項目に関する重要な説明が得られるために、CONSORT 2010 声明の「解説と詳細」⁵⁾と合わせて確認することが強く推奨されている。

b: 「Reporting of Multi-Arm Parallel-Group Randomized Trials Extension of the CONSORT 2010 Statement」で拡張された項目のみを抽出した。

解説：電子データベースでランダム化比較試験の報告を特定できるかどうかは、その報告がどのように索引付けされたかに大きく依存する。索引作成者は、著者がそのような情報を明示していない場合、その報告をランダム化比較試験として分類しないことがある。研究が適切に索引付けされ、容易に特定されるようにするために、著者はタイトルに「ランダム化比較試験」という言葉を使い、参加者がランダムに割り付けられた群（治療群）の数を示すべきである¹²⁾。この問題は多群間比較試験にも当てはまる。論文のタイトルには通常、文字数制限があり、介入の一部またはすべてを列挙するのは煩雑であるため、代わりに「多群間」という単語を加えるのが効率的で参考にもなる。

4-2. 背景と目的 [Background and Objectives]

4-2-1. 項目 2a. CONSORT 2010: 科学的背景と論拠の説明 [Item 2a. CONSORT 2010: Scientific background and explanation of rationale.]

多群間比較試験の拡張版：多群間デザインを用いる根拠。

「多くの患者は単剤療法に反応せず、エビデンスがほとんどないにもかかわらず、薬剤の併用がよく推奨される。リチウムとバルプロ酸の併用療法は、ファーストラインの単剤療法が無効であった場合によく推奨される。もしこの併用療法に薬理学的な相加効果があり、単剤療法よりも優れていることが証明されれば、適切なファーストライン療法となりうる。我々はここで、BALANCE (Bipolar Affective disorder: Lithium/ANti-Convulsant Evaluation) 研究について報告する。これは、双極性障害 I 型の再発予防に対して、リチウムとバルプロ酸セミナトリウムの併用療法がどちらか一方の単独療法に比べて優れているかどうかを明らかにするために計画されたランダム化比較試験である¹³⁾。」

解説：試験が2つの並行群間で比較する場合、それらの群を比較することが目的であることは明らかである。しかし、3群以上の介入群では、意図している主な比較は明確でない場合がある。各介入群は特定の研究課題を検討すべきであるので、それぞれの群は少なくとも1つの事前に計画された比較を検

討する必要がある。著者は、多群間比較デザインの使用を正当化し、論文の序論で、なぜ研究された介入を調べることに決めたか、どの特定の比較が計画されたかを示すべきである。例えば、1つの計画された介入に2つのアクティブな介入を組み合わせた場合、著者は要因試験を行わなかった理由を説明すべきである。一般的に、このような“不完全な”要因計画は、患者に対してアクティブな治療を差し控えることが非倫理的である場合に用いられる。

4-2-2. 項目 2b. CONSORT 2010: 特定の目的または仮説 [Item 2b. CONSORT 2010: Specific objectives or hypotheses.]

多群間比較試験の拡張版：すべての治療群の研究課題を明記。検証されるすべての仮説と主要な比較を明確に記述。

「要約（目的）：緩和ケアを受けている患者において、プラセボと比較して、苦痛を伴うせん妄の標的症狀の緩和におけるリスペリドンまたはハロペリドールの有効性を検討する。」

「はじめに：この研究の目的は、リスペリドンまたはハロペリドールを、せん妄発症リスクの管理 (managing precipitants of delirium) と個別化された支持的看護ケアと併用する場合、プラセボと比較して、苦痛を伴うせん妄の標的症狀の軽減にさらなる効果が得られるかどうかを明らかにすることであった。主要な帰無仮説は、リスペリドンとプラセボの間に差がないことであり、副次的な帰無仮説は、ハロペリドールとプラセボの間に差がないことであった¹⁰⁾。」

解説：3群間比較試験（A群、B群、C群）から8通りの解析が可能であり、ほとんどの試験では2群または3群が含まれる (Box)。比較可能な回数は、介入群の数が増えるにつれて急速に増加する；各群は少なくとも1つの比較に現れるべきである。従って、全群を一度に比較することのみを意図しない限り (用量反応試験以外、これは適切な方法ではない)、 k 個の治療群を持つ試験の解析では、少なくとも $k-1$ 通りの比較を行うことができる。2群/ペア比較の最大回数は $k \times (k-1) / 2$ である (例えば、4群

間比較試験では、6通りの2群間比較が考えられる)。

したがって、解析を事前に決めることが特に重要であり、著者は計画された主要な比較、副次的な比較、探索的な比較をすべて報告すべきである。そうでなければ、多数の可能な解析を考慮せずに、観察された差を強調し、それに迷わされるリスクが大いにある。すべての場合において、特に多くの比較が計画されている場合は、主要な比較を示すことが助けになる。これらの比較は、計画された症例数(項目7a)の説明にも記載されるべきである。計画された比較が同じように重要であるとは限らない。例えば、1つの2群間比較が試験の主な焦点となる場合がある。この特徴は、多重比較の調整を行うかどうかを検討する際に重要である。また、仮説検定に階層的方法を用いれば、多重比較の問題を防ぐことができる(項目12a)。優越性と非劣性の検証を組み合わせた多群間比較試験もある。例えば、Foaらは、心的外傷後ストレス障害の現役軍人における症状の重症度を軽減するために、2週間にわたって10セッションの長時間暴露療法(トラウマに焦点を当てた認知行動療法)を行うこと(集中型療法)が、最小限の接触(コントロール)よりも有効であり、8週間にわたって10セッションの長時間暴露療法を行うこと(間隔型療法)よりも劣っていないかどうかを検討した¹⁶⁾。

Box. 多群間ランダム化比較試験の方法論的問題点

デザイン [Design]

研究目的 [Research Objectives]

2群より多い治療群を有する試験は一般的に、2群間比較試験よりも複雑な問題を検討し、通常、1つより多い介入に関する研究課題を検討しようと試みる。著者は、多重性の影響とデータ・ドレッシング(すなわち、計画されていない探索的解析)の疑いを部分的に軽減するために、試験におけるすべての群を明記し、計画されたすべての介入群の比較を事前に決め、多群間比較試験の目的を明確に定義するべきである。

適格基準 [Eligibility Criteria]

複数の薬剤を使用する試験では、安全性・毒性プロファイルが参加候補者のプールを減らし、募集と一般化可能性に悪影響を及ぼす可能性がある。

患者および/または募集施設の特性、均衡性やリソースの不足により、いずれかの群にランダムに割り付けることができない場合がある。多群間比較試験では、禁忌のある2つの治療群が含まれているが、患者を他の群にランダムに割り付けることができる。例えば、すべての人がある外科手術には適さないが、ある薬剤の評価には参加できるかもしれないので、患者を「コントロール対手術対薬剤」、「コントロール対手術」(薬剤療法に適さない場合)、「コントロール対薬剤」(手術に適さない場合)としてランダムに割り付けることができる。

サンプルサイズ [Sample Size]

第3相の多群間比較試験のサンプルサイズは、計画された主要な比較によって異なるべきである。1群あたりのサンプルサイズは、事前に決めた主要な比較が十分な検出力を有するように十分大きくすべきである。

盲検化のためのプラセボの使用

[Use of Placebos for Blinding]

多群間比較デザインでは、群を識別できないように盲検化する必要がある。2つの実験薬剤の投与経路が異なる場合(例えば、経口投与と静脈内投与)、盲検化は侵襲的で費用がかかり、参加者にさらなる負担をかけることになる。実験薬剤群の数が増えれば増えるほど、盲検化がより問題となる可能性がある。

3群間比較試験：プラセボと実薬対照

[3-Arm Trial: Placebo and Active Control]

プラセボ群だけでなく、実薬対照群を含む3群間比較試験により、被験治療の効果をプラセボと区別できなかった場合、新しい被験治療が無効であったのか、それとも単に実薬を同定する能力が欠如していた結果なのかを確認するこ

とができる。このようなデザインにおけるプラセボと実薬（標準薬剤）の比較は、分析感度（有効な治療と有効性の低いあるいは無効な介入を区別する能力として定義される臨床試験の特性）の内的証拠を示している。実薬治療群の人数をプラセボ群より多めに設定し、実薬を比較する精度を向上させるために、不均等割付比を用いることができる。この割付比は、プラセボ群に割り付けられる確率が低くなるため、参加者と研究者に受け入れやすくなる可能性がある¹⁴⁾。

実施 [Conduct]

中間解析と中止のガイドライン

[Interim Analysis and Stopping Guidelines]

多くの臨床試験は、中間モニタリングと早期中止のガイドラインについて、正式な方法を採用している。これらのガイドラインは、有益性、害、あるいは無益性に関する強いエビデンスがある場合、早期募集中止の検討を促すものである。早期中止のガイドラインに基づいた中間解析の解釈は、複数の治療群の存在によって複雑になる。多群間比較試験の種類や構造によっては、中間解析の結果として、倫理的ジレンマが生じることがある。例えば、治療介入の一つがコントロール群に比べて有益であるという十分に強いエビデンスが認められた場合がある。この介入がコントロール群に対して有意な改善とみなされた場合、コントロール群の募集を中止しなければならない可能性があり、その結果、他の治療介入群の募集が、並行のコントロール群がないために中止される可能性がある。複数の治療介入とコントロール間の比較のいずれかが有効性の早期中止の境界を超えた場合、試験は中止される可能性があるため、有効性の境界には多重性の調整が必要である。

解析 [Analysis]

解析戦略 [Analysis Strategy]

介入がどのように異なるかを検討せず、異なるかどうかを検討するのが主な目的の場合は、1つの包括的な有意差検定を用いてすべての群を一度に比較するのが適切であろう。主な目的が傾向を調べることにある場合、用量反応モデル

を用いるべきである。通常、2つ以上の特定の比較が、特定のペアまたは治療の組み合わせの間で行われる。しかし、可能な比較の数は相当多くなる可能性がある。

複数の治療の比較

[Multiple Treatments Comparisons]

3群間比較試験（例えば、治療A, B, C）では、下記のように、いくつかの比較の可能性がある：

1. 3群すべてを一度に比較する（A対B対C）；順序付けされていない群の包括的な検定、または順序付けされた群間の傾向の検定。
2. 1つの群と組み合わせた他の2つの群（A+B対C）を比較し、次に組み合わせた群同士（A対B）を比較する；AとBは同じ薬剤の低用量と高用量であり、最初は治療群と未治療群を比較し、次に治療群の2群を比較する、あるいはAとB（同じクラスの2種類の抗生物質）、C（異なるクラスの抗生物質）を比較する（注：この例でのラベリングは任意である）。
3. すべてのペアごとの比較：A対B, A対C, B対C。
4. A対C, B対Cは比較するが、A対Bは比較しない。例えば、2つの治療をそれぞれコントロールと比較するが、2つの治療を互いに比較しない。

報告と解釈 [Reporting and Interpretation]

多群間比較試験は、複雑で込み入った問題を並行して扱うことが多いため、2群間比較試験とは異なる焦点が当てられる。例えば、事前に決めたすべての群の比較の後、統計的に有意な包括的な検定は率直な解釈とは言えない。研究者は、すべての介入が等しく効果的であったという仮説を否定するエビデンスを持っているが、差がどこにあるのかについては明確に示していない。観察されたデータそのものを用いてより正確な結論を導き出すことは、魅力的ではあるが間違っている。例えば、最も良好な結果を得た介入は他の介入よりも優れていると推論するのは正しくない、その理由は、この問題はまだ明確には検証されていないからである。

さらに、複数のペアごとの比較は、一見矛盾した結果をもたらすことがある。例えば、2つの

アクティブな介入（A と B）対プラセボの試験において、A はプラセボより有意に優れているが、B は A またはプラセボのいずれとも有意差がないことがありうる。また、有意な包括的な検定にもかかわらず、ペアごとの比較が有意でないこともありうる。このような問題は、農業研究など正式な多群間比較が一般的である研究分野ではよく知られているが、臨床研究ではほとんど経験がない。

比較の多重性に関する解釈の問題は、一般的なものである。臨床医は、ある群を試験に追加することで、得られる情報が減少するのではなく、むしろ増加することをよく感じる¹⁵⁾。このような臨床試験の多くでは、多重性の調整をした結果を解釈することは、解釈上の問題を解決するのではなく、むしろ問題を引き起こすことが多い。しかし、特定の解析で多重比較の調整が必要となる場合もある。そのような調整が実に不十分で過度に補正されやすいものであれば、著者は解釈においてその点を考慮すべきである。

多群間比較試験の報告書の読者は、複数の介入群から生じる主要な比較と副次的な比較がどのように扱われたのかの説明を期待するだろう。ほとんどの著者や読者は、正式な調整が行われたかどうかにかかわらず、実施された解析の数を念頭に置いていると思われる。

4-3. 方法 [Methods]

4-3-1. 項目 3a. CONSORT 2010：試験デザインの記述（並行群間、要因分析など）、割付比を含む。[Item 3a. CONSORT 2010：Description of trial design (such as parallel, factorial) including allocation ratio.]

多群間比較試験の拡張版：治療群数を明記。

この実用的な非盲検ランダム化試験では、新規にパーキンソン病と診断された患者を、レボドパスベアリング療法（ドパミンアゴニストまたは MAOBI [モノアミン酸化酵素 B 阻害薬]）とレボドパ単独療法にランダムに割り付けた（中央オフィスへの電話連絡により実施される、1:1:1)¹⁷⁾。

本試験は、2009年2月17日～2013年3月21日に実施された第3相のランダム化二重盲検プラセボ・実薬対照並行群間比較試験である。適格患者

は、HbA1c 値 (<8.0% [<63.9 mmol/mol] 対 $\geq 8.0\%$ [≥ 63.9 mmol/mol])、心筋梗塞 (MI) の既往歴、年齢 (<65 歳対 ≥ 65 歳) で層別化され、ランダムに割り付けられた (3:3:3:1)。バックグラウンドのメトホルミン投与に加え、ベースライン時にアルビグルチド 30 mg、シタグリプチン 100 mg、グリメピリド 2 mg、プラセボの4つのうち1つが投与された。アルビグルチド、シタグリプチン、グリメピリドに対して釣り合う用量のプラセボは、試験の盲検化を維持するために使用された⁹⁾。[なぜ 3:3:3:1 の割付比が使われたのかを説明することが、報告の改善点と考えられる]

解説：読みやすさを向上させることや多群間比較試験のデザインと根拠を把握する観点から、治療群数を明記することは不可欠である。割付比を記述することで、深い理解と明瞭さがもたらされ、特に不均等割付比が採用された場合は、説明が必要である。

通常、多群間比較試験の構造と参加者の流れを図示することで、読者に深い理解を与えることができる。試験の構造と参加者の流れを示す例を図に示す (図 A, B)⁹⁾。しかし、この例における試験の構造と参加者の流れを示す内容では、ラベリング (例えば、図 A における「追跡 (Follow-up)」の位置)、ランダム化ボックスから続く 2 本の矢印の欠如、提供される情報の説明 (例えば、図 B における「資金提供者により中止 (Terminated by sponsor)」とは何を意味するのか) などの点について改善される可能性がある。

4-3-2. 項目 3b. CONSORT 2010：試験開始後の方法上の重要な変更（適格基準など）とその理由。[Item 3b. CONSORT 2010: Important changes to methods after trial commencement (such as eligibility criteria), with reasons.]

多群間比較試験の拡張版：追加または中止された治療群の詳細（該当する場合）とその理由、および/または割付比の変更。

例（群を中止した場合）。

当初の試験は多施設盲検ランダム化並行群間比較試験で、患者は 4 週間の心理社会的治療を受けた後、

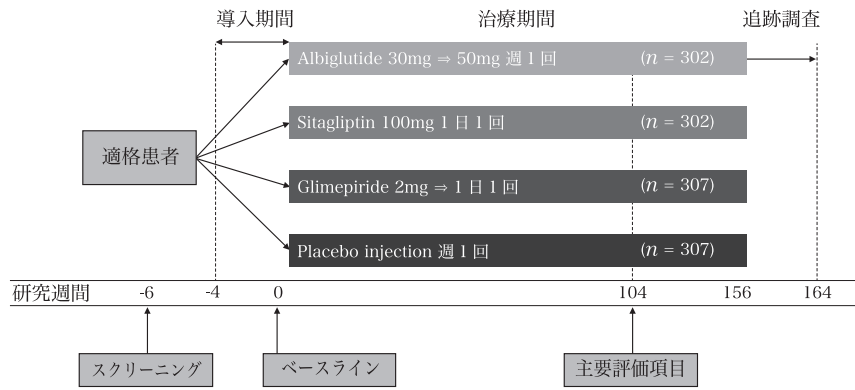


図 A. HARMONY 3 試験の構成

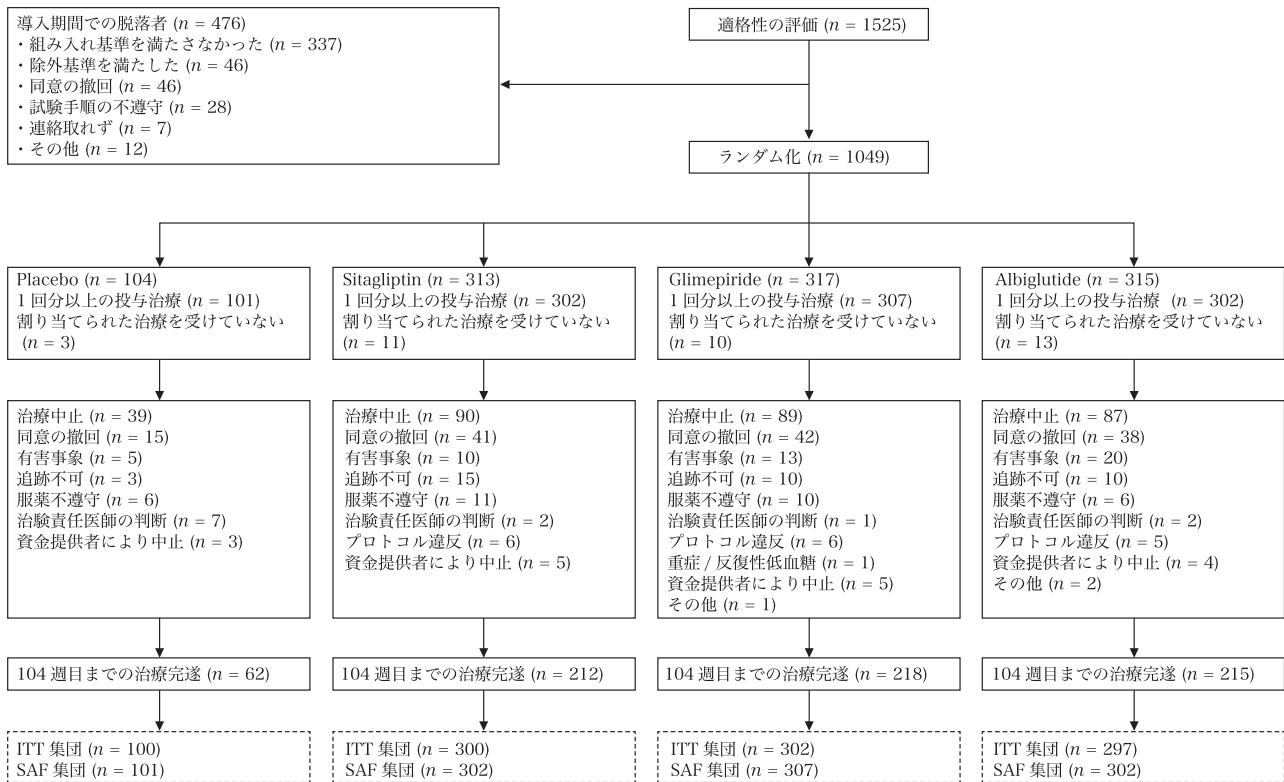


図 B. HARMONY 3 参加者フロー図

Juszczak E, Altman DG, Hopewell S, Schulz K. 多群並行群無作為化試験の報告：CONSORT 2010 声明の拡張。JAMA. doi:10.1001/jama.2019.3087

この補足資料は、読者に彼らの研究についての追加情報を提供するために、著者らによって提供されたものである。

リスペリドン（リスパダール，エーザイ），ドネペジルまたはプラセボを12週間投与する群に割り付けられた。目標のサンプルサイズは285名のアルツハイマー病患者であった。募集は2003年11月に開始されたが、英国医薬品安全性委員会（the United Kingdom Committee for Safety of Medicines）がリスペリドンとオランザピンを認知症の行動症状の治療に使用しないよう勧告したため、2004年3月に中断された。試験は2004年7月に2群のデザイン（ドネペジルとプラセボ）で再開され、2005年9月に

募集が終了した¹⁸⁾。

例（群を追加した場合）。

合計1493名の統合失調症患者を米国の57施設で募集し、オランザピン（1日7.5～30mg）、ペルフェナジン（1日8～32mg）、クエチアピン（1日200～800mg）、リスペリドン（1日1.5～6.0mg）のいずれかを最長18ヵ月間投与する群にランダムに割り付けた。ジプラシドン（1日40～160mg）は米国食品医薬品局（FDA）の承認後に追加された。主要

な目的は、これら5つの治療の全般的な有効性の違いを明らかにすることであった¹⁹⁾。

解説：治療群が追加または中止された場合、偏りがなく妥当な比較に利用できる参加者の数が影響される（すなわち、同時にランダム化された参加者のみが比較されるべきである）。上記の治療群が中止された例では、割り付け比率が1:1:1から1:1になったので（参加者の流れと結果表から一目瞭然）、どちらかの介入を受ける確率は0.33から0.50に変わったが、ランダム化は継続され、どちらかの介入を受ける確率がほぼ等しい状態でした。群が追加された例では、割り付け比率が明言されることはなかった。

この項目は従来の多群間比較試験に関するものであり、事前に決めた規則で群を中止できるアダプティブデザイン（adaptive design）に関するものではない。このようなデザインは、ランダム化が必要な参加者数を最小限に抑えながら、より高い効率を実現する。アダプティブデザイン試験の報告に関するガイドラインは、アダプティブデザインに関するCONSORTの拡張版でカバーされる²⁰⁾。

4-3-3. 項目7a. CONSORT 2010：どのように目標症例数が決められたか。 [Item 7a. CONSORT 2010: How sample size was determined.]

多群間比較試験の拡張版：各主要な比較について、計画されたサンプルサイズとその決定方法の詳細。

サンプルサイズの計算は、プラセボ治療患者の34%、タダラフィル治療患者（1日1回またはオンデマンド）の54～64%が、DFW（薬物を使用しないウォッシュアウト）の後にIIEF-EFスコア（国際勃起機能スコア-勃起機能）を達成するという仮定に基づいて行われた。ランダムに割り付けられた患者412名のサンプルサイズでは、タダラフィル（1日1回またはオンデマンド）対プラセボの2つのペアごとの比較で20%の割合での差を検出し、検出力が84%でした（20%の脱落率を想定）²¹⁾。

日常的な臨床診療を変えるためには高度の有益性が必要とされるため、コントロール群における推定

発症率11%（相対減少率30%；オッズ比[OR]0.67）に基づき、絶対減少率を3.3%と決めた。90%の検出力、2.5%の有意水準で2つの比較を想定し、15%の離脱率を考慮すると、各群に2,345名（全体で7,035名）の参加者を募集する必要がある。この試験では、2つの同等に重要な比較が検証された：銀合金カテーテル対PTFE [ポリテトラフルオロエチレン] カテーテル、ニトロフラールカテーテル対PTFEカテーテル²²⁾。

解説：多群間比較試験のサンプルサイズは、計画された主要な比較（項目2b）に対応すべきである。サンプルサイズの考え方は、比較される介入の構造と計画された解析の性質によって決定される（Box）。ペアごとの比較が計画されている場合、サンプルサイズは通常、意図された主要な比較をそれぞれ評価するのに十分な検出力を与えるように決定される。研究者が多重比較を調整する必要があると判断した場合、計画されたサンプルサイズはその調整を考慮して大きくすることができる（項目7a）。

4-3-4. 項目12a CONSORT 2010：主要・副次的アウトカムの群間比較に用いられた統計学的手法。 [Item 12a. CONSORT 2010: Statistical methods used to compare groups for primary and secondary outcomes.]

多群間比較試験の拡張版：多重性の調整が適用されなかった場合はその旨を明示し、調整が行われた場合はその方法を明記する。

例（調整が行われなかった場合）。

仮説としては、高心拍数型治療群、遅延型治療群、またはその両方が、通常型治療群と比較して、不適切な治療の初回発生リスクを減少させるということであった。2つの試験は並行して実施され、それぞれで推論が行われたため、多重比較の調整なしは適切であると判断された²³⁾。

P値はすべて両側で、多重比較の調整は行っていない²⁴⁾。

例（調整が行われた場合）。

尿路感染症のアウトカムをロジスティック回帰で評価し、結果を絶対リスク減少率およびORとして

要約し、いずれも 95% 信頼区間を 97.5% 信頼区間として算出し、2つの比較を調整した。主要な解析では、 $P=0.025$ を有意とみなした²²⁾。

VAS 型即時疼痛評価と圧力データの両方について、Shapiro-Wilk 正規性検定を通った場合、一元配置反復測定分散分析 (ANOVA) とボンフェローニ補正 (Bonferroni correction) 後にペアごとの比較を行い、試験条件間の有意差 ($P < 0.05$) を調べた²⁵⁾。

(前略) 我々は、Hochberg 多重検定法を用いて、0.05 の両側有意水準、90% の検出力で、アピキサバンのプラセボに対する優越性を示すためには、各群に 810 人の患者を組み入れる必要があると計算した²⁶⁾。

解説：一般的に、多群間比較試験の解析戦略には 2つの大きな目的がある。第 1 に、研究者は複数の介入の有効性の変動を検討するが、これはすべての群を一度に比較する全体的な解析で行うことができる。このような解析は、差のある領域を示していないため、十分とは言いがたい。第 2 に、上述のように特定の治療間で 2つ以上の特定のペアごとの比較を行うことができる。特定の試験では、両方のタイプの分析を行うことができる。1つの戦略としては(農学の分析やその他の試験で一般的に推奨されている)、最初にすべての群にわたって包括的な統計検定を行い、包括的な検定が統計的に有意な場合にのみペアごとの比較を行うということである。この戦略は、治療の比較により焦点を絞った評価方法を必要とする臨床試験の分析には、特に望ましくないように思われる。

さらに 2つの複雑な状況が考えられる。第 1 に、2つ (またはそれ以上) の治療が、同じ薬剤または介入の異なる用量または期間である場合がある。このような場合、単純に治療ペア間の有意差を検出することよりも、用量反応関係があるかどうかを検討することに最も関心があるであろう。第 2 に、2つの群が同じ基本的な介入のバリエーションを受ける場合がある。例えば、同じ薬剤を経口または静脈内投与する。研究者は、2つのバリエーションが異なるかどうかを検討する前に、まずこれらの群を組み

合わせて比較群 (通常はプラセボまたは標準治療) と比較する。異なる用量を投与された群もこの方法で検討されることがある。このような解析が計画された場合、最初の比較の検出力を最大にするために、群を 1:1:2 の比率で割り付けるべきと考える研究者もいる。多重比較の統計的調整は、方法論学者の間で議論を呼び、コンセンサスは得られていない。このような調整を用いる人もいれば、決して調整を行わない人もいる^{27,28)}。研究者は、様々な解析方法を駆使し、多重性問題を回避することができる。以下に例を挙げる：

- 比較群間で単一の包括的な有意性検定を使用し (例えば、3 群間比較試験で A 対 B 対 C を比較する)、多重比較を避ける。注目すべきは、すべての治療にわたる単一の包括的な検定は、限られた用途しかないということである²⁹⁾。
- 用量反応関係をモデル化し、多重比較を排除する³⁰⁾。
- 検定の優先順位を決める。例えば、研究者は 300 mg の抗生物質の新しい治療と標準治療の比較を優先検定として決定し、その比較が統計的に有意であれば、200 mg の抗生物質治療と標準治療の比較を行う。優先順位をつけた一連の検定は、調整なしで多重性問題を解決できる³¹⁾。
- すべての比較について透明性をもって報告し、多重性調整を行わないこと。多くの多群間比較試験は、3 群間比較試験における A 対 C、B 対 C の比較のように、関連性のない治療群とコントロール群を直接比較するように計画されている。このような多群間比較試験では、一般的に多重比較の調整は不要である^{2, 15, 32, 33)}。

多重性を考慮した正式な調整が避けられないこともある。欧州医薬品庁 (European Medicines Agency) を含む一部の規制当局では、そのような調整を求めている。欧州医薬品庁の臨床試験における多重性問題に関するガイドラインには、「一般的な規則として、試験ごとの第一種過誤の制御は、確証を得るための最低限の前提条件であると言える」と記載されている³⁴⁾。しかし、調整が適切であっても、その

実施には問題がある。ボンフェローニ調整は、単純であるため、推奨されることが多い。しかし、他の調整戦略は、第一種過誤を全体的に制御するには優れているが（一般にファミリーワイズエラー率（FWER）と呼ばれる）^{32, 35-37}、一方で、1つ以上の偽陽性の確率を悪化させる³²。特にボンフェローニ調整では、多重性に対する過剰補正が頻繁に行われている。この方法は、比較間の相関が高くなるにつれて過度に保守的になる。ホルム（Holm）法、Hochberg法、ダネット（Dunnett）検定、調整Hochberg法など、他の方法をボンフェローニ法と比較することが行われている³²。どの方法もボンフェローニ法よりも保守的でないように見える。

4-4. 結果 [Results]

4-4-1. 項目 14a. CONSORT 2010：参加者の募集期間と追跡期間を特定する日付。[Item 14a. CONSORT 2010: Dates defining the periods of recruitment and follow-up.]

多群間比較試験の拡張版：募集期間と追跡期間が治療群間で異なる場合（例えば、群の追加や中止）、募集期間と追跡期間、その差異の理由、統計学的関連事項を説明する。

方法（試験設定とデザイン）：本試験は2001年1月から2004年12月にかけて、米国内の57の臨床施設（16の大学クリニック、10の州精神保健機関、7の退役軍人医療センター、6の民間非営利機関、4の個人開業診療所、14の混合システム施設）で実施された。患者はまず二重盲検下でオランザピン、ペルフェナジン、クエチアピン、リスペリドンのいずれかを投与する群にランダムに割り付けられ、最長18ヵ月または何らかの理由で治療が中止されるまで追跡された（第1相）。（ジプラシドンは、試験開始後に米国食品医薬品局（FDA）により使用が承認され、2002年1月に40mgを含む同一外観のカプセルの形で試験に追加された）。

方法（統計解析）：（前略）ジプラシドンは患者の約40%が登録された後に試験に追加された（中略）そして、ジプラシドン群を含む比較は、ジプラシドンが追加された後にランダムに割り付けられた患者コホート（ジプラシドンコホート）に限定され

た。通常、本試験は2つの非定型薬剤間の中止率における12%の絶対差を検出し、その統計的検出力が85%であった（中略）しかし、ジプラシドンを含む比較における統計的検出力は58%であった（中略）オランザピン、クエチアピン、リスペリドン、ペルフェナジン群間の全体的な差は、自由度（df）3の検定を用いて評価された。もしその差がP値0.05未満で有意であった場合、3つの非定型薬剤群は、ステップダウン法（step-down）または閉検定手順（closed testing）によって互いに比較され、P値0.05未満は統計的有意差を示すとみなされた（中略）ジプラシドン群は、4つのペアごとの比較に対するHochberg調整によって、ジプラシドンコホート内で他の3つの非定型薬剤群やペルフェナジン群と直接比較された。比較で得られた最も小さいP値は0.013（0.05÷4）であった。[Intention-To-Treat (ITT) 集団における有効性のアウトカム評価に関連する表2¹⁹ および図2¹⁹ のレジェンドの脚注に再説されている。]

結果（治療の中止）：（前略）ジプラシドンが試験に追加された後にランダムに割り付けられた889名の患者コホートにおいて、オランザピン投与群はジプラシドン投与群よりも何らかの理由で治療を中止するまでの期間が長かった（ハザード比0.76； $P=0.028$ ）。しかし、この差は多重比較の調整後では有意ではなかった（必要なP値 ≤ 0.013 ）¹⁹。

解説：募集中の臨床試験において、新興の治療法を新たなランダム化群として追加することは、試験が最新の状態に保たれ、新しい治療が可能な限り迅速に評価され、最適な治療法を決定するための時間と費用が最小限に抑えられるようにするために、研究者、規制当局、患者にとって望ましいことである³⁸。様々な方法論的、統計学的関連事項を検討する必要がある。これらの関連事項には、

- (1) 段階効果 (stage effects) と多重性に起因するファミリーワイズエラー率の制御、
- (2) 追加群との偏りのない比較のために同時対照群データのみを使用すること³⁹、
- (3) 統計的検出力（同時対照群データとの比較には十分な検出力が必要）、
- (4) 各群への割り付け比率および/または募集期

間（主要な仮説を検証するために必要な参加者の総数や募集に費やす時間を調整することにより、効率の改善が実現できる）、

- (5) コントロール群における変更の可能性（既存のコントロール群が劣っていることが示される可能性があり、よって、理論上コントロール群を変更しなければならない可能性がある）、
- (6) ロジスティカル上の考慮（例えば、追加のファンディング、すべての必要な承認・修正に要する時間、薬剤の調達、試験のランダム化や臨床データベースシステムの更新、盲検化、試験の監督、募集への影響の可能性）³⁸⁾、

などが含まれる。これらの関連事項をどの程度考慮すべきかは、試験の性質と構造によって異なる。アダプティブデザインに関する CONSORT の拡張版と重なり合う可能性がある²⁰⁾。

多群間比較試験で複数の治療群への募集が早期に中止された場合、理由が異なる可能性があるため、それらを記載することが重要である。また、CONSORT 標準版の項目 15（すなわち、各群のベースラインにおける人口統計学的、臨床的な特性を示す表）に関して、すべての治療群への募集が同時期に行われない状況では、単一または複数のベースラインの表を使用できる。著者は、各群のどの比較にどの参加者が含まれるかを明示しなければならない。

4-4-2. 項目 17a CONSORT 2010：主要・副次的アウトカムのそれぞれについて、各群の結果、介入のエフェクト・サイズの推定とその精度（95% 信頼区間など）。[Item 17a. CONSORT 2010: For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval).]

多群間比較試験の拡張版：事前に決めた各治療群の比較結果。

主要アウトカム - 6 ヶ月時点では、フォーム群におけるアバディーン静脈瘤質問票 (Aberdeen Varicose Veins Questionnaire, AVVQ) のスコアは手術群よりも有意に高かったが（疾患特異的な QOL の悪化を示す）、その差は中程度であった（効果量 -

1.74; 95% 信頼区間 -2.97 ~ -0.50; $P = 0.006$ ）。レーザー群における AVVQ スコアの改善は、手術群と有意差はなかった。EQ-5D スコア（一般的な健康状態を測定するための標準化尺度）および SF-36 (Short Form Health Survey) の身体的側面スコアについては、両群間に有意差はみられなかった。レーザー治療とフォーム治療の事後解析では、唯一の有意差は SF-36 の精神的側面スコアであり、フォーム群よりレーザー群でわずかに高かった（一般的な QOL の改善）（効果量 1.54; 95% 信頼区間 0.01 ~ 3.06; $P = 0.048$ ）。（中略）副次的アウトカム - QOL。6 週時点で、下記の項目について有意な群間差が認められた ($P < 0.005$)：AVVQ スコアがフォーム群より手術群で低かったこと（疾患特異的な QOL の改善を示す）（効果量 -2.3; 95% 信頼区間 -3.7 ~ -0.9）、SF-36 スコアがレーザー群より手術群で低かったこと（一般的な QOL の悪化を示す）（効果量 -2.7; 95% 信頼区間 -4.4 ~ -0.9）、活力（効果量 -2.3; 95% 信頼区間 -3.9 ~ -0.8）、精神的健康による日常役割の制限（効果量 -2.4; 95% 信頼区間 -4.0 ~ -0.8）、身体的健康による日常役割の制限（効果量 -3.5; 95% 信頼区間 -5.2 ~ -1.8）。これら 4 つの SF-36 の下位尺度のスコアは、6 ヶ月時点では、群間有意差はなかった ($P < 0.005$ は統計的有意差を示す)。レーザー治療とフォーム治療の事後比較では、EQ-5D スコアのみが、6 週時点でフォーム治療群で有意に低かった（一般的な QOL の悪化を示す）($P = 0.044$; 95% 信頼区間 0.014 ~ 0.074)⁴⁰⁾。

解説：研究者は意図した比較を計画し、それをプロトコルと統計解析計画書に記載し、適切な解釈とともに試験報告書にすべて報告すべきである。試験中に介入群を追加または中止した場合は、解析でその意味を説明することが重要である。研究者が全体的な有意水準を制御する対策を講じた場合は（例えば、比較群間で単一の包括的な有意性検定を実施した場合、用量反応関係をモデル化した場合、優先順位をつけた一連の検定を使用した場合）、その詳細を報告すべきである。研究者が多重性の正式な調整を必要とする解析を行った場合は、その方法と限界を報告すべきである。前述したように（項目 12a）、多くの多群間比較試験では多重性の正式な調整を行わない。そのような場合でも、計画的、非計画的

問わず、実施されたすべての比較について透明性をもって報告し、結果の適切な解釈を示すべきである。

4-5. 考察 [Discussion]

4-5-1. 項目 20. CONSORT 2010: 試験の限界、可能性のあるバイアスや精度低下の原因、関連する場合は解析の多重性の原因を記載。 [Item 20. CONSORT 2010: Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses.]

ここでは、標準的な CONSORT 項目に対する具体的な拡張内容は提案されていないが、著者は Box に詳述されている問題に関して、多群間比較試験の長所と限界に言及すべきである。

5. 考察 [Discussion]

多群間比較試験には慎重な検討と計画が必要である。多群間比較試験は、1つ以上の研究課題に取り組む機会を提供し、新たな介入の評価を加速させ、競合する治療選択肢との直接比較を容易にし、資源利用を最適化しながら患者に利益をもたらす可能性がある。通常、標準治療よりも実験的介入を受ける確率が高いため、参加者や臨床医にとって多群間比較試験は2群間比較試験よりも魅力的かもしれない。しかし、研究者は、多群間比較試験の効率的な利点と、より多くの介入をより短期間で評価する機会には、必要な参加者数を募集することアウトカムと照合することが条件であることに常に留意すべきである。

多群間ランダム化比較試験は一般的であり、これらの試験の報告には、読者が試験の実施と結果の解釈を正確に評価できるように、試験デザインに関する特徴的な情報を説明することが重要である。透明かつ完全な報告は、再現性の前提条件である。また、優れた報告は、システマティックレビューにおける多群間比較試験の同定と採用を容易にする。しかし、多群間比較試験、特に3群を超える治療群を有する試験は、デザインも解析も困難である。

本稿は、多群間ランダム化比較試験の完全かつ正確な報告を可能にするため、広く採用されている CONSORT 2010 声明の拡張版を示しているもの

である。このような試験では、すべての治療群について明確な目的と仮説が必要であり、主要な比較を明記する必要もある。サンプルサイズは事前に決めるべきであり、多重性検定の調整に関する問題も少なくとも認識されるべきである。募集期間と追跡期間が治療群間で異なる場合（例えば、群の追加や中止）、それぞれの期間とその差異の理由を報告し、統計学的関連事項を説明すべきである。

多群間ランダム化比較試験では、群間の複数の比較における多重性調整は、依然として困難な問題である。多くの多群間比較試験は効率性のために実施されている。それらの試験は、単一のコントロール群に対して異なる治療・介入を比較するものであり、一つの多群間比較試験ではなく、複数の個別の試験で容易に実施できる。複数の実験的介入がコントロール群を共有する多群間比較試験のデザインでは、各介入の研究課題を個別に評価することに焦点が当てられる。通常、1つの比較結果の解釈は、他の比較の解釈に直接関係しない。介入が個別の試験で比較された場合、そのような調整は必要ないため、そのような試験では多重性調整も必要ないと主張する試験専門家・方法論学者は多い^{2, 15, 32, 33, 41, 42}。一部の多群間比較試験では、同じ薬剤における複数の異なる用量を一つのコントロール群と比較して評価し、これは関連のある比較を表している^{2, 32, 33, 37}。このような状況では、試験専門家や方法論学者は多重性の調整を推奨する傾向がある。一例として、規制当局への医薬品承認申請におけるある判断基準からこのような状況が生じる。資金提供者が1つを超える治療の比較を指定し、1つ以上の用量が統計的に有意な場合に治療効果を主張することを申し出た場合、ほとんどの試験専門家や方法論学者は多重性調整を提案している^{2, 15, 32, 33, 41, 42}。しかし、多重性検定の調整が常に必要である、絶対に必要でないといった大げさな主張は無視すべきであり、調整に関する決定は、目的、デザイン、解析によって決まる。

多群間比較試験の中には、クロスオーバー試験、クラスター試験、要因試験など、特徴を持つものもある。そのような試験では、試験の種別ごとに特定の推奨事項が重要である。2群並行群間比較試験の報告に CONSORT 声明を使用することは、すでに

報告の質の向上に寄与することが示されている⁴³⁾。

今回提案された CONSORT 声明の拡張版の日常的な使用には、同様の改善を促進することが期待されている。

CONSORT グループは、今後も推奨事項のモニタリングと改訂を続け、様々なデザインの臨床試験の報告の質を向上させるためのチェックリストやフローチャートを作成している。その他の類似した拡張版や更新版も準備中であり、すべての CONSORT

の推奨事項の最新版は CONSORT ウェブサイト (<http://www.consort-statement.org>) で確認できる。

6. 結論 [Conclusions]

この CONSORT 2010 声明の拡張版は、多群ランダム化並行群間比較臨床試験の報告に関する具体的なガイダンスを示し、この種の臨床試験の報告における透明性と正確性の向上に役立つことが期待される。

参考文献

1. Parmar MK, Carpenter J, Sydes MR: More multiarm randomised trials of superiority are needed. *Lancet*. 2014; **384**(9940): 283-284. doi:10.1016/S0140-6736(14)61122-3
2. Freidlin B, Korn EL, Gray R, Martin A: Multi-arm clinical trials of new agents: some design considerations. *Clin Cancer Res*. 2008; **14**(14):4368-4371. doi:10.1158/1078-0432.CCR-08-0325
3. Moher D, Dulberg CS, Wells GA: Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994; **272**(2):122-124. doi:10.1001/jama.1994.03520020048013
4. Odutayo A, Emdin CA, Hsiao AJ, *et al.*: Association between trial registration and positive study findings: cross sectional study (Epidemiological Study of Randomized Trials-ESORT). *BMJ*. 2017; **356**: j917. doi:10.1136/bmj.j917
5. Moher D, Hopewell S, Schulz KF, *et al.*: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010; **340**: c869. doi:10.1136/bmj.c869
6. Schulz KF, Altman DG, Moher D: CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010; **340**: c332. doi:10.1136/bmj.c332
7. Hopewell S, Clarke M, Moher D, *et al.*: CONSORT Group. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med*. 2008; **5**(1): e20. doi:10.1371/journal.pmed.0050020
8. Ioannidis JP, Evans SJ, Gøtzsche PC, *et al.*: CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004; **141**(10): 781-788. doi:10.7326/0003-4819-141-10-200411160-00009
9. Ahrén B, Johnson SL, Stewart M, *et al.*: HARMONY 3 Study Group. HARMONY 3: 104-week randomized, double-blind, placebo-controlled active-controlled trial assessing the efficacy and safety of albiglutide compared with placebo, sitagliptin, and glimepiride in patients with type 2 diabetes taking metformin. *Diabetes Care*. 2014; **37**(8): 2141-2148. doi:10.2337/dc14-0024
10. Agar MR, Lawlor PG, Quinn S, *et al.*: Efficacy of oral risperidone, haloperidol, or placebo for symptoms of delirium among patients in palliative care: a randomized clinical trial. *JAMA Intern Med*. 2017; **177**(1): 34-42. doi:10.1001/jamainternmed.2016.7491
11. Connick P, De Angelis F, Parker RA, *et al.*: UK Multiple Sclerosis Society Clinical Trials Network. Multiple Sclerosis-Secondary Progressive Multi-Arm Randomisation Trial (MS-SMART): a multiarm phase IIb randomised, double-blind, placebo-controlled clinical trial comparing the efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis. *BMJ Open*. 2018; **8**(8): e021944. doi:10.1136/bmjopen-2018-021944
12. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S: Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof*. 2002; **25**(1):38-64.
13. Geddes JR, Goodwin GM, Rendell J, *et al.*: BALANCE investigators and collaborators. Lithium plus valproate combination therapy versus monotherapy for relapse prevention in bipolar I disorder (BALANCE): a randomised open-label trial. *Lancet*. 2010; **375**(9712): 385-395. doi:10.1016/S0140-6736(09)61828-6
14. European Medicines Agency. ICH Topic E 10: Choice of Control Group in Clinical Trials. Canary Wharf, London, United Kingdom: European Medicines Agency; 2001. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf. Accessed February 14, 2018.
15. Schulz KF, Grimes DA: Multiplicity in randomised trials I: endpoints and treatments. *Lancet*. 2005; **365**(9470): 1591-1595. doi:10.1016/S0140-6736(05)66461-6
16. Foa EB, McLean CP, Zang Y, *et al.*: STRONG STAR Consortium. Effect of prolonged exposure therapy delivered over 2 weeks vs 8 weeks vs present-centered therapy on PTSD symptom severity in military personnel: a randomized clinical trial. *JAMA*. 2018; **319**(4): 354-364. doi:10.1001/jama.2017.21242
17. Gray R, Ives N, Rick C, *et al.*: PD Med Collaborative Group. Long-term effectiveness of dopamine agonists and monoamine oxidase B inhibitors compared with levodopa as initial treatment for Parkinson's disease (PD MED): a large, open-label, pragmatic randomised trial. *Lancet*. 2014; **384**(9949): 1196-1205. doi:10.1016/S0140-6736(14)60683-8
18. Howard RJ, Juszcak E, Ballard CG, *et al.*: CALM-AD Trial Group. Donepezil for the treatment of agitation in Alzheimer's disease. *N Engl J Med*. 2007; **357**(14): 1382-1392. doi:10.1056/NEJMoa066583

19. Lieberman JA, Stroup TS, McEvoy JP, *et al.*: Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med.* 2005; **353**(12): 1209-1223. doi:10.1056/NEJMoa051688
20. Dimario M, Todd S, Julious S, *et al.*: Adaptive designs CONSORT Extension (ACE) Project Protocol: version 2.3. <http://www.equator-network.org/wp-content/uploads/2017/12/ACE-Project-Protocol-v2.3.pdf>. Published 2016. Accessed February 14, 2018.
21. Montorsi F, Brock G, Stolzenburg JU, *et al.*: Effects of tadalafil treatment on erectile function recovery following bilateral nerve-sparing radical prostatectomy: a randomised placebo-controlled study (REACTT). *Eur Urol.* 2014; **65**(3): 587-596. doi:10.1016/j.eururo.2013.09.051
22. Pickard R, Lam T, MacLennan G, *et al.*: Antimicrobial catheters for reduction of symptomatic urinary tract infection in adults requiring short-term catheterisation in hospital: a multicentre randomised controlled trial. *Lancet.* 2012; **380**(9857): 1927-1935. doi:10.1016/S0140-6736(12)61380-4
23. Moss AJ, Schuger C, Beck CA, *et al.*: MADIT-RIT Trial Investigators. Reduction in inappropriate therapy and mortality through ICD programming. *N Engl J Med.* 2012; **367**(24): 2275-2283. doi:10.1056/NEJMoa1211107
24. Ndiranza J, Mpairwe H, Webb EL, *et al.*: Impact of anthelmintic treatment in pregnancy and childhood on immunisations, infections and eczema in childhood: a randomised controlled trial. *PLoS One.* 2012; **7**(12): e50325. doi:10.1371/journal.pone.0050325
25. Fong DT, Pang KY, Chung MM, Hung AS, Chan KM: Evaluation of combined prescription of rocker sole shoes and custom-made foot orthoses for the treatment of plantar fasciitis. *Clin Biomech* (Bristol, Avon). 2012; **27**(10): 1072-1077. doi:10.1016/j.clinbiomech.2012.08.003
26. Agnelli G, Buller HR, Cohen A, *et al.*: AMPLIFY-EXT Investigators. Apixaban for extended treatment of venous thromboembolism. *N Engl J Med.* 2013; **368**(8): 699-708. doi:10.1056/NEJMoa1207541
27. Perneger TV: What's wrong with Bonferroni adjustments. *BMJ.* 1998; **316**(7139): 1236-1238. doi:10.1136/bmj.316.7139.1236
28. Rothman KJ: No adjustments are needed for multiple comparisons. *Epidemiology.* 1990; **1**(1): 43-46. doi:10.1097/00001648-199001000-00010
29. Pocock SJ: *Clinical Trials: A Practical Approach*. Chichester, UK: John Wiley & Sons Ltd; 1983.
30. Senn S: *Statistical Issues in Drug Development*. Chichester, UK: John Wiley & Sons Ltd; 1997.
31. Bauer P, Chi G, Geller N, *et al.*: Industry, government, and academic panel discussion on multiple comparisons in a "real" phase three clinical trial. *J Biopharm Stat.* 2003; **13**(4): 691-701. doi:10.1081/BIP-120024203
32. Howard DR, Brown JM, Todd S, Gregory WM: Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *Stat Methods Med Res.* 2018; **27**(5): 1513-1530. doi:10.1177/0962280216664759
33. Wason JM, Stecher L, Mander AP: Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? *Trials.* 2014; **15**: 364. doi:10.1186/1745-6215-15-364
34. European Medicines Agency: *Guideline on Multiplicity Issues in Clinical Trials [draft]*. Canary Wharf, London, United Kingdom: European Medicines Agency; 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf. Accessed February 14, 2018.
35. Sankoh AJ, D'Agostino RB Sr, Huque MF: Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. *Stat Med.* 2003; **22**(20): 3133-3150. doi:10.1002/sim.1557
36. Hsu JC: *Multiple Comparisons: Theory and Methods*. New York, NY: Chapman & Hall; 1996.
37. Proschan MA, Waclawiw MA: Practical guidelines for multiplicity adjustment in clinical trials. *Control Clin Trials.* 2000; **21**(6): 527-539. doi:10.1016/S0197-2456(00)00106-9
38. Cohen DR, Todd S, Gregory WM, Brown JM: Adding a treatment arm to an ongoing clinical trial: a review of methodology and practice. *Trials.* 2015; **16**: 179. doi:10.1186/s13063-015-0697-y
39. Altman DG: Avoiding bias in trials in which allocation ratio is varied. *J R Soc Med.* 2018; **111**(4): 143-144. doi:10.1177/0141076818764320
40. Britten J, Cotton SC, Elders A, *et al.*: A randomized trial comparing treatments for varicose veins. *N Engl J Med.* 2014; **371**(13): 1218-1227. doi:10.1056/NEJMoa1400781
41. Duncan DB: Multiple range and multiple F tests. *Biometrics.* 1955; **11**: 1-42. doi:10.2307/3001478
42. Cook R, Jones V: Multiplicity considerations in the design and analysis of clinical trials. *J Royal Stat Soc A Stat Soc.* 1996; **159**(1): 93-110. doi:10.2307/2983471
43. Turner L, Shamseer L, Altman DG, *et al.*: Consolidated Standards of Reporting Trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev.* 2012; **11**: MR000030.

骨の健康と食因子

大好評
発売中

骨代謝研究の第一人者
渾身の一冊！

骨粗鬆症の予防と修復へのアプローチ

骨の健康と食因子

骨粗鬆症の予防と修復へのアプローチ



山口 正義
Masayoshi Yamaguchi

国内外から高い評価を得ている著者が長年の研究成果をもとに、生体機能がホルモン（内分泌因子）によってダイナミックに調節されていることを紹介し、その一端としてのカルシウム代謝および骨代謝のホルモン調節のしくみを骨粗鬆症の理解のために概説した。さらに、骨粗鬆症の予防と修復に役立つ食因子の知見とその周辺について記述した全10章からなる貴重な一冊。

- 第1章 ホルモンと生体機能調節
- 第2章 ホルモンの細胞内への情報伝達とそのしくみ
- 第3章 カルシウム代謝とそのホルモン調節
- 第4章 骨代謝とそのホルモン調節
- 第5章 老化と骨カルシウムホメオスタシス
- 第6章 栄養性ミネラルと骨粗鬆症の予防
- 第7章 生体微量元素と骨粗鬆症の予防
- 第8章 骨粗鬆症を予防する食品由来生理活性因子
- 第9章 骨粗鬆症を予防する食品素材
- 第10章 複合食因子の骨効果と新規サプリメントの開発

■ 著者 / 山口 正義 (やまぐち まさよし)

◆薬学博士。米国エモリー大学医学部医学科内分泌代謝学部門 客員教授 (任用)

■ A5 版 / 248 ページ
■ 定価：(本体 3,500円+ 税)
■ 発行：食品資材研究会

◆ご注文は FAX またはメールにて FAX:042-312-0845 info@nfi-llc.co.jp