

シリーズ EQUATOR Network が提供するガイドラインの紹介

Consensus Statement for Protocols of Factorial Randomized Trials: Extension of the SPIRIT 2013 Statement の和訳

著者 Brennan C. Kahan, PhD¹; Sophie S. Hall, PhD²; Elaine M. Beller, MAppStat³; Megan Birchenall, BSc²; Diana Elbourne, PhD⁴; Edmund Juszczak, MSc⁵; Paul Little, MD; John Fletcher, MPH⁶; Robert M Golub, MD⁷; Beatriz Goulao, PhD⁸; Sally Hopewell, DPhil⁹; Nazrul Islam, PhD^{5,6}; Merrick Zwarenstein, MBBCh, PhD¹⁰; An-Wen Chan, MD, DPhil¹¹; Alan A. Montgomery, PhD²

翻訳 馬場 亜沙美 (*BABA Asami*)^{1*}, 鈴木 直子 (*SUZUKI Naoko*)¹,
田中 瑞穂 (*TANAKA Mizuho*)¹, 山本 和雄 (*YAMAMOTO Kazuo*)¹

Key Words: ヒト臨床試験, 特定保健用食品, 機能性表示食品, ファクトリアルデザイン, CONSORT, SPIRIT, 試験プロトコル

Japanese translation of “Consensus Statement for Protocols of Factorial Randomized Trials: Extension of the SPIRIT 2013 Statement”

Authors:

Brennan C. Kahan, PhD¹; Sophie S. Hall, PhD²; Elaine M. Beller, MAppStat³; Megan Birchenall, BSc²; Diana Elbourne, PhD⁴; Edmund Juszczak, MSc⁵; Paul Little, MD; John Fletcher, MPH⁶; Robert M Golub, MD⁷; Beatriz Goulao, PhD⁸; Sally Hopewell, DPhil⁹; Nazrul Islam, PhD^{5,6}; Merrick Zwarenstein, MBBCh, PhD¹⁰; An-Wen Chan, MD, DPhil¹¹; Alan A. Montgomery, PhD²

Translators:

Asami Baba^{1*}, Naoko Suzuki¹, Mizuho Tanaka¹, Kazuo Yamamoto¹

*Correspondence author: Asami Baba

Keywords: Human clinical trial, Food for Specified Health Uses, Food with Function Claims, Factorial design, CONSORT, SPIRIT, Trial protocol

Affiliations (Authors):

¹ MRC Clinical Trials Unit at UCL, London, United Kingdom

² Nottingham Clinical Trials Unit, School of Medicine, University of Nottingham, Nottingham, United Kingdom

³ Institute for Evidence-Based Healthcare, Bond University, Robina, Australia

⁴ London School of Hygiene and Tropical Medicine, London, United Kingdom

⁵ Primary Care Research Centre, School of Primary Care, Population Sciences and Medical Education, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

⁶ The BMJ, London, United Kingdom

* 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

⁷ Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

⁸ Health Services Research Unit, University of Aberdeen, Aberdeen, Scotland

⁹ Oxford Clinical Trials Research Unit, University of Oxford, Oxford, United Kingdom

¹⁰ Centre For Studies in Family Medicine, Schulich School of Medicine and Dentistry, Western University, London, Canada

¹¹ Women’s College Research Institute, University of Toronto, Toronto, Canada

Affiliated institution (Translators)

¹ ORTHOMEDICO Inc.

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

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「Consensus Statement for Protocols of Factorial Randomized Trials: Extension of the SPIRIT 2013 Statement」を翻訳・要約したものである。

キーポイント [Key Points]

論点: ファクトリアルランダム化比較試験のプロトコルには、どのような追加情報を記載すべきか？

結果: 本コンセンサス声明は、介入研究のための標準的プロトコル項目に関する推奨事項(SPIRIT 2013 声明)の拡張版を提示するものである。SPIRIT 項目のうち 9 項目が修正されている。

解釈: 本 SPIRIT 拡張版チェックリストは、ファクトリアル試験プロトコルにおける透明性ある記載を促進し、試験の有用性向上に寄与する可能性がある。

要旨 [Abstract]

背景: 試験プロトコルは、試験の目的と用いられる方法(設計、実施、解析)を概説するものであり、プロトコルによる透明性のある報告は目的を明確にし、研究方法の妥当性を評価するための手助けとなる。ファクトリアル試験では 2 つ以上の介入を同一の参加者群で評価するため、試験を設計する上での留意点が存在する。しかし、プロトコルの報告に関する指針を提供する「標準的なプロトコル項目：介入試験のための推奨(SPIRIT 2013 声明)」には、ファクトリアル試験に特化した拡張ガイドラインはこれまで存在していない。

目的: ファクトリアル試験に対応した SPIRIT 2013 声明の合意形成に基づく拡張ガイドラインを開発する。

エビデンスの検討: ファクトリアル試験のための SPIRIT 拡張ガイドラインは、医療研究の質と透明性の向上を目的とした EQUATOR 方法論フレームワークを用いて開発された。まず、報告推奨項目のリストは、MEDLINE(開始から 2019 年 5 月まで)の検索によって特定された方法論的文献のスコピングレビューを通じて作成され、著者らの個人的な文献資料からの関連文献によって補足された。次に、14 か国 104 名の調査パネリストが参加したデルファイ調査が 2022 年 1 月から 6 月にかけて 3 回実施され、各項目の重要性の評価と追加推奨事項の特定が行われた。最後に、15 名の調査パネリストが参加した合同会議が開催され、チェックリストの項目選定と文言の最終決定が実施された。

結果: ファクトリアル試験に対応する本 SPIRIT 拡張版では、SPIRIT 2013 のチェックリスト全 33 項目のうち、9 項目に修正を加えた。報告に関する主要な推奨事項としては、ファクトリアルデザインを採用する根拠(交互作用の仮定の有無を含む)を明示すること、主要な比較対象となる群を特定すること、そして各主要比較に対して用いる統計解析手法(交互作用の評価方法を含む)を記述することが求められる。

結論と研究意義: 本コンセンサス声明では、ファクトリアル試験の有用性と透明性を高めるために、すべてのファクトリアル試験のプロトコルにおいて記載すべき 9 つの項目が提示された。

序論 [Introduction]

試験プロトコルは、研究の根拠、目的、ならびに統計解析を含む計画された方法を明示する^{1,2)}。試験プロトコルは、研究実施者にとって試験実施の指針となるものであり、倫理委員会による倫理的妥当性の確認、ならびに学術誌・規制当局・査読者による試験の実施および報告の評価に用いられる^{1,2)}。これらの目的を満たすのに適した試験プロトコル作成を支援するために、「標準的なプロトコル項目：介入試験のための推奨（SPIRIT 2013 声明）」が策定された^{1,2)}。SPIRIT 2013 声明では、試験プロトコルに記載すべき 33 の項目を体系的に整

理したチェックリストを提示している。SPIRIT は基本的に 2 群並行群間比較試験を対象としているが、複雑な試験デザインにおいては項目の適用に際して調整や追加が必要となる可能性がある。

ファクトリアル試験とは、単一の試験において、同一の参加者集団において 2 つ以上の介入が評価される試験である³⁻¹⁶⁾。因子 A および B を組み合わせた 2×2 ファクトリアル試験の例を表 1 に示す。参加者は介入 A またはその比較対象に加え、介入 B またはその比較対象にも割り付けられるため、結果として 4 つの治療群（A 単独、B 単独、A+B 併用、A・B いずれも非投与 [二重対照群]）のいずれか

表 1 2×2 ファクトリアルランダム化試験の例

治療	因子	治療 B ^a	
		アクティブ ^b	コントロール ^b
治療 A ^a	アクティブ ^b	アクティブ A+ アクティブ B ^c	アクティブ A+ コントロール B ^c
	コントロール ^b	コントロール A+ アクティブ B ^c	コントロール A+ コントロール B ^c

a: A と B は因子である。

b: アクティブ A とコントロール A は因子 A のレベルであり、アクティブ B とコントロール B は因子 B のレベルである。

c: これらの項目は 4 つの治療群を表す。完全ファクトリアル試験では、すべての参加者が 4 つの治療群のいずれかにランダムに割り付けられる。部分ファクトリアル試験では、参加者の一部はアクティブ A とコントロール A の間でのみランダム化され、B 因子についてはランダム化されず、自動的にコントロール B に割り付けられる。ファクトリアル解析では、介入 A に割り付けられたすべての参加者（アクティブ A+ アクティブ B、およびアクティブ A+ コントロール B）を、介入 A に割り付けられていない参加者（コントロール A+ アクティブ B、およびコントロール A+ コントロール B）と比較する。同様にして、介入 B についての比較も行われる。マルチアーム解析では、各治療群をコントロール群と比較する（例：アクティブ A+ アクティブ B、アクティブ A+ コントロール B、コントロール A+ アクティブ B の各群は、いずれもコントロール A+ コントロール B と比較される）。

表 2 用語集

用語	定義
ファクトリアル試験	単一の研究で、同一の参加者に対して 2 つ以上の介入を組み合わせて評価される試験デザイン
因子	各介入とその比較対象を含む概念（例：因子 A にはアクティブ A とコントロール A が含まれる）
因子内のレベル	因子を構成する具体的な介入（例：アクティブ A とコントロール A は因子 A の 2 つのレベルである）
治療群	参加者がランダムに割り付けられる因子とレベルの一意的な組み合わせ（例：アクティブ A+ アクティブ B は 1 つの治療群である）
完全ファクトリアルデザイン	すべての参加者が、因子とそのレベルのあらゆる組み合わせのいずれかにランダムに割り付けられる
部分ファクトリアルデザイン	一部の参加者は、特定の因子についてはランダム化されない
分割ファクトリアルデザイン	因子の組み合わせの一部が省略される
比較	どの治療群同士を比較するか
主要な比較	各介入の有効性に関する結論を導くために主に用いられる比較
推定値	試験から推定される治療効果の記述
ファクトリアル解析	マージン解析（at-the-margins analysis）とも呼ばれ、アクティブ A に割り付けられた全参加者をコントロール A に割り付けられた全参加者と比較し、因子 B の比較も同様に行う
マルチアーム解析	インサイド・ザ・テーブル解析（inside-the-table analysis）とも呼ばれ、各治療群（例：アクティブ A+ コントロール B、コントロール A+ アクティブ B）がコントロール群などの他の治療群と比較される
相互作用	ある治療の効果が、参加者が他の治療も受けるかどうかによって変化する場合に生じる

表3 ファクトリアルランダム化試験の報告チェックリスト:SPIRIT 2013 声明の拡張版^{a,b,c}

章	項目番号	SPIRIT 2013 チェックリスト項目	ファクトリアル試験に関する拡張版
管理情報			
タイトル	1	試験デザイン、対象集団、介入、および該当する場合は試験の略称を示す記述的タイトル	試験がファクトリアルランダム化試験であること、対象集団、介入、および該当する場合は試験の略称を示す記述的タイトル
はじめに			
背景と根拠	6a	研究課題と試験を実施する正当性の記述、各介入の有益性および有害性を検討した関連研究(公表・未公表)の要約を含む	研究課題と試験を実施する正当性の記述、各介入の有益性および有害性を検討した関連研究(公表・未公表)の要約を含む。さらに、相互作用が仮定されているかどうかを含め、ファクトリアルデザインを採用する理由の説明
目的	7	具体的な目的または仮説	具体的な目的または仮説、および主要な比較を構成する治療群の記述 ^b
試験デザイン	8	試験の種類(例:並行群間、クロスオーバー、ファクトリアル、単一群)、割付比率、および枠組み(例:優越性、同等性、非劣性、探索的)を含む試験デザインの記述	ファクトリアル試験の種類(例:完全または部分ファクトリアル、因子の数、各因子内のレベル)、割付比率、および枠組み(例:優越性、同等性、非劣性、探索的)を含む試験デザインの記述
方法: 参加者、介入、アウトカム			
適格基準	10	参加者の選択基準および除外基準。該当する場合は、研究施設および介入を実施する個人の適格基準(例:外科医、心理療法士)	各因子の選択基準および除外基準(該当する場合は違いを明記)、および研究施設や介入を実施する個人の適格基準(例:外科医、心理療法士)
サンプルサイズ	14	研究目的を達成するために必要な推定参加者数およびその決定方法(サンプルサイズ計算を裏付ける臨床的・統計的仮定を含む)	研究目的を達成するために必要な推定参加者数および主要な各比較に対する決定方法(サンプルサイズ計算を裏付ける臨床的・統計的仮定、計算で相互作用が想定されたかどうかを含む)
介入の割付け(対照試験の場合)			
シーケンス生成	16a	割付けシーケンスの生成方法(例:コンピュータ生成の乱数)および層別化のための因子の一覧。ランダム化シーケンスの予測可能性を低減するため、計画された制限(例:ブロッキング)の詳細は、参加者登録者や介入割付担当者が閲覧できない別文書で提供すること	割付けシーケンスの生成方法(例:コンピュータ生成の乱数)、層別化のための変数の一覧、および該当する場合は参加者が異なる時点で因子に割り当てられたかどうか。ランダム化シーケンスの予測可能性を低減するため、計画された制限(例:ブロッキング)の詳細は、参加者登録者や介入割付担当者が閲覧できない別文書で提供すること
データ収集、管理、解析			
統計学的手法	20a	主要アウトカムおよび副次アウトカムを分析するための統計的方法。プロトコルに記載されていない場合は、統計解析計画の詳細の所在を明示	<p>主要アウトカムおよび副次的アウトカムについて、各主要比較に用いる統計学的方法(以下を含む)</p> <ul style="list-style-type: none"> ●目標治療効果が、他の因子の有無によって影響を受けるかどうか ●解析アプローチ(例:ファクトリアル解析、多群解析など) ●アプローチの選択方法(事前指定か、推定された相互作用に基づくかなど) ●ファクトリアル解析を用いる場合、因子間で相互に調整するかどうか ●統計的相互作用を評価する方法、および主要アウトカムに加えてどのアウトカムに適用するか ●該当する場合、因子への非同時募集をどのように扱うか ●統計解析計画のその他の詳細がプロトコルに記載されていない場合の参考先
モニタリング			
データモニタリング	21b	中間解析および試験中止のガイドラインの記述(中間解析結果にアクセスする者および試験中止の最終決定を下す者を含む)	主要比較間の相違点およびその理由を含み、中間解析結果にアクセスする者および試験中止の最終決定を下す者を明示した中間解析および試験中止のガイドラインの記述

略語: IRB (institutional review board = 施設内倫理審査委員会), REB (research ethics board = 研究倫理審査委員会), SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials = 介入試験の標準プロトコル項目に関する推奨事項)

a: このチェックリストは、項目に関する重要な説明を理解するために、SPIRIT 2013 声明のガイドライン1)と併せて確認することが推奨される。

b: 比較される各介入群全体は因子となる(例:アクティブAとコントロールAを合わせて1つの因子、アクティブBとコントロールBを合わせて別の因子)。因子内の具体的な介入はそのレベルに相当する(例:アクティブAとコントロールAが因子Aの2つのレベル)。治療群は因子とレベルのユニークな組み合わせで構成される{例:因子AとBの2×2試験では、4つの治療群が存在する(アクティブA+コントロールB, アクティブA+アクティブBなど)}。主要な比較とは、各介入の有効性に関する主要な結論を導くために、どの治療群同士を比較するかを指す。

c: 「Consensus Statement for Protocols of Factorial Randomized Trials: Extension of the SPIRIT 2013 Statement」で拡張された項目のみを抽出した。

に割り当てられることになる。ファクトリアル試験は、複数の介入因子を同時に評価する特性上、並行群間比較試験よりも設計・解析の面で複雑性が増す。これらの試験は、異なる研究目的（推定量）に応じてそれぞれ異なる解析方法を適用することで、複数の課題に同時に対応可能である（Box 参照）。ファクトリアル試験は単一の試験で複数の介入を同時に評価できるだけでなく、ある治療の効果が他の治療の投与状況に依存するかどうか、すなわち治療間の交互作用の有無を検討することも可能である^{10, 15, 17, 18)}。主要比較群の定義、交互作用の解析戦略、非同時登録によるバイアスの管理など、ファクトリアル試験には並行群間比較試験にはない設計上の留意点が存在する^{3, 4, 6, 8, 12-15, 19)}。

本コンセンサス文書では、SPIRIT 2013 チェックリストをファクトリアル試験に適用するための拡張版を提案するものである^{1, 2)}。ファクトリアル試験において、“因子”は介入とその比較対象を含む単位を指し、“治療群”は因子の水準の組み合わせによって定義される。2×2 デザインでは、A 単独、B 単独、A+B 併用、二重対照の 4 群が設定される。主要な用語の定義をまとめた用語集を表 2 に示す。本コンセンサス文書は2×2 のファクトリアル試験を中心に論じているが、提示された報告指針は因子数や水準数が増加する複雑なファクトリアルデザインにも応用可能である。

方法 [Methods]

ファクトリアル試験に対応する SPIRIT 拡張版は、同様の目的を持つ「臨床試験報告に関する統合基準 (CONSORT)」ガイドライン拡張版と並行して策定された²⁰⁾。本拡張は EQUATOR ネットワークの方法論的枠組みに則って開発されており、報告にあたっては「品質改善報告の卓越性に関する基準 (SQUIRE)」ガイドラインに準拠している²¹⁾。詳細な方法論については、別資料にて提示している²²⁾。ファクトリアル試験プロトコルの報告推奨項目の初期リストを作成するためにスコーピングレビューを実施した。レビュー対象には、2019 年 5 月までに発表された方法論的文献に加え、著者らの個人所蔵文献も含まれている。推奨項目の体系化に向けて資金提供を受けた後、各項目の重要度評価および追加項目の提案を得ることを目的として、2022 年 1 月

から 6 月にかけて 3 ラウンドのデルファイ調査を実施した。最終的なチェックリストの構成と文言に関する合意形成のため、2022 年 9 月 6, 7 日に 15 名のパネリストによるハイブリッド会議を開催し、以後は電子メールを通じた議論を行った。

結果 [Results]

ファクトリアル試験プロトコルの報告に対応する修正チェックリストを表 3 に提示している。本チェックリストには、SPIRIT 2013 声明の既存項目を修正した 9 項目が含まれている。

Box. ファクトリアル試験における推定対象の概要

ファクトリアル試験における推定対象

- 試験が解決を目指す研究課題を明確に定義するために、推定対象が用いられる。
- ファクトリアル試験では、研究目的に応じて異なる種類の推定対象を定義することができる。
- 2 因子試験において、推定対象は通常、治療 A と非治療 A との比較（他の因子についても同様）に基づいて定義される。しかし、この推定対象は複数の方法で定義可能である。たとえば、治療 B を誰も受けていない場合における治療 A と非治療 A の比較に基づくことも、全員が治療 B を受けている場合における治療 A と非治療 A の効果として定義することも可能である。
- あるいは、治療 A に関する推定対象は、治療 B を受ける群と受けない群の両方を含めた平均的な治療 A と非治療 A の比較に基づいて定義することもできる^{a)}。しかし、このような推定対象は、実際の臨床現場における治療の使用実態を必ずしも反映しておらず、そのため 2 因子試験においては他の種類の推定対象の方がより適切である場合が多い。
- 治療間の交互作用の有無を評価することを目的とした試験においては、推定対象は、治療 B を誰も受けていない場合における治療 A の効果と、全員が治療 B を受けている場合における治療 A の効果との差に基づいて定義されることがある。

統計解析への含意

- 統計解析の手法は、あらかじめ定義された推定対象に基づいて選択されるべきである。

● 2因子試験では、効率性の観点から、ファクトリアル解析（別名：マージン解析）が一般的に用いられる。しかし、この解析は、治療Bを割り付けられた群と割り付けられていない群の2つの層にわたって平均化を行うため、治療AとBの間に交互作用が存在しないという前提のもとでは、治療Bを誰も受けていない場合における治療Aの効果のみを推定することになる。治療間に交互作用が存在する場合、この解析は治療Bの層にわたる治療Aの平均効果を推定することになるが、これは通常、主要な関心対象とはならない。

●マルチアーム解析（別名：インサイド・ザ・テーブル解析）では、治療AとBの間に交互作用が存在する場合でも、治療Bを誰も受けていない条件下における治療Aの効果を推定することが可能である。しかし、ファクトリアル解析と比較して効率性が劣るため、2因子試験ではインサイド・ザ・テーブル（inside-the-table analysis）が用いられる頻度は低い。

a: この分布は、研究者が任意に定義した割合に基づくことだけではなく、試験における治療Bおよび非治療Bへの割り付け比率に基づくことも可能である。したがって、平均化の具体的な方法は明確に記述されるべきである。平均が試験における割り付け比率に基づいて定義される場合には、それが事前に指定された割り付け比（例：1:1）に基づくのか、各層における最終的な観察比率に基づくのかを明確にする必要がある。例えば、安全性上の理由により試験途中で因子Bへの割り付けが中止された場合には、これらの比率は大きく変化する可能性がある。

スコーピングレビューでは、ファクトリアル試験プロトコルに関する19の推奨事項が特定され、これらはデルファイ調査において評価された。複数の推奨事項が同一のSPIRIT項目に関連していた場合においても、推奨事項は個別に評価された。デルファイ調査には104名が参加した。内訳は、統計専門家60名、臨床試験研究者25名、試験マネージャー7名、主任研究者としての経験を有する者19名、学術誌編集者としての経験を有する者17名、患者・市民参画メンバー2名であった（複数の役割を選択可）²²⁾。合同会議において評価対象となる基準を満たした推奨事項は20件であり、そのう

ち1件は第2回デルファイ調査で追加されたものである。合同会議の後、オンラインおよび電子メールによる議論を経て、拡張版チェックリストが最終化された。

ファクトリアル試験を記述する際に用いられる用語にはばらつきがあるため、本声明の各項目は、SPIRITの原文項目を置き換える形で作成されている。改訂版チェックリストを使用する際は、表2に示された主要用語の定義を参照することが推奨される。

本報告書には、SPIRITファクトリアル試験拡張版において修正された項目の簡潔な説明が含まれている。各項目の解釈に関する詳細および適切な報告例は、別途作成予定の解説・補足記事にて提示する。

ファクトリアル試験プロトコルのためのSPIRITチェックリスト拡張版【SPIRIT Checklist Extension for Factorial Trial Protocols】

項目1. SPIRIT 2013 原文項目：試験デザイン、対象集団、介入内容、および該当する場合は試験略称を明示する記述的なタイトル

ファクトリアル試験プロトコル拡張版：タイトル中に「ファクトリアルランダム化試験」であることを明示する。

ファクトリアル試験には特有の方法論的特徴があるため、試験デザインを明示することで読者がその含意や潜在的な限界を考慮する可能となる^{4,6,7,10,23,24)}。

項目6a. SPIRIT 2013 原文項目：研究課題の記述および試験を実施する根拠の提示。

ファクトリアル試験プロトコル拡張項目：ファクトリアルデザインを採用する理由の説明（交互作用の仮説の有無を含む）。

ファクトリアル試験は、異なる研究仮説（すなわち、推定対象）に対応するために用いることができる（Box参照）。例えば、因子試験ではサンプルサイズを増加させることなく、1つの試験内で複数の介入を評価することができる（いわゆる「2因子試験」として言及されることが多い）。これにより、介入間の交互作用（すなわち、治療Aの効果が他の因子の有無によって変化するかどうか）を評価したり、最適な介入の組み合わせを特定したりすることができる。ファクトリアルデザインを採用す

る理由および交互作用の仮説の有無を明確にすることで、読者が本試験の主要な目的や、ファクトリアルデザイン採用にあたっての前提条件を理解することができる^{3, 6-8, 24)}。

項目 7. SPIRIT 2013 原文項目：具体的な研究目的または仮説の記述。

ファクトリアル試験プロトコル拡張項目：主要な比較を構成する治療群の組み合わせを明示する記述。

ファクトリアル試験は、複数の介入を多様な方法で比較・評価することを可能にする。例えば、因子A および B を用いた 2×2 のファクトリアル試験において、介入 A の治療効果は以下のいずれかの比較により推定可能である：(1) A 群対 A 非介入群、(2) A 単独群対 A および B いずれの介入も受けない群、(3) A と B の併用群対 B 単独群。これらの比較はそれぞれ異なる推定を対象とし、異なる仮定を必要とする可能性がある^{6, 8, 13)}。推定対象とは、試験を通じて研究者が推定しようとする治療効果を定義する概念である^{13, 25, 26)}。

項目 8. SPIRIT 2013 原文項目：試験の種類、割付比、解析枠組みを含む試験デザインの記述。

ファクトリアル試験プロトコルの拡張項目：ファクトリアル試験の種類（完全因子デザインであるか部分因子デザインであるか）、因子の数、および各因子内の水準数を含む記述。

ファクトリアル試験には、さまざまな設計形式を用いることが可能である。最も基本的な設計は完全ファクトリアルデザインであり、参加者がすべての因子およびその水準の組み合わせに割り付けられる可能性を有する^{11, 27, 28)}。一部の因子組み合わせを省略する縮約ファクトリアルデザインや、特定の因子にのみ割り付け可能な参加者を含む部分ファクトリアルデザインでは、異なる方法論的アプローチが求められる^{3, 29)}。

項目 10. SPIRIT 2013 原文項目：参加者の選択基準および除外基準の記載。

ファクトリアル試験プロトコル拡張項目：各因子に対する適格基準の記載（該当する場合は因子間での差異も明示）。

因子間で適格基準に差異がある場合、サンプルサ

イズや解析方法の調整が必要となる可能性があり、解析時に適切に扱われないとバイアスが生じるおそれがある。特定の因子への割付に適格でない参加者は、当該因子に関する比較（および交互作用の評価）から除外するべきである。これらの参加者を含めて解析を行うと、比較がランダム化に基づかないものとなり、交絡バイアスを生じる可能性がある^{3, 29)}。

項目 14. SPIRIT 2013 原文項目：研究目的を達成するためには必要な参加者数の推定値と、その算出方法の記載。

ファクトリアル試験プロトコル拡張項目：各主要な比較に対してサンプルサイズの算出方法を記載（交互作用を仮定したかどうかも含めて明示）。

適切なサンプルサイズの算出には、ファクトリアルデザインを採用する具体的な理由と、試験を実施する際に用いる方法論の両方を考慮する必要がある。例えば、介入間の交互作用の有無を評価することを目的とした試験では、各介入の主効果を評価する試験よりも、一般により大きなサンプルサイズを必要とする。また、2 因子試験においては、解析方法（ファクトリアル試験として扱うか、多群比較試験として扱うか）によって、必要なサンプルサイズが異なる。さらに、ファクトリアル試験によっては計画された主要な比較ごとに必要なサンプルサイズが異なる場合がある。これは、各比較で期待される効果量が異なる場合や因子ごとに主要アウトカムが異なる場合に生じ得る^{8, 30)}。

項目 16a. SPIRIT 2013 原文項目：割付方法の生成手順、および層別に用いる因子の一覧を記載する。

ファクトリアル試験プロトコル拡張項目：該当する場合は、参加者が異なる時点で各因子に割り付けられるかどうかを記載する。

一部のファクトリアル試験では、参加者が異なる時点で各因子に対してランダム化される場合がある。例えば、診断時に因子 A に対するランダム化を行い、治療 A 完遂後に因子 B に対するランダム化を実施する場合がある。各因子に対するランダム化の時点は、ベースライン期間、追跡期間の長さ、介入間の交互作用が生じる可能性といった試験の主要な設計要素に影響を与える⁴⁾。

項目 20a. SPIRIT 2013 原文項目：主要アウトカムおよび副次アウトカムの解析に用いる統計手法を記載する。統計解析計画書（SAP）がプロトコル外に存在する場合は、その参照先を明示する。

ファクトリアル試験プロトコル拡張項目：各主要な比較に対して、主要および副次アウトカムの解析に用いる統計手法を記載する。

- ・各主要な比較における目標とする治療効果が、他の因子の存在下での効果を指すのか、それとも他の因子が存在しない場合の効果を指すのかを明確にする。

推定される治療効果の定義を正確に理解することは、試験結果を適切に解釈するために不可欠である。しかし、こうした点は試験方法の記載だけでは必ずしも明確ではない場合がある³¹⁻³³⁾。ファクトリアル試験に特有の課題として、治療効果の推定対象となる群と実際に比較に用いられる治療群で一致しない場合がある^{13, 34)}。例えば、多くのファクトリアル試験では効率性の観点から、A 群と非 A 群を比較するファクトリアル解析が用いられるが、実際に関心があるのは、A 単独とコントロールとの比較（すなわち B が存在しない状況での A の効果）、あるいは、B の有効性が示されている場合には、A と B の併用と B 単独との比較（すなわち B が存在する状況での A の効果）である¹³⁾。目標とする治療効果について、他の因子の存在下での効果あるいは非存在下での効果であるかを含めて明確に記載することで、読者は試験で問われている正確な研究課題を理解することができる^{13, 25, 31, 32)}。目標とする治療効果は「推定量」と呼ばれ、各比較に対して明確に定義する必要がある^{13, 25)}。

・解析の基本方針

関心のある推定量に応じて、ファクトリアル試験の解析には異なる統計手法が用いられる。介入の評価において最も一般的な方法は、ファクトリアル解析（マージナル解析）^{4, 6, 8, 13, 35, 36)}と、マルチアーム解析（インサイド・ザ・テーブル解析）の 2 つである^{4, 6-8, 12-14, 19, 23, 35, 36)}。表 1 を例にとると、ファクトリアル解析では、因子 A に割り付けられたすべての参加者（すなわち、A 有効 +B 有効群および A 有効 +B 対照群）を、A に割り付けられていないすべ

ての参加者（A 対照 +B 有効群および A 対照 +B 対照群）と比較する。多群解析では、各治療群を基準群と比較する。例えば、A 有効 +B 対照群、A 対照 +B 有効群、A 有効 +B 有効群を、それぞれ A 対照 +B 対照群と比較する。これら 2 つの解析アプローチはそれぞれ異なる利点を有し、前提とする仮定も異なる（Box 参照）。

・解析アプローチの選定方法

解析手法としてファクトリアル解析または多群解析のいずれを採用するかを判断するために、研究者が事前に交互作用の有無を検定することがある。しかし、このようなアプローチではバイアスが生まれる可能性がある¹⁹⁾。そのため、一般的にはこのアプローチは推奨されない。しかしながら、もし当該アプローチを採用する場合には、解析手法の統計学的含意を読者が正しく理解できるよう明確に記載することが重要である。

・統計的交互作用の評価方法

ファクトリアル試験においては、治療間の交互作用の有無を評価することが求められる。これは、解析が「治療間に交互作用が存在しない」という仮定に依拠している場合、あるいは交互作用そのものが主要な関心対象である場合による^{4, 6-8, 12, 13, 24)}。交互作用の評価方法を明示することで、読者は採用された解析手法の妥当性を理解することができる。

・各因子の効果を推定する際に、他の因子を共変量として調整するかどうか

ファクトリアル解析では、統計モデルに他因子への割り付けを表す項を含めることで、参加者が他の因子に割り付けられていたかどうかを調整することが可能である^{4, 8, 13, 30)}。この調整により統計的検出力が向上する可能性があり、場合によっては他の因子を調整しないことにより特定の推定量に対してバイアスが生じることがある¹³⁾。

・因子への随時割り付け（随時募集）をどのように取り扱うか

随時割り付けとは、一部の参加者が特定の因子に対してランダム化されない状況（例：ある因子への登録が一時中断または終了された場合など）を指し、

解析時に適切に取り扱わなければバイアスが生じる可能性がある^{3, 29)}。したがって、特定の因子に対してランダム化されなかった参加者が、その因子に関する解析から除外されたかどうかを明示することはバイアスのリスクを理解する上で重要である。

項目 21b. SPIRIT 2013 原文項目：試験中止に関する最終決定を行う者を含め、どの者が中間解析の結果にアクセスするかを明示した上で、すべての中間解析および中止基準の内容を記載する。

ファクトリアル試験に対する拡張項目：該当する場合には、すべての中間解析および中止基準について説明する。主要な比較間で中止基準に差異がある場合には、違いと理由をそれぞれ明確に記載する。

中間解析は、安全性、有効性、または無益性の評価を目的として実施されることが多い。中止基準は因子ごとに異なる場合がある²⁹⁾。1つの因子が他の因子より先に中止された場合、ランダム化の方法や比較対象の選定、解析対象集団に影響を及ぼす可能性がある^{3, 29, 37)}。

議論 [Discussion]

SPIRIT 2013 声明は臨床試験プロトコルの報告に関する包括的なチェックリストを提供しており、試験の計画、実施、解析に関する明確性を確保することで、適切な試験運営および評価を促進することを目的としている^{1, 2)}。本稿は、SPIRIT 2013 声明の拡張版として、ファクトリアル試験プロトコルの報告に関する指針を提供するものである。ファクトリアル試験プロトコルの明確な記載は、試験計画に基づく手順の明瞭性と網羅性を担保し、研究倫理審査委員会や査読者など、プロトコルを評価する読者による適切な審査を促進する上で有用である。本声明は、ファクトリアル試験プロトコルにおける追加的な報告要件の概要を示すものであるが、各項目の詳細な解説および適切な記載例を提供する補足文書と併せて、本チェックリストを活用することを推奨する。

本チェックリストは、ファクトリアル試験プロトコルの報告において、最低限記載すべき必須項目を示している。一部の試験においては、プロトコルに追加項目を記載する必要がある。たとえば、主要アウトカムまたは副次アウトカムが因子ごとに異なる

場合には、その旨をプロトコルに明記しなければならない。同様に、多重検定が問題となる可能性がある場合には、その対処方法についてプロトコル内で明示する必要がある。

本拡張版は、ファクトリアル試験の報告に関する CONSORT 拡張版と連携して作成された。これら 2 つの拡張ガイドラインは、試験プロトコルから試験結果の最終報告に至るまで一貫性のある報告を行うための枠組みを提供する。本ガイドラインおよびその他の SPIRIT 声明の最新版は、オンラインで閲覧可能である (<https://www.spirit-statement.org/>)。

限界 [Limitations]

本拡張版は、EQUATOR の最良の方法論的枠組みに基づいて作成されているが、いくつか限界がある。まず、本拡張版は各因子の結果が同一の論文内で同時に報告される試験を対象として作成されている。たとえば、因子ごとに必要なサンプルサイズや追跡期間が異なる場合など、同一論文内での同時報告が常にできるとは限らない。各因子の結果を別々の論文で報告する計画がある場合には、その旨をプロトコル内に記載する必要がある。第二に、本調査には多様な背景を有する多数の対象者がデルファイ調査に参加したが、参加者は自己選択によるものであり、そのことが結果に影響を及ぼした可能性がある。第三に、合同会議のパネリストは、ランダム化比較試験に関連する専門性および役割（例：学術誌編集者）に基づいて選定されており、ファクトリアル試験を実施する研究者全体の意見を必ずしも反映しているとは限らない。ただし、本ガイドラインの策定には、ファクトリアル試験に関する報告推奨事項の厳密なスコーピングレビューを含むエビデンスに基づくアプローチが採用されており、これにより前述の限界による影響が少なくなる可能性がある。

結論 [Conclusions]

本コンセンサス声明は、SPIRIT 2013 声明の拡張版として、ファクトリアル試験プロトコルの報告に関する具体的な指針を示すものである。本指針は、これらのプロトコルの報告における透明性と完全性の向上に寄与するものである。

参考文献 [REFERENCES]

1. Chan AW, Tetzlaff JM, Altman DG, et al.: SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med.* 2013; **158**(3): 200–207. doi: 10.7326/0003-4819-158-3-201302050-00583.
2. Chan AW, Tetzlaff JM, Gøtzsche PC, et al.: SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ.* 2013; **346**: e7586. doi: 10.1136/bmj.e7586.
3. Green S, Liu PY, O’Sullivan J: Factorial design considerations. *J Clin Oncol.* 2002; **20**(16): 3424–3430. doi: 10.1200/JCO.2002.03.003.
4. Kahan BC, Tsui M, Jairath V, et al.: Reporting of randomized factorial trials was frequently inadequate. *J Clin Epidemiol.* 2020; **117**(117): 52–59. doi: 10.1016/j.jclinepi.2019.09.018.
5. Lubsen J, Pocock SJ: Factorial trials in cardiology: pros and cons. *Eur Heart J.* 1994; **15**(5): 585–588. doi: 10.1093/oxfordjournals.eurheartj.a060552.
6. McAlister FA, Straus SE, Sackett DL, Altman DG: Analysis and reporting of factorial trials: a systematic review. *JAMA.* 2003; **289**(19): 2545–2553. doi: 10.1001/jama.289.19.2545.
7. Montgomery AA, Astin MP, Peters TJ: Reporting of factorial trials of complex interventions in community settings: a systematic review. *Trials.* 2011; **12**: 179. doi: 10.1186/1745-6215-12-179.
8. Montgomery AA, Peters TJ, Little P: Design, analysis and presentation of factorial randomised controlled trials. *BMC Med Res Methodol.* 2003; **3**: 26. doi: 10.1186/1471-2288-3-26.
9. Byth K, Gebski V: Factorial designs: a graphical aid for choosing study designs accounting for interaction. *Clin Trials.* 2004; **1**(3): 315–325. doi: 10.1191/1740774504cn026oa.
10. Dakin H, Gray A: Economic evaluation of factorial randomised controlled trials: challenges, methods and recommendations. *Stat Med.* 2017; **36**(18): 2814–2830. doi: 10.1002/sim.7322.
11. Dakin HA, Gray AM, MacLennan GS, Morris RW, Murray DW: Partial factorial trials: comparing methods for statistical analysis and economic evaluation. *Trials.* 2018; **19**(1): 442. doi: 10.1186/s13063-018-2818-x.
12. Freidlin B, Korn EL: Two-by-two factorial cancer treatment trials: is sufficient attention being paid to possible interactions? *J Natl Cancer Inst.* 2017; **109**(9):109. doi: 10.1093/jnci/djx146.
13. Kahan BC, Morris TP, Goulão B, Carpenter J: Estimands for factorial trials. *Stat Med.* 2022; **41**(22): 4299–4310. doi: 10.1002/sim.9510.
14. Korn EL, Freidlin B: Non-factorial analyses of two-by-two factorial trial designs. *Clin Trials.* 2016; **13**(6): 651–659. doi: 10.1177/1740774516659472.
15. McClure LA, Coffey CS, Howard G: Monitoring futility in a two-by-two factorial design: the SPS3 experience. *Clin Trials.* 2013; **10**(2): 250–256. doi: 10.1177/1740774512474374.
16. Lin DY, Gong J, Gallo P, Bunn PH, Couper D: Simultaneous inference on treatment effects in survival studies with factorial designs. *Biometrics.* 2016; **72**(4): 1078–1085. doi: 10.1111/biom.12507.
17. Bria E, Di Maio M, Nisticò C, et al.: Factorial design for randomized clinical trials. *Ann Oncol.* 2006; **17**(10): 1607–1608. doi: 10.1093/annonc/mdl106.
18. Byar DP: Some statistical considerations for design of cancer prevention trials. *Prev Med.* 1989; **18**(5): 688–699. doi: 10.1016/0091-7435(89)90040-6.
19. Kahan BC: Bias in randomised factorial trials. *Stat Med.* 2013; **32**(26): 4540–4549. doi: 10.1002/sim.5869.
20. Kahan BC, Hall SS, Beller EM, et al.: Reporting of factorial randomized trials: extension of the CONSORT 2010 Statement. *JAMA.* doi: 10.1001/jama.2023.19793. Published online December 5, 2023.
21. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D: SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf.* 2016; **25**(12): 986–992. doi: 10.1136/bmqs-2015-004411.
22. Hall SS, Juszczak E, Birchenall M, et al.: Development of extensions to SPIRIT and CONSORT guidelines: the Reporting Factorial Trials (RAFT) Study. [Accessed November 3, 2023].
23. Juszczak E, Altman DG, Hopewell S, Schulz K: Reporting of multi-arm parallel-group randomized trials: extension of the CONSORT 2010 statement. *JAMA.* 2019; **321**(16): 1610–1620. doi: 10.1001/jama.2019.3087.
24. Mdege ND, Brabyn S, Hewitt C, Richardson R, Torgerson DJ: The 2 × 2 cluster randomized controlled factorial trial design is mainly used for efficiency and to explore intervention interactions: a systematic review. *J Clin Epidemiol.* 2014; **67**(10): 1083–1092. doi: 10.1016/j.jclinepi.2014.06.004.
25. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. [Accessed November 3, 2023].
26. Kahan BC, Cro S, Li F, Harhay MO: Eliminating ambiguous treatment effects using estimands. *Am J Epidemiol.* 2023; **192**(6): 987–994. doi: 10.1093/aje/kwad036.
27. Allore HG, Murphy TE: An examination of effect estimation in factorial and standardly-tailored designs. *Clin Trials.* 2008; **5**(2):121–130. doi: 10.1177/1740774508089278.
28. Baker TB, Smith SS, Bolt DM, et al.: Implementing clinical research using factorial designs: a primer. *Behav Ther.* 2017; **48**(4):567–

580. doi: 10.1016/j.beth.2016.12.005.
29. White IR, Choodari-Oskooei B, Sydes MR, *et al.*: Combining factorial and multi-arm multi-stage platform designs to evaluate multiple interventions efficiently. *Clin Trials*. 2022; **19**(4): 432–441. doi: 10.1177/17407745221093577.
30. Curran D, Sylvester RJ, Hoctin Boes G: Sample size estimation in phase III cancer clinical trials. *Eur J Surg Oncol*. 1999; **25**(3): 244–250. doi: 10.1053/ejso.1998.0635.
31. Cro S, Kahan BC, Rehal S, *et al.*: Evaluating how clear the questions being investigated in randomised trials are: systematic review of estimands. *BMJ*. 2022; **378**: e070146. doi: 10.1136/bmj-2022-070146.
32. Kahan BC, Morris TP, White IR, Carpenter J, Cro S: Estimands in published protocols of randomised trials: urgent improvement needed. *Trials*. 2021; **22**(1): 686. doi: 10.1186/s13063-021-05644-4.
33. Mitroiu M, Teerenstra S, Oude Rengerink K, Pétavy F, Roes KCB: Estimation of treatment effects in short-term depression studies: an evaluation based on the ICH E9(R1) estimands framework. *Pharm Stat*. 2022; **21**(5): 1037–1057. doi: 10.1002/pst.2214.
34. Tian Z, Esserman D, Tong G, *et al.*: Sample size calculation in hierarchical 2×2 factorial trials with unequal cluster sizes. *Stat Med*. 2022; **41**: 645–664. doi: 10.1002/sim.9284.
35. Pocock SJ, Clayton TC, Stone GW: Challenging Issues in clinical trial design: part 4 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol*. 2015; **66**(25): 2886–2898. doi: 10.1016/j.jacc.2015.10.051.
36. Leifer ES, Troendle JF, Kolecki A, Follmann DA: Joint testing of overall and simple effects for the two-by-two factorial trial design. *Clin Trials*. 2021; **18**(5): 521–528. doi: 10.1177/17407745211014493.
37. Slud EV: Analysis of factorial survival experiments. *Biometrics*. 1994; **50**(1): 25–38.