

## CONSORT 2010 statement: extension to randomised crossover trials の和訳

著者 NKerry Dwan<sup>1</sup>, Tianjing Li<sup>2</sup>, Douglas G Altman<sup>3</sup>, Diana Elbourne<sup>4</sup>

翻訳 馬場 亜沙美 (BABA Asami)<sup>1\*</sup>, 鈴木 直子 (SUZUKI Naoko)<sup>1</sup>, 野田 和彦 (NODA Kazuhiko)<sup>1</sup>, 波多野 絵梨 (HATANO Eri)<sup>1</sup>, 高橋 徳行 (TAKAHASHI Noriyuki)<sup>1</sup>, 新林 史悠 (SHINBAYASHI Fumiharu)<sup>1</sup>, 板橋 怜央 (ITABASHI Reo)<sup>1</sup>, 柿沼 俊光 (KAKINUMA Toshihiro)<sup>1</sup>, 山本 和雄 (YAMAMOTO Kazuo)<sup>1</sup>

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, クロスオーバー

### Japanese translation of “CONSORT 2010 statement: extension to randomised crossover trials”

#### Authors

Kerry Dwan<sup>1</sup>, Tianjing Li<sup>2</sup>, Douglas G Altman<sup>3</sup>, Diana Elbourne<sup>4</sup>

#### Translator

Asami Baba<sup>1\*</sup>, Naoko Suzuki<sup>1</sup>, Kazuhiko Noda<sup>1</sup>, Eri Hatano<sup>1</sup>, Noriyuki Takahashi<sup>1</sup>, Fumiharu Shinbayashi<sup>1</sup>, Reo Itabashi<sup>1</sup>, Toshihiro Kakinuma<sup>1</sup>, Kazuo Yamamoto<sup>1</sup>

\*Correspondence author: Asami Baba

Keywords: Human clinical trial, Human trials, Food for specified health uses, Food with function claims, Research design, CONSORT, Guideline, Outcome, SPIRIT, Crossover

#### Affiliation (Author)

<sup>1</sup> Review Production and Quality Unit, Editorial and Methods Department, Cochrane Central Executive, Cochrane, St Alban's House, London SW1Y 4QX, UK

<sup>2</sup> Center for Clinical Trials and Evidence Synthesis, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

<sup>3</sup> Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

<sup>4</sup> London School of Hygiene and Tropical Medicine, Department of Medical Statistics, London, UK

#### Affiliated institution

<sup>1</sup> ORTHOMEDICO Inc.

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

\* 責任著者: 馬場 亜沙美

所属 (翻訳)

<sup>1</sup> 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「CONSORT 2010 statement: extension to randomised crossover trials」の本文を翻訳したものである。

## 要約 [Abstract]

ランダム化比較試験の報告の質が最適ではないことを示すエビデンスがある。透明性が欠如している報告は、読者が試験結果の信頼性と妥当性を判断することを妨げ、研究者がシステマティックレビューのために情報を抽出できなくなり、その結果、研究の無駄をもたらす。CONSORT (Consolidated Standards of Reporting Trials) 声明は、ランダム化比較試験の報告を改善するために作成された。この声明の主な焦点は、2つの治療群を持つ並行群間比較試験である。クロスオーバー試験は、参加者が一連の介入にランダムに割り付けられる、慢性疾患に特化したタイプの試験である。このタイプの試験は参加者自身が対照となるため、有用で効率的なデザインである。しかし、クロスオーバー試験の報告はばらつきがあり、不完全であったため、臨床上の意思決定や将来の研究者による使用の妨げとなっていた。我々は、ランダム化クロスオーバー試験のより良い報告の促進を目的とした、CONSORT 声明の拡張版を紹介する。CONSORT 2010 チェックリストはクロスオーバーデザイン用に改訂され、透明性を高めるために修正されたフローチャートとベースライン表が導入されている。CONSORT のクロスオーバーチェックリスト項目に関する優れた報告例とエビデンスに基づく根拠が示されている。

### 要約ポイント [Summary points]

- CONSORT 声明は、すべてのランダム化試験について、報告すべき最小限の 25 項目を根拠と模範例とともに示している。
- CONSORT のクロスオーバー試験拡張版は、CONSORT 声明の 14 項目を拡張するものである。
- CONSORT のクロスオーバー試験拡張版の使用により、ランダム化クロスオーバー試験の報告が改善される。

ランダム化比較試験 (RCT) の不十分な報告は、介入効果の推定におけるバイアスと関連している<sup>1,2)</sup>。また、ランダム化試験の質の批判的評価も損なわれる。これは、個々の試験結果の妥当性を評価する際や、システマティックレビューを実施する上で重要である。この問題に対処するため、CONSORT 声明が作成され、RCT の報告に関する一連の推奨事項が含まれている<sup>3)</sup>。CONSORT 声明は、RCT の報告書に含めるべき必須項目のチェックリストと、群割り付け前から最終解析までの試験参加者の流れを記録する図で構成されている。これらの項目は、可能な場合はエビデンスに基づいている。チェックリスト項目の根拠に関する「解説と詳細」は、付随する論文として示されている<sup>4)</sup>。現在、多くのジャーナルが RCT の報告を CONSORT 声明の推奨事項に準拠することを求めている<sup>5)</sup>。

CONSORT 声明の主な焦点は、個別ランダム化、並行群間、優越性デザインを用いた 2つの治療群 (2つの「群」) を有する最も一般的なタイプの RCT である<sup>3)</sup>。CONSORT 声明のほぼすべての要素は、他のデザインの RCT にも同様に適用されるが、一部の要素は適応が必要であり、場合によっては追加の問題を検討する必要がある。CONSORT グループのメンバーは、様々なタイプの介入やデータについて CONSORT 声明を補強するいくつかの拡張版論文を発表している。また、クラスターランダム化試験<sup>6)</sup>、非劣性試験と同等性試験<sup>7)</sup>、N-of-1 試験<sup>8)</sup>、プラグマティック試験 (pragmatic trials)<sup>9)</sup>、個人内試験 (within person trials)<sup>10)</sup> など、様々な試験デザインについて CONSORT 2010 の拡張版も発表されている。このシリーズの一部として、本稿では CONSORT 2010 声明の推奨事項を単純クロスオーバー RCT に拡張する。その試験デザインとし

ては、参加者が2つの期間にわたって2つの治療を順番に受け、治療を受ける順序がランダム化されている。

### 本稿の対象範囲【Scope of this paper】

まず、クロスオーバー試験の主要な方法論的特徴を要約する。次に、クロスオーバー試験がどの程度一般的であるかについての経験的証拠を検討し、そのような試験の報告の質に関する公表された研究をレビューする。これらの文献レビューの後、クロスオーバー試験に適応した CONSORT チェックリストの修正を提案し、優れた報告の実例を示している。本ガイドラインでは、すべての参加者が2つの順序のいずれかで2つの介入を受ける（2×2 または AB/BA デザインとして知られる）試験に焦点を当てる。これは最も単純で一般的なランダム化クロスオーバー試験である。推奨事項のほとんどは、より複雑なデザイン（2つ以上の介入、期間、または順序）にも適用される。また、別のセクションでは、2つ以上の介入を比較する試験で生じる特定の問題について簡単に説明する。

### ランダム化クロスオーバー試験の方法論的特徴【Methodological features of randomised crossover trials】

並行群間比較試験とは対照的に、クロスオーバー試験では各個人が複数の介入を受けるが、順番はランダムである。このように、各参加者は自分自身の対照として働く。このような事前に決められたデザインは、一部の参加者がコンプライアンス違反やレスキュー治療により「クロスオーバー」する試験や、対照群の参加者全員が本試験の終了時に実験的治療に「クロスオーバー」する機会が与えられる試験と混同してはならない。Zeng らの調査により、「クロスオーバー割付け」と表示された試験登録記録のほぼ4分の1（ $n=17/72$ ）では、参加者を介入順序にランダムに割り付けるランダム化クロスオーバーのデザインが使用されておらず、代わりに、参加者が試験の途中で介入を変更できることが発覚した<sup>11)</sup>。

ランダム化クロスオーバー試験には特有の課題がある。1つ目の課題は、「持ち越し効果（carry over effect）」の可能性である。すなわち、最初の介入の効果が2番目の介入期間まで持続し、観察された治

療間の差は、治療を受けた順番によって異なる（用語集は box 1 を参照）。持ち越し効果には様々な原因が考えられる。薬剤や他の治療が体内に残るという明らかな問題に加え、参加者における遅発的な反応は、以前の治療による副作用やその他の反応に影響される可能性がある。クロスオーバー試験では、最初の介入の終了と2番目の介入の開始の間に十分な「ウォッシュアウト」期間を設け、最初の介入による効果が2番目の介入期間のアウトカム測定に「持ち越し」しないようにすることが推奨される。

#### Box 1

##### 用語説明【Glossary】

**期間（period）**：1回の治療を受けた時間。

**順序（sequence）**：治療順序（AB, BA）、AB 試験群に割り付けられた参加者は最初に治療 A を受け、次に治療 B を受ける、BA 群ではその逆。

**参加者内変動（within participant variability）**：参加者内差の予想標準偏差

**ウォッシュアウト（washout）**：治療効果がなくなるまで治療を受けない、治療期間の間の時間。

**持ち越し効果（carry over effect）**：最初の介入の効果が2番目の介入期間にも持続すること。

**時期効果（period effect）**：介入効果とは無関係に、関心のあるアウトカムが時間とともに変化すること。

**参加者内比較（within participant comparison）**：参加者内比較では、各参加者が自分自身の対照として働くため、測定値が独立しておらず、各参加者の測定値間の相関が考慮されること。

もう一つの問題は「時期効果」であり、これは介入効果とは関係なく、関心のあるアウトカムが時間とともに変化する場合に起こる。例えば、病状が安定していない場合や、介入効果に季節性がある場合などが考えられる。

さらに問題なのは、最初の介入が大成功または失敗した場合、参加者が試験から脱落する可能性であり、これらの参加者の結果は解析に使用できない。

### デザイン【Design】

単純な AB/BA クロスオーバーデザインの特に優れた点は、両方の介入が同じ参加者を用いて評価さ

れることであり、これは群レベルではなく個人レベルでの比較を可能にする。また、クロスオーバー試験の参加者は、2つの介入の経験を比較することで好みを明らかにすることができるが、これは参加者が1つの介入しか受けない並行群間デザインでは不可能である<sup>12)</sup>。

方法論的に重要な問題は、クロスオーバーデザインの使用が正当化されるかどうかである。クロスオーバー試験は、少なくとも試験期間中、慢性的または比較的安定した病状や疾患（例えば、多発性硬化症や関節リウマチ）の対症療法（すなわち、痛みなどの症状に対する治療）に最も適しており、さらに治療効果が可逆的で短期間である場合にも適している。クロスオーバーデザインは、対象の病状が治療可能である場合や、参加者が試験期間中に死亡する可能性がある場合には不適切である。しかし、クロスオーバーデザインは、あまり適切でない状況でも一般的に用いられている。例えば、妊娠は不妊治療の意図されたアウトカムである。もし女性が試験の最初の期間中（すなわち、クロスオーバーの前）に妊娠した場合、その女性はそれ以降のフェーズの試験から除外される。それにもかかわらず、クロスオーバーデザインはこの分野で擁護されており<sup>13)</sup>（例えば、妊娠は統計的に「ランダムな欠測（missing at random）」として扱えることが示唆されている<sup>14)</sup>）、一般的であり続けている<sup>15)</sup>。

このような試験のサンプルサイズの計算は、反応の参加者内変動に基づいている。異なる治療に対する参加者の反応の間に高い正の相関がある場合、クロスオーバーデザインは並行群間デザインよりもはるかに効率的である。並行群間デザインに比べると、クロスオーバー試験では、ある目標の効果量と第1種の過誤率に対して同じ検出力を得るために、必要な参加者数が少なくなる。

クロスオーバー試験には弱点がある。特に、前述したように持ち越し効果が起こりうる。参加者は最初の治療で脱落し、2番目の治療を受けられなくなる可能性がある。脱落は副作用に関係するかもしれない。

## 解析【Analysis】

クロスオーバー試験の解析は、ペアデータに基づくべきである<sup>16-18)</sup>。推定アプローチは、同じ個人に

おける繰り返し測定の間接性を考慮すべきである。有意差検定は、連続的な反応の場合は参加者内差に基づく対応のある $t$ 検定（持ち越し効果や時期効果がないと仮定する）、二項反応の場合はMainland-Gart検定などの手順を用いるべきである<sup>19,20)</sup>。

以前、推奨されていたが後に批判されたクロスオーバー試験の解析方法は、持ち越し効果を検定し、統計的に有意であれば2番目の期間のデータを破棄し、最初の期間のデータのみを解析するものであった。すなわち、最初の期間のデータはあたかも並行群間比較試験のように解析される。Freeman<sup>21)</sup>は、この方法には欠陥があり、バイアスのかかった解答につながることを示した（これは、2つの解析の選択が予備的な仮説検定の結果に基づいている場合に起こりうる）。Senn<sup>17)</sup>らは、2期間2治療クロスオーバーデザインの使用は、持ち越し効果が最小であるという前提の上に成り立っていると主張している。

クロスオーバー試験特有のもう一つの統計学の問題は、起こりうる時期効果や持ち越し効果の調整が必要なことである。統計モデルには、持ち越し効果のパラメータを含めることができる。AB/BAクロスオーバーデザインでは、「持ち越し」と「治療と期間の交互作用」という用語が互換的に使用されることがある。持ち越し効果は推定可能であるが、Senn<sup>17)</sup>らは、持ち越し効果を用いて治療効果を調整する価値はほとんどないと主張している。その理由として、このような調整は、起こりうる持ち越し効果の性質に関する仮定に依存し、主な治療効果を推定するための統計的効率を低下させるからである。

時期効果は、解析で対処し調整することができる。AB/BAクロスオーバーデザインでは、同数の参加者が各順序に割り当てられている場合、平均して時期効果は治療効果の推定にバイアスをかけない。しかし、時期効果は、どの程度の治療効果がランダム変動に起因するかを妨げるため、分散推定に影響を与える。著者は、読者が時期効果の程度を理解できるようにデータを提示し、時期効果が解析で調整されたかどうか、またそのような決定が事前に行われたかどうかを明確に伝えることが重要である。

## ランダム化クロスオーバー試験はどの程度行われているのか？

### 【How common are randomised crossover trials?】

2000年12月に発表されたPubMedに索引付けされたすべてのRCTに関する詳細なレビューでは、74% (383/519) の試験が並行群間デザインで、22% (116/519) がクロスオーバー試験であった<sup>22)</sup>。2000年12月にMedlineに索引付けされた試験のうち、22% (116/526) がクロスオーバー試験であり、そのほとんどが2つの治療を用い (72%)、2つの期間を設けていた (64%)<sup>23)</sup>。2006年12月に発表されたPubMedに索引付けされたすべてのRCTに関するレビューでは、77% (477/616) が並行群間デザインで、16% (100/616) がクロスオーバー試験であった<sup>24)</sup>。2007年から2010年の間にClinicalTrials.govに登録された介入試験に関するレビューでは、11.2% (4351/38969) がクロスオーバー試験であった<sup>25)</sup>。2012年12月に発表されたPubMedに関する最新のレビューでは、RCTの8.7% (98/1122) がクロスオーバーデザインであった<sup>6)</sup>。

## ランダム化クロスオーバー試験の報告の質とは？

### 【What is the quality of reporting of randomised crossover trials?】

CONSORTに関連したRCTの報告の質に関する論文は比較的多いが、クロスオーバー試験の報告の質を明確に検討した論文はほとんどない。Millsらは、2000年12月にMedlineに索引付けされたランダム化クロスオーバー試験では、デザイン、解析、解釈に関する詳細が省略されていることが多いことを報告した<sup>23)</sup>。一方、ほとんどの試験でウォッシュアウト期間が報告されており (69%, 87/127)、解析にペアデータを用いたことも報告されていた (95%, 121/127)。Gewandterらは、1993年から2013年の間に発表された慢性疼痛の薬物治療に関する124のクロスオーバー試験を調査した。その結果、28% (35/124) の試験がベースラインとウォッシュアウト期間後の疼痛レベルを報告しており、参加者内の変動に基づくことを明確に示したサンプルサイズの計算を報告していたのはわずか31% (23/75) であることがわかった<sup>27)</sup>。Straubeらは、1990年から2014年の間に発表され、PubMedに索引付けされた慢性疼痛疾患に関する98件のクロス

オーバー試験を検討した。その結果、有害事象は試験報告の抄録ではあまり報告されておらず、論文の全文でもほとんど報告されていないことと、治療期間別の内訳が示されているのはわずか23% (23/98) であることがわかった<sup>28)</sup>。Zengらは、2014年9月にClinicalTrials.govから分析した54件の第III相ランダム化クロスオーバー試験のうち、3分の2近くが単純なAB/BAデザインであり、ほとんどの試験 (87%, 47/54件) が試験期間中の参加者の流れについて十分な情報を示していることを明らかにした<sup>11)</sup>。ベースライン特性は、すべての参加者を一つの群として報告されることが最も多く (59%, 32/54例)、主要アウトカムと有害事象は「介入ごと」に報告されることが最も多かった (それぞれ81%, 44/54例; 83%, 45/54例)。ベースライン特性、アウトカム評価指標、有害事象の結果報告は、ほとんどクロスオーバーデザインを十分に反映していないようであった。

いくつかの研究では、メタアナリシスに関連してランダム化クロスオーバー試験の報告について検討を行った<sup>29-31)</sup>。その結果、データがメタアナリシスに含めることができるように不適切に報告されることが多いことがわかった。

これらの研究は、その問題が数年にわたって改善されていないことを示しており、これらの研究のほとんどは、ランダム化クロスオーバー試験の報告に関するガイダンスを求めている。

## 本 CONSORT 拡張版の作成に用いた方法

### 【Methods used to develop this CONSORT extension】

2002年5月、CONSORTの著者数名が米国バージニア州アーリントンに集まり、2001年のCONSORT声明を様々な試験デザインに拡張することを検討した。2002年から2003年にかけて、Doug AltmanとDiana Elbourneによって、CONSORT声明をクロスオーバー試験に拡張する最初の論文草稿が作成された。2010年、CONSORT声明が更新された。2014年にKerry DwanとTianjing Liがグループに加わったことにより、クロスオーバー試験への拡張版の作業が進展した。チェックリストと解説文は、公表されたランダム化試験のレビュー (上記参照) に基づいて検討され、2014年から2018年にかけて、著者間で行われた複数回の電話会議により完

成した。CONSORT グループのガイダンスに従い、EQUATOR 運営グループの議長でもある CONSORT グループの幹部メンバー (Doug Altman) を参加させた。論文草稿は、CONSORT グループ全体とその他の選ばれた個人に配布され、そのフィードバックを考慮して論文が改訂され、CONSORT グループの幹部によって承認された。

## ランダム化クロスオーバー RCT の CONSORT チェックリスト

### 【CONSORT checklist for randomised crossover RCTs】

チェックリストの項目について検討を行い、ランダム化クロスオーバー試験に対して標準的な CONSORT 項目からの変更点に焦点を当てる。背景を説明し、優れた報告例を1つ以上示している。また、修正を提案しないが、クロスオーバー RCT を実施するにあたって、特別な考慮が必要とするその他のチェックリスト項目についても述べている。表1は、ランダム化クロスオーバー試験に対して標準的な CONSORT チェックリストからの修正案を示している。

## タイトルと抄録【Title and abstract】

### 項目 1a：タイトル【Item 1a: Title】

タイトルにランダム化クロスオーバー試験であることを記載。【Identification as a randomised crossover trial in the title.】

**標準的な CONSORT 項目** - タイトルにランダム化試験であることを記載。

**例 1-**「イチョウ葉が中国の正常眼圧緑内障患者の視野とコントラスト感度に及ぼす効果：ランダム化クロスオーバー試験」<sup>33)</sup>。

**例 2-**「標準化された朝食の一部として提供される未発酵および発酵全粒ライ麦クリスピーパンが食欲、食後血糖値、インスリン反応に及ぼす影響：ランダム化クロスオーバー試験」<sup>34)</sup>。

**解説-**タイトルにデザインを明記する主な理由は、読者が試験デザインを特定しやすくするためである。ランダム化クロスオーバー試験であることを明記することで、読者がサンプルサイズや解析に関連して

表1 クロスオーバーランダム化試験を報告する際に含めるべき CONSORT チェックリストの情報

章 (Section/topic)	項目番号 (Item No.)	説明 (Description)
タイトル <sup>†</sup>	1a	タイトルにランダム化クロスオーバー試験であることを記載
抄録 <sup>†</sup>	1b	クロスオーバーデザインを明記し、表2にあるすべての情報を報告する
はじめに：		
背景 <sup>‡</sup>	2a	科学的背景と論拠の説明
目的 <sup>‡</sup>	2b	特定の目的または仮説
方法：		
試験デザイン <sup>†</sup>	3a	クロスオーバーデザインの根拠。そのデザインの特徴の記述、割付比率、特に介入期間の数や長さ、ウォッシュアウト期間の長さ、持ち越し効果への考慮を含む
プロトコルからの変更 <sup>‡</sup>	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
参加者 <sup>‡</sup>	4a	参加者の適格基準
設定と場所 <sup>‡</sup>	4b	データが収集された設定と場所
介入 <sup>†</sup>	5	再現可能となるような詳細な介入。実際にいつどのように実施されたかを含む
アウトカム <sup>‡</sup>	6a	事前に特定され明確に定義された主要・副次的アウトカム評価項目。いつどのように評価されたかを含む
アウトカムの変更 <sup>‡</sup>	6b	試験開始後のアウトカムの変更とその理由
サンプルサイズ <sup>†</sup>	7a	参加者内変動を考慮した上、どのようにサンプルサイズが決められたか
中間解析と中止基準 <sup>‡</sup>	7b	あてはまる場合には、中間解析と中止基準の説明
ランダム化：		
順序の生成 <sup>‡</sup>	8a	割付順序を作成した方法

順序の生成 <sup>‡</sup>	8b	ランダム化のタイプ:制限の詳細(ブロック化, ブロックサイズなど)
割振りの隠蔽機構 <sup>‡</sup>	9	ランダム割振り順番の実施に用いられた機構(番号付き容器など), 各群の割付けが終了するまで割振り順番が隠蔽されていたかどうかの記述
実施 <sup>†</sup>	10	誰が割付順序を作成したか, 誰が参加者を組入れたか, 誰が参加者を各介入の順序に割り当てたか
ブラインディング <sup>‡</sup>	11a	ブラインド化されていた場合, 介入に割付け後, 誰がどのようにブラインドかされていたか(参加者, 介入実施者, アウトカムの評価者など)
介入の類似性 <sup>‡</sup>	11b	関連する場合, 介入の類似性の記述
統計的手法 <sup>†</sup>	12a	クロスオーバーデザインに適切な(すなわち, 参加者内での比較に基づく)主要アウトカムおよび副次的アウトカムの群間比較に用いられる統計的手法
追加解析 <sup>‡</sup>	12b	サブグループ解析や調整解析のような追加的解析の手法
結果		
参加者の流れ(フローチャートを強く推奨) <sup>†</sup>	13a	ランダム割付けされた人数, 意図された治療を受けた人数, 主要アウトカムの解析に用いられた人数を順序と期間ごとに分けて記述
追跡不能と除外 <sup>†</sup>	13b	各順序と期間ごとに, 各段階で除外された参加者の数とその理由を記述
募集 <sup>‡</sup>	14a	参加者の募集期間と追跡期間を特定する日付
試験終了 <sup>‡</sup>	14b	試験が終了または中止した理由
ベースライン・データ <sup>†</sup>	15	ベースラインにおける人口統計学的, 臨床的な特性を順序・期間ごとに示す表
解析された人数 <sup>†</sup>	16	各解析における参加者数(分母), 解析が元の割付け群によるものであるか
アウトカムと推定 <sup>†</sup>	17a	主要・副次的アウトカムのそれぞれについて, 介入の推定効果量とその精度(95%信頼区間など)を含む結果は, 参加者内比較に基づくべきである。また, 各期間における各介入の結果が推奨される
二値アウトカム <sup>‡</sup>	17b	二値アウトカムについては, 絶対効果量と相対効果量の両方を記載することが推奨される
補助的解析 <sup>‡</sup>	18	サブグループ解析や調整解析を含む, 実施した他の解析の結果。事前に特定された解析と探索的解析を区別する
害 <sup>†</sup>	19	すべての重要な害または意図しない効果を, デザインに配慮した形で記述(詳細は害に関する CONSORT を参照)
考察:		
限界 <sup>†</sup>	20	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。持ち越し効果の可能性を考慮
一般化可能性 <sup>‡</sup>	21	試験結果の一般化可能性(外的妥当性, 適用性)
解釈 <sup>‡</sup>	22	結果の解釈, 有益性と有害性のバランス, 他の関連するエビデンス
その他:		
登録 <sup>‡</sup>	23	登録番号と試験登録名
プロトコル <sup>‡</sup>	24	可能であれば, 完全なプロトコルの入手方法
資金提供者 <sup>‡</sup>	25	資金提供者と他の支援者(薬剤の供給者など), 資金提供者の役割。

略語: CONSORT = Consolidated Standards of Reporting Trials

\*注: ジャーナルの要件に応じてページ番号はオプションになる。

<sup>†</sup>元の CONSORT から変更された項目。

<sup>‡</sup>変更されていない CONSORT 項目。

§ このランダム順序とは, 一般的にコンピュータプログラムによって生成されるランダムな順番リストのことである。これは, ランダム化クロスオーバー試験における介入の順序, 例えばある試験参加者が A → B という順番で介入を受けることと混同してはならない。

¶ 参加者内比較では, 各参加者が自分自身の対照として働くため, 各参加者の測定値間の相関を考慮する。したがって, 測定値は独立していない。

このデザインの意味合いを考え始めるようになる。

**項目 1b: 抄録 [Item 1b: Abstract]**

クロスオーバーデザインを明記し、表 2 にあるすべての情報を報告する。【Specify a crossover design and report all information outlined in table 2.】

**標準的な CONSORT 項目** - 試験デザイン, 方法, 結果, 結論の構造化抄録 (詳細は抄録に関する CONSORT を参照<sup>3)</sup>)。

**例**

**背景:** 冠動脈疾患 (CAD) の男性におけるシルデナフィルクエン酸塩の使用と報告されている心血管系の有害事象との関係は不明である。

**目的:** CAD の男性における運動中のシルデナフィルの心血管系への影響を評価する。

**デザイン, 設定, 被験者:** 2000 年 3 月から 10 月にかけて, 米国の外来診療紹介センターにおいて, 勃

起不全を有し, CAD であるまたは強く疑われる, 平均年齢 66 (± 9) 歳の男性 105 名 (最初にシルデナフィルを投与される 55 名と最初にプラセボを投与される 55 名) を対象に, ランダム化二重盲検プラセボ対照 2 期クロスオーバー試験が実施された。

**介入:** すべての患者は, 各運動負荷試験の 1 時間前にシルデナフィル (50 mg または 100 mg) またはプラセボを単回投与された後, 1~3 日の間隔をあけて症状制限のある仰臥位自転車負荷心エコー (symptom limited supine bicycle echocardiograms) を 2 回受けた。

**主なアウトカム評価指標:** 運動負荷心エコーで評価した運動中のシルデナフィルの血行動態への影響 (虚血の発症, 範囲, 重症度)。

**結果:** シルデナフィル使用後とプラセボ使用後の平均変化量の差は 4.3 (95%CI 0.9~7.7; P=0.01) であった。運動能力はシルデナフィル使用とプラセボ使用

表 2 ランダム化クロスオーバー試験の抄録に含めるべき情報: 抄録用 CONSORT チェックリストの拡張版 35)

項目 (Item)	説明 (Description)
タイトル *	ランダム化クロスオーバー試験であることを記載
試験デザイン *	試験デザイン (クロスオーバー試験と期間の数) の説明
方法:	
参加者 <sup>†</sup>	参加者の資格基準とデータが収集された設定
介入 *	参加者全員を対象とした介入
目的 <sup>†</sup>	具体的な目的または仮説
アウトカム <sup>†</sup>	明確に定義された本報告の主要アウトカム
ランダム化 *	参加者がどのように順序に割り付けられたか
ブラインディング (マスキング) *	参加者, 介護者, アウトカム評価者が介入についてブラインディングされているかどうか
結果:	
ランダム化された人数 *	各順序にランダムに割り付けられた参加者数
募集 <sup>†</sup>	試験状況
解析された数 *	解析された参加者数
アウトカム *	主要アウトカムについては, 参加者内比較に基づく推定効果量とその精度
害 <sup>†</sup>	重要な有害事象または副作用
結論 <sup>†</sup>	一般的な結果の解釈
試験登録 <sup>†</sup>	登録番号と試験登録名
資金提供者 <sup>†</sup>	資金源

略語: CONSORT = Consolidated Standards of Reporting Trials

\* 元の CONSORT から変更された項目。

<sup>†</sup> 変更されていない CONSORT 項目。

<sup>‡</sup> これは学会抄録にも適用される。



で同程度であった（平均差 0.07; 95%CI -0.06~0.19;  $P = 0.29$ ）。運動時の血圧と心拍数の増加も同様であった。呼吸困難または狭心症の発症は、シルデナフィル使用群 69 名、プラセボ使用群 70 名に認められた ( $P = 0.89$ )。運動負荷心電図は、シルデナフィル使用群 12 名 (11%)、プラセボ服用群 17 名 (16%) に陽性であった ( $P = 0.09$ )。運動誘発性壁運動異常は、シルデナフィル使用後とプラセボ使用後に同程度の症例数であった（それぞれ 84 名と 86 名;  $P = 0.53$ ）。最大運動時の壁運動スコア指数は、シルデナフィル使用後とプラセボ使用後に同程度であった（平均差 0.01; 95%CI -0.01~0.03;  $P = 0.40$ ）。

**結論：**安定した CAD 男性患者において、運動負荷心エコー検査で評価した結果、シルデナフィルは症状、運動持続時間、運動誘発性虚血の有無や程度に影響を及ぼさなかった。（Arruda-Olson らの論文より適応<sup>36)</sup>）

**解説-**明確で透明性があり、十分に詳細な抄録は重要である。読者は抄録にしかアクセスできない可能性があり、多くの読者は抄録の流し読みをしてから、さらに読むかどうかを決める。適切に書かれた抄録は、電子データベースから関連する報告を検索するのにも役立つ。2008 年に、ランダム化試験の抄録の報告に関する CONSORT 拡張版が発表され<sup>35)</sup>、その推奨事項は CONSORT 2010 に盛り込まれた<sup>3)</sup>。クロスオーバー RCT の抄録には、試験のデザイン、すなわち参加者内比較を考慮したランダム化順序、解析方法を示すべきである。表 2 は、クロスオーバー試験の抄録に含めるべき情報を示している。

必要なすべての項目に取り組んでいる優れた報告例を見つけることはできなかった。そのため、発表されている抄録を適応させた（例を参照）。

## 方法【Methods】

### 項目 3a: 試験デザイン【Item 3a: Trial design】

クロスオーバーデザインの根拠。そのデザインの特徴の記述、割付比率、特に介入期間の数や長さ、ウォッシュアウト期間の長さ、持ち越し効果への考慮を含む。【Rationale for a crossover design. Description of the design features including allocation ratio, especially the number and duration of periods,

duration of washout period, and consideration of carry over effect.】

**標準的な CONSORT 項目-**試験デザインの記述（並行群間、要因分析など）、割付比率を含む。

**例 1-**「この試験は、ランダム化二重盲検プラセボ対照クロスオーバーデザインで、期間は 15 か月であった…ランダム化 (1 か月)、治療期間 1 (6 か月)、ウォッシュアウト期間 (2 か月)、最後に治療期間 2 (6 か月) …患者は、治療期間 1 にアジスロマイシン、治療期間 2 にプラセボを投与される順序、または治療期間 1 にプラセボ、治療期間 2 にアジスロマイシンを投与される順序にランダムに割り付けられた<sup>37)</sup>」。

**例 2-**「この研究では、患者内変動が患者間変動よりも小さく、必要な患者数が少なくなるため、従来のランダム化並行群間デザインではなくクロスオーバーデザインが選択された。また、クロスオーバーデザインの既知の欠点のいくつか（例えば、脱落率が高い、患者病状の不安定性、持ち越し効果の可能性）は、この研究では予想されなかった<sup>38)</sup>」。

**例 3-**「各治療期間は、いずれの治療においても半減期の 5 倍以上にあたる 2 週間のウォッシュアウト期間で区切られており、これにより、次の治療を開始する前に薬剤が効果的に全身から排出されるようになっている<sup>39)</sup>」。

**例 4-**「患者の安全性を高めるため、治療の間に休薬期間を設けなかった。また、8 週間の治療期間は、第 2 期終了時の有効性測定前に、最初の治療効果のウォッシュアウトを行うのに十分であると考えた<sup>40)</sup>」。

**解説-**方法は、特定の設定においてクロスオーバーデザインを使用する根拠を含むべきである。特に、2x2 のクロスオーバーデザインでは、十分な検出力をもって持ち越し効果を同定できず、そのような効果を調整することもできないことから、持ち越し効果は無視できるものであると仮定し、その根拠を示す必要がある。デザインの説明では、治療期間、導入期間、ウォッシュアウト期間（ある場合）の長さ

に関する情報を含め、いくつかの介入がいくつかの期間にわたって検証されたかを明確にすべきである。

### 項目 3b: 方法の変更【Item 3b: Changes to methods】

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。【Important changes to methods after trial commencement (such as eligibility criteria), with reasons.】

標準的な CONSORT 項目から変更なし。

**解説**—持ち越し効果の検定は推奨されない。ただし、持ち越し効果の検定が実施され、その結果、著者が第1期のデータのみを使用している場合は、その旨を報告すべきである。この検定の使用について、項目 12a（統計学的手法）でも説明する必要がある。持ち越し効果が存在する理由についても説明する必要がある。

### 項目 5：介入【Item 5: Interventions】

再現可能となるような詳細な介入。実際にいつどのように実施されたかを含む。【The interventions with sufficient details to allow replication, including how and when they were actually administered.】

**標準的な CONSORT 項目**—再現可能となるような詳細な各群の介入。実際にいつどのように実施されたかを含む。

**解説**—この項目については、2×2 ランダム化クロスオーバー試験への拡張のために「各群の」が削除された。すなわち、すべての参加者が両方の介入を受けることが意図されている。

### 項目 7a：サンプルサイズ【Item 7a: Sample size】

参加者内変動を考慮した上、どのようにサンプルサイズが決められたか。【How sample size was determined, accounting for within participant variability.】

**標準的な CONSORT 項目**—どのようにサンプルサイズが決められたか。

**例**—「ケンブリッジ研究施設の先行研究（未発表データ）では、Apathy Evaluation Scale [AES] の平均スコアは 31 点（標準偏差 SD = 15.6）であった。AES-I の臨床的に有意な改善を平均スコアの 35% 低下と

定義すると、絶対効果量は  $0.35 \times 31$  点 = 10.85 点となる。したがって、サンプルサイズの推定には 10 ユニットという控えめの推定値が用いられる。

さらに、被験者内 SD = 15.0 が仮定される。各順序の群のサンプルサイズが 19（合計サンプルサイズ<sup>38)</sup> の場合、2×2 クロスオーバーデザインは、両側有意水準 0.050 の 2 群 t 検定（クロスオーバー ANOVA）を使用し、クロスオーバー ANOVA [分散分析]  $\sqrt{\text{MSE}}$  [平均二乗誤差] が 15.000（差の標準偏差 SD は 21.213）であると仮定すると、平均の差 10.000（治療 1 の平均値  $\mu_1$  の 31 と治療 2 の平均値  $\mu_2$  の 21 との差）の検出力が 80% になる。脱落の可能性を考慮し、40 名の患者がランダム化される。サンプルサイズの計算は nQuery 7.0 で行った<sup>41)</sup>。」

**解説**—クロスオーバーデザインの主な利点は、与えられた有意水準、検出力、効果量に対して、各参加者が 1 つの治療のみを受ける並行群間デザインと比べて、必要なサンプルサイズが小さくなることである。これは、各参加者が自分自身の対照として働き（各参加者は被験とコントロール介入を受ける）、参加者内の変動が取り除かれるためである。

試験の著者は、有意水準や検出力など、サンプルサイズの計算に必要な通常量を報告するだけでなく、連続変数の場合は例に示したように参加者内変動も報告する必要がある。サンプルサイズの計算に必要な参加者内情報を得ることは、しばしば困難である。発表するクロスオーバー試験の報告書では、サンプルサイズがどのように決められたかを明らかにすべきであり、理想的には参加者内変動の適切な推定値が使用されたことを示すべきである。連続変数のアウトカムがあるクロスオーバー試験の場合、サンプルサイズの推定に組み込む必要があるのは、参加者内差の予想標準偏差である。実際には、多くの試験でこの値の現実的な推定値を示唆する根拠データが存在していないが、この値を無視するとクロスオーバー試験のサンプルサイズが過大評価する可能性があるため、控えめである<sup>42)</sup>。参加者内差の標準偏差の推定を試みるべきである（または相関を考慮する）。

同様に、二値アウトカムの場合、データのペア性を考慮しないと、ペアデザインから生じる参加者内

比較を考慮することができず、サンプルサイズが不必要に大きくなる。著者は、サンプルサイズの計算が再現できるように、適切な詳細を示すことが求められている。

サンプルサイズ計算上の追跡不能に対する許容量も報告されるべきである。

#### 項目 8a：順序の生成【Item 8a: Sequence generation】

割付順序を作成した方法。【Method used to generate the random allocation sequence.】

標準的な CONSORT 項目から変更なし。

例 1- 「4 週間のプラセボ導入期間後、適格患者はコンピューター生成の割付スケジュールに従い、2 つの治療順序（モンテルカストとプラセボマッチングサルメテロール、またはサルメテロールとプラセボマッチングモンテルカスト）のいずれかにランダムに割り当てられた。2 週間のウォッシュアウト期間後、患者は他の治療に移行した<sup>43)</sup>。」

例 2- 「適格者は 2 つの治療順序（デノスマブ / アレンドロネートまたはアレンドロネート / デノスマブ）のいずれかに 1:1 でランダムに割り付け、それぞれの治療を 1 年間受けた<sup>44)</sup>。」

解説-クロスオーバー RCT では、割付順序は介入を受ける順番を指している。割付けは、参加者が A の後に B を受ける順序 1、または参加者が B の後に A を受ける順序 2 になる可能性がある。

#### 項目 10：実施【Item 10: Implementation】

誰が割付順序を作成したか、誰が参加者を組入れたか、誰が参加者を各介入の順序に割り当てたか。

【Who generated the random allocation sequence, who enrolled participants, and who assigned participants to the sequence of interventions.】

標準的な CONSORT 項目-誰が割付順序を作成したか、誰が参加者を組入れたか、誰が参加者を各介入に割り当てたか。

解説-この項目では、参加者は 1 つの介入ではなく、一連の介入にランダムに割り付けられるため、介入の後ろに「順序」を追記した。

#### 項目 12a：統計学的手法【Item 12a: Statistical methods】

クロスオーバーデザインに適切な（すなわち、参加者内での比較に基づく）主要アウトカムおよび副次的アウトカムの群間比較に用いられる統計的手法。【Statistical methods used to compare groups for primary and secondary outcomes which are appropriate for crossover design (that is, based on within participant comparison).】

標準的な CONSORT 項目-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。

例 1- 「健康関連 QOL スコアのクロスオーバー解析では、各順序内で各患者の治療間差を平均し、次に両順序間で平均し、治療効果の推定値を示した。推定治療差、95% 信頼区間および P 値は、分散分析モデルで期間と順序の影響に合わせて調整された」（強調追加）<sup>39)</sup>。

例 2- 「一般化線形混合モデル法を用いて、クロスオーバーデザインから生じる被験者内相関を考慮しながら、電気刺激期間と無刺激期間の差を推定した」（強調追加）<sup>45)</sup>。

例 3- 「統計解析により、ベースライン情報について両治療群を比較し、治療効果について 2 週間後と 4 週間後に比較することができた。治療に対する治験責任医師の評価と患者の評価は、治療順番 [厳密には期間] を考慮した二値回答の Gart 検定を用いて分析した」（強調追加）<sup>46)</sup>。

例 4- 「副作用と患者の嗜好は記述的に分析され、McNemar 検定を用いた」（強調追加）<sup>47)</sup>。

例 5- 「Prescott 検定は、時期効果が存在する場合の 2 つの治療法間の有意差検定を行い、主要エンドポイントを解析するために使用された」（強調追加）<sup>39)</sup>。

解説-国際医学雑誌編集者委員会（International Committee for Medical Journal Editors）と CONSORT グループによる推奨事項に従い、解析方法は「元のデータにアクセスできる知識のある読者が、報告された結果を検証できるように十分詳細に」記述すべきである（<http://www.icmje.org/recommendations/>

browse/manuscript-preparation/preparing-for-submission.html#d)。クロスオーバーデザインと使用された統計学的手法を明らかにすることで、読者は解析方法を評価することができる。

クロスオーバー試験の解析は、比較の「参加者内」という性質を尊重すべきである。「方法」のセクションには、どの解析方法を用いたかを明記すべきである。参加者内解析がどのように構築されたかを明確に示す必要があり、例えば、参加者内差に関する *t* 検定や、参加者、時期、介入効果を用いた分散分析などを使用することである。時期効果と持ち越し効果がモデル化されている場合は、これを報告すべきである。同様に、二値アウトカムの場合、条件付きロジスティック回帰は Mainland-Gart 検定を実施するための代替方法を提供している。参加者内比較を考慮しない解析結果は、治療効果の分散を過大評価する可能性がある。

一部のクロスオーバー試験では、参加者は両期間の開始時と終了時に結果変数について測定され、治療効果は各期間の変化スコアを用いて推定される。この直感的なアプローチは、持ち越し効果を排除すると主張されているが、治療効果の推定精度が低く、バイアスが生じる可能性もあるため<sup>48,49</sup>、推奨されていない。

クロスオーバー試験において、他のデザインと同様に欠測データによって一般的な問題を引き起こすが、具体的にはより複雑である。欠測データがない場合の解析モデルを特定し、ベースライン・データの役割を慎重に検討する必要がある。これは、ベースライン調整によって標準誤差が増加することが多いためである。利用可能なすべてのデータの混合モデル（例えば、この場合には、固定効果とランダム効果の混合）は、通常、被験者内依存性について状況に応じて適切な調整を行った最初のステップとして推奨され、Rubin の「ランダムな欠測」仮定の下で有効である。大まかに言うと、これは、治療順序と以前のデータが与えられた場合、後のアウトカムのデータの分布は、それらのデータが観察されているかどうかにかかわらず同じであることを示している。完全な記録を解析すると、ベースラインと治療順序が与えられた場合のアウトカムの分布は、それらが観察されるかどうか（すなわち、ランダムな欠測）に関係なく同じであると仮定することで、有

効な intention to treat の推定値が得られる。この検証不可能な仮定に対する結論の頑健性は、データを多重代入し、代入されたアウトカムの分布が、ベースラインと治療順序が与えられた時に観察されたアウトカムと異なるように強制することで調べることができる。多重代入、患者のサブセットからの代入（単一平均代入、最終値の繰り越し、または最良/最悪の代入以外）の使用は、代入されたデータが文脈的に妥当であり、変動性を適切に反映するため歓迎される<sup>50</sup>。

## 結果【Results】

### 項目 13a：参加者の流れ【Item 13a: Participant flow】

ランダム割付けされた人数、意図された治療を受けた人数、主要アウトカムの解析に用いられた人数を順序と期間ごとに分けて記述（フロー

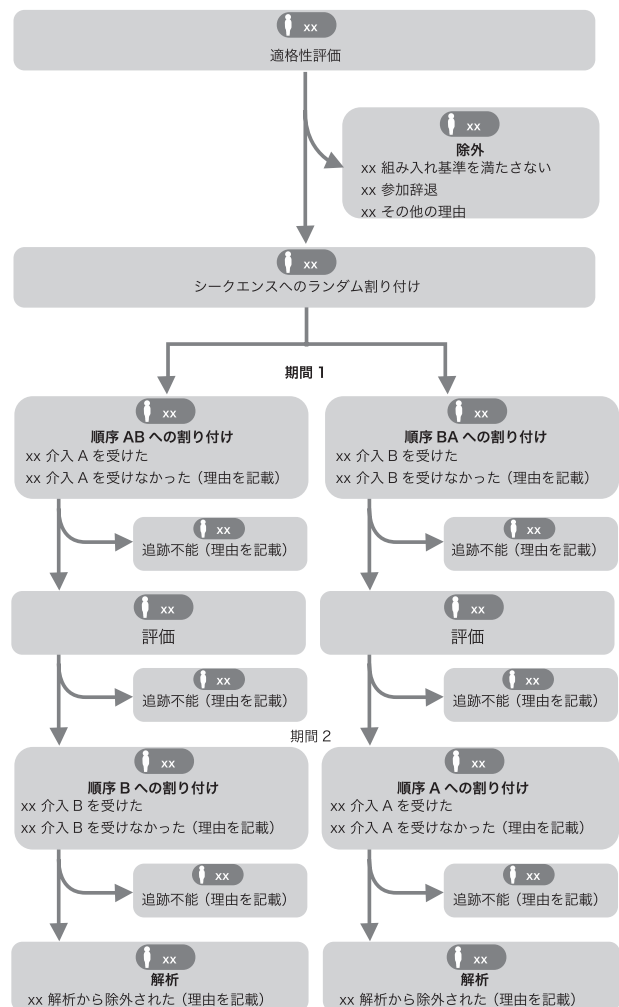


図1 クロスオーバー試験の CONSORT フローチャート（2つの群および/または期間を超える場合は修正が必要）。A= 介入、B= 対照またはその他の介入、CONSORT=Consolidated Standards of Reporting Trials。多くの場合、ウォッシュアウト期間を示すことができる。

チャートを強く推奨；図 1 を参照）。【The numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome, separately for each sequence and period (a flow diagram is strongly recommended; see fig 1).】

標準的な CONSORT 項目-各群について，ランダム割付けされた人数，意図された治療を受けた人数，主要アウトカムの解析に用いられた人数の記述。

例 1-図 2 を参照（Chen ら<sup>51)</sup>）。

例 2-図 3 を参照（Marchetti ら<sup>52)</sup>）。

解説-フローチャートは CONSORT 声明の重要な要素であり，広く採用されている。クロスオーバー試験では，各期間間の参加者の流れを理解することが

重要である。試験全体を通じて参加者の流れを説明するためにフローチャートを推奨しているが，正確な形式と内容は試験の特徴によって異なる場合がある。縦組みにし，タイムスケールを含めることを推奨する。

項目 13b：追跡不能と除外【Item 13b: Losses and exclusions】

各順序と期間ごとに，各段階で除外された参加者の数とその理由を記述。【Number of participants excluded at each stage, with reasons, separately for each sequence and period.】

標準的な CONSORT 項目-各群について，追跡不能例とランダム化後の除外例を理由とともに記述。

例 1-「最初にプラセボを投与する群に割り付けら

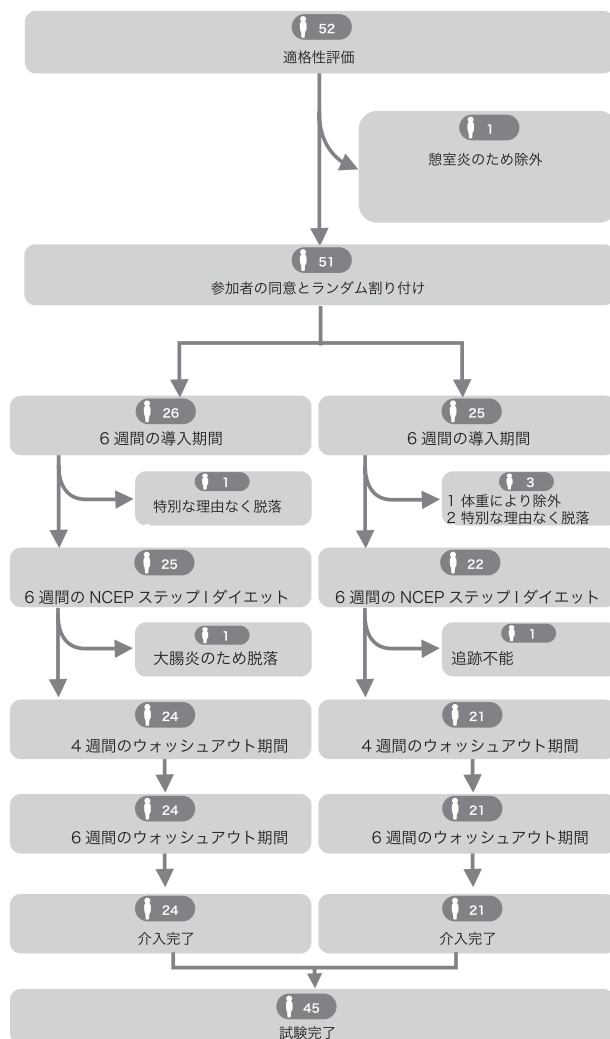


図 2 | フローチャートの例  
 NCEP = 国立コレステロール教育プログラム

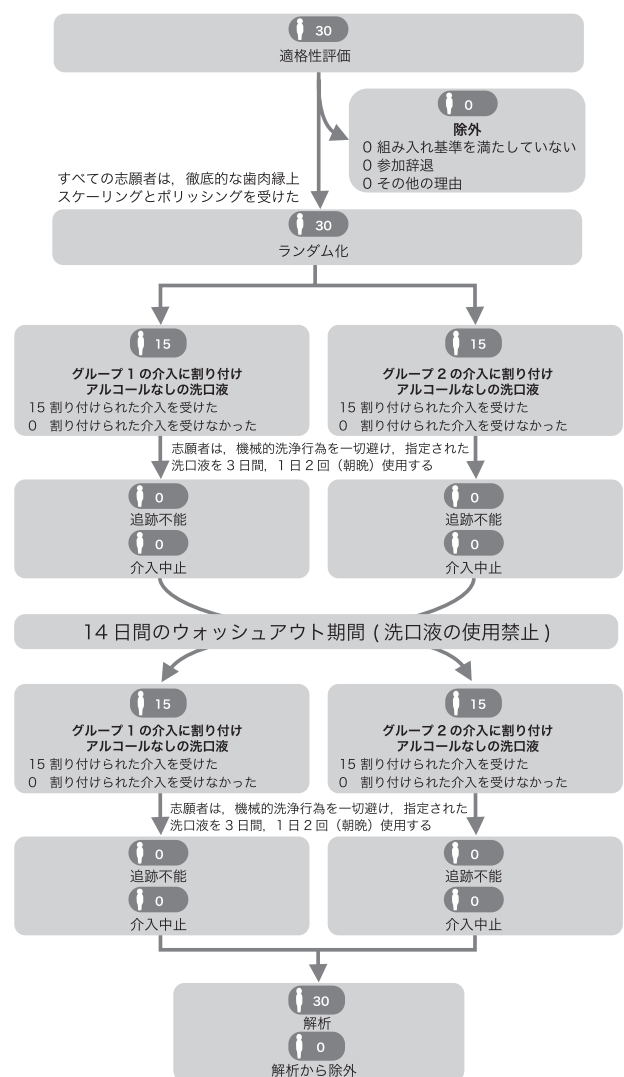


図 3 | フローチャートの例

れた被験者1名は、スケジュールの都合により、試験薬を投与する前に試験を辞退した。最初にプレガバリン投与群に割り付けられた被験者2名は、有害事象により、最初の期間に試験を辞退した。残りの26名は試験を完了した<sup>53)</sup>。」

**例2-**「同意を得た23名の患者のうち、17名がランダムに各治療順序に割り付けられた(9名がパンクレリパーゼ→プラセボ, 8名がプラセボ→パンクレリパーゼ)。16名の患者が試験を完了し、1名の患者(パンクレリパーゼ→プラセボの順序)は最初の治療期間の2日目に同意を取り下げた<sup>54)</sup>。」

**解説-**試験の途中で脱落した参加者は、1つの介入についてのみアウトカムが評価される。脱落は有益かもしれない。例えば、参加者は与えられた治療に不満があり、他の治療を試したくない可能性がある。

これにより結果にバイアスが生じる可能性がある。

著者は、各介入における参加者の追跡不能を、順序・期間ごとに分けて、可能であれば理由とともにフローチャート内で示すべきである。

不完全なデータを扱う統計学的手法がある(項目12aを参照)。

**項目15: ベースライン・データ【Item 15: Baseline data】**

ベースラインにおける人口統計学的、臨床的な特性を順序・期間ごとに示す表。【A table showing baseline demographic and clinical characteristics by sequence and period.】

**標準的な CONSORT 項目-**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。

**例1-**表3を参照(Fogelら<sup>43)</sup>)。

表3 ベースラインの人口統計学的、臨床的特性(順序別)

特性 (Characteristics)	Montelukast-salmeterol (n=78)	Salmeterol-montelukast (n=76)
性別:		
男性	43 (55.1)	30 (39.5)
女性	35 (44.9)	30 (39.5)
人種:		
アジア人	1 (1.3)	0 (0.0)
黒人	11 (4.1)	7 (9.2)
白人	38 (48.7)	41 (53.9)
その他	28 (35.9)	28 (36.8)
平均値 (SD) 年齢 (歳)	10.2 (2.0)	9.8 (2.0)
平均値 (SD) 運動前の FEV1 (L)	2.30 (1.1)	2.2 (0.6)
平均値 (SD) 運動前の FEV1 (% 予測)	96.3 (31.8)	92.8 (12.4)
平均値 (SD) 運動後の FEV1 の最大減少率	24.8 (10.3)	25.4 (9.0)
平均値 (SD) AUC0-20 分 (%-min)	320.1 (208.6)	317.7 (165.7)
平均値 * (SD) 回復までの時間 (min)	23.5 (10.5)	21.5 (8.3)
平均値 (SD) FEV1 の最大値 (% 予測)	99.9 (32.5)	100.5 (15.6)
平均値 (SD) SABA 初回使用後の FEV1 の平均変化率	1.4 (11.0)	4.8 (10.9)
チャレンジ後のレスキュー治療の必要性:		
なし	77 (98.7)	75 (98.7)
あり	1 (1.3)	1 (1.3)
喘息の増悪が通常の身体活動を制限する:		
全くない	2 (2.6)	4 (5.3)
少しある	21 (26.9)	20 (26.3)
程よくある	46 (59.0)	44 (57.9)
ひどくある	9 (11.5)	8 (10.5)

AUC0-20min = 運動後 20 分間の曲線下面積; FEV1 = 1 秒間の呼気量; SABA = 短時間作用型 β アゴニスト

特に明記されていない限り、データは数値 (%) である

\* ベースラインの FEV1 値の 5% 以内に戻った患者の数に基づく

表 4 ベースラインの人口統計学的，臨床的特性（順序別，合計別）

特性 (Characteristics)	100 IU kg <sup>-1</sup> 週 1 回→ 50 IU kg <sup>-1</sup> 週 2 回 (n=22)	50 IU kg <sup>-1</sup> 週 2 回→ 100 IU kg <sup>-1</sup> 週 1 回 (n=25)	合計 (n=50)*
平均値 (SD) 年齢 (歳)	31.7 (13.4)	25.1 (14.4)	27.7 (13.9)
男性	22 (100.0)	25 (100.0)	50 (100.0)
人種:			
白人	21 (95.5)	25 (100.0)	49 (98.0)
黒人	1 (4.5)	0	1 (2.0)
ヒスパニック系またはラテン系	5 (22.7)	2 (8.0)	7 (14.0)
非ヒスパニック系または非ラテン系	17 (77.3)	23 (92.0)	43 (86.0)
平均値 (SD) 体重 (kg)	72.3 (14.2)	64.6 (26.0)	69.2 (21.3)
対象関節 <sup>†</sup>	20 (90.9)	19 (76.0)	42 (84.0)
血友病性関節症 <sup>†</sup>	20 (90.9)	17 (68.0)	40 (80.0)
血友病性関節症による運動機能低下 <sup>†</sup>	18 (81.8)	14 (56.0)	34 (68.0)

SD = 標準偏差

特に明記されていない限り，データは数値 (%) である

\* 最初のオンデマンド期間に試験薬を投与されたが，ランダム化されなかった被験者 3 名を含む

<sup>†</sup> 試験開始時

例 2-表 4 を参照 (Valentino ら<sup>55)</sup>)。

解説-個人によるランダム割付けは，ベースライン時の群特性の差が何らかの系統的バイアスではなく偶然の結果であることを保証する<sup>2)</sup>。ランダム化クロスオーバー試験では，介入によって影響を受ける可能性のあるベースライン特性が，2 番目の期間の開始時に初期状態に戻っているかどうかを把握することが望ましい。順序別の情報は，ランダム化によって試験開始時に重要な変数の順序間バランスを達成したかどうかを評価するために必要である。期間別の情報は，次の期間における治療効果が，期間間で

変化する参加者の特性によって交絡されているかどうかを読者が理解するのに有用である。性別や年齢など，2 つの期間の開始時に同じである特性は，一度だけ示すことができる。しかし，不安定な予後因子と主なアウトカムのベースライン値は，各期間の開始時に確認する必要がある。もし特性が時間とともに変化する可能性がある場合，順序別のベースライン表では，期間（すなわち，介入）間の差に関する推論のみが排除される。

**項目 16：解析された人数【Item 16: Numbers analysed】**

各解析における参加者数（分母），解析が元の割

表 5 膵外分泌機能不全の治療と重症度による脂肪吸収係数の結果

変動 (Variable)	Pancrelipase (n=16)	Placebo (n=16)	Treatment difference (pancrelipase-placebo) (n=16) 治療差	P
CFA (%):				
LS 平均値 (SE)	82.8 (2.7)	47.4 (2.7)	35.4 (3.8)	<0.001
95% CI	77.0 to 88.6	41.6 to 53.2	27.2 to 43.6	-
EPI の重症度別の CFA (%):				
プラセボ CFA ≤ 50%	n=10	n=10	n=10	<0.001
LS 平均値 (SE)	81.8 (1.7)	37.3 (1.7)	44.5 (2.4)	-
95% CI	77.9 to 85.7	33.4 to 41.2	39.0 to 50.0	-
プラセボ CFA ≤ 50%	n=6	n=6	n=6	0.008
LS 平均値 (SE)	84.5 (2.9)	64.3 (2.9)	20.2 (4.1)	-
95% CI	76.5 to 92.5	55.3 to 72.3	8.9 to 31.6	-

95% CI = 95% 信頼区間 ; CFA = 脂肪吸収係数 ; EPI = 膵外分泌機能不全 ; LS, 最小二乗 ; SE = 標準誤差

付け群によるものであるか。【Number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups.】

**標準的な CONSORT 項目**-各群について、各解析における参加者数（分母）、解析が元の割付け群によるものであるか。

**解説**-結果の解釈には、試験解析に寄与する参加者の数が不可欠である。クロスオーバー試験の解析は、デザインのペア性質を考慮しなければならない。各アウトカムについて解析された人数は、参加者内の可能な差または対比の数と等しくなければならない。しかし、すべての参加者が各結果の解析に寄与するとは限らない。クロスオーバー試験では、ある期間で参加者が解析に寄与しなかった場合、対応する期間が失われる可能性がある。持ち越し効果や時期効果がないと仮定すると、代入が行われればデータを救済できるが、代入が行われない場合はデータが失われ、これは検出力の問題となる。サンプルサイズ、ひいては研究の検出力は、すべての参加者が情報を提供するという前提で計算されるため、統計的検出力の潜在的な低下を評価できるように、特定の解析に寄与する参加者の数を報告する必要がある。持ち越し効果や時期効果がある場合、欠損データはバイアスのかかった推定値につながる。また、CONSORT 2010 ガイドライン<sup>2)</sup>で詳細に説明されているように、per protocol解析が行われたのか、intention to treat解析が行われたのかを明記すべきである。

#### 項目 17a：アウトカムと推定【Item 17a: Outcomes and estimation】

主要・副次的アウトカムのそれぞれについて、介入の推定効果量とその精度（95%信頼区間など）を含む結果は、参加者内比較に基づくべきである。また、各期間における各介入の結果が推奨される。【For each primary and secondary outcome, results including estimated effect size and its precision (such as 95% confidence interval) should be based on within participant comparisons. In addition, results for each intervention in each period are recommended.】

**標準的な CONSORT 項目**-主要・副次的アウトカムのそれぞれについて、各群の結果、介入の推定効果

量とその精度（95%信頼区間など）。

例 1-表 5 を参照（Graff ら<sup>54)</sup>）。

例 2-表 6 を参照（RubioAurioles ら<sup>38)</sup>）。

例 3-「患者 80 名（70%）がパゾパニブを希望した。最も多い理由は、全体的に QOL の向上と疲労の軽減であった。患者 25 名（22%）がスニチニブを希望した。最も多い理由は、下痢が少ないことと全体的に QOL の向上であった。医師の好みは患者の好みと一致していた。パゾパニブの継続を希望する医師（61%）はスニチニブの継続を希望する医師（22%）より多く、残りの 17% は好みがないと回答した<sup>39)</sup>。」

例 4-表 7 を参照（O'Connor ら<sup>56)</sup>）。

**解説**-ランダム化クロスオーバー試験の結果を報告する場合、主要と副次的アウトカムについて、信頼区間付きの点推定値を報告すべきである。これは、これらの結果が適切な参加者内解析に基づく必要があることを除いて、標準的な CONSORT ガイドラインと同じである。結果は、並行群間比較試験の結果であるかのように示すことや、参加者を二重カウントしてはならない。相関が試験の検出力に影響するため、理想的には、将来のクロスオーバー試験の計画に役立つように、解析された各主要アウトカムの相関係数も提供すべきである。

二値アウトカムについては、読者が一致と不一致のペアを確認できるため、マッチ集計（matched tabulation）形式を用いた提示が望ましい。マッチ集計は、参加者内の相関を考慮することで（利用できない場合でも）、適切な数式を使用して治療間分散を下方修正できるため、将来的なメタアナリシスでこのような試験を使用することが容易になる<sup>57-59)</sup>。クロスオーバーデザインから得られた結果を並行群間比較試験形式で 2×2 表で表示すると、治療間分散の適切な調整ができない<sup>57)</sup>。ペア表示は、将来のサンプルサイズの計算にも役立つ。しかし、多くの場合、データはデザインを考慮したモデルで解析され、例 4 に示すように表示される。

各期間における各介入の結果を提示することが推



表6 副次的アウトカムの変化とベースラインから治療エンドポイントまでの変化

副次的アウトカム (Secondary outcome)	副次的アウトカムのベースラインからエンドポイントまでの変化* (Secondary outcome Changes between baseline and endpoint)						治療比較† (Treatment comparisons)			
	Sildenafil prn	Tadalafil OaD	Tadalafil prn	OaD-sildenafil prn	Tadalafil OaD-tadalafil prn	Tadalafil prn-sildenafil prn	OaD-sildenafil prn	Tadalafil prn	OaD-tadalafil prn	Tadalafil prn-sildenafil prn
SEAR スケール	25.40 (1.36) n=347	25.56 (1.36) n=348	26.92 (1.35) n=355	0.23 (1.11) [-1.95 to 2.42; P=0.834]	-1.47 (1.11) [-3.65 to 0.70; P=0.185]	1.71 (1.10) [-0.46 to 3.87; P=0.123]	0.23 (1.11) [-1.95 to 2.42; P=0.834]	-1.47 (1.11) [-3.65 to 0.70; P=0.185]	-1.47 (1.11) [-3.65 to 0.70; P=0.185]	1.71 (1.10) [-0.46 to 3.87; P=0.123]
性的関係	19.50 (1.31) n=347	19.40 (1.31) n=349	20.42 (1.30) n=355	-0.07 (1.07) [-2.17 to 2.04; P=0.951]	-1.12 (1.06) [-3.22 to 0.97; P=0.291]	1.06 (1.06) [-1.03 to 3.15; P=0.320]	-0.07 (1.07) [-2.17 to 2.04; P=0.951]	-1.12 (1.06) [-3.22 to 0.97; P=0.291]	-1.12 (1.06) [-3.22 to 0.97; P=0.291]	1.06 (1.06) [-1.03 to 3.15; P=0.320]
信頼度の合計	22.87 (1.29) n=347	22.94 (1.29) n=348	24.13 (1.29) n=355	0.11 (1.050) [-1.95 to 2.17; P=0.915]	-1.30 (1.040) [-3.35 to 0.74; P=0.212]	1.42 (1.04) [-0.63 to 3.46; P=0.174]	0.11 (1.050) [-1.95 to 2.17; P=0.915]	-1.30 (1.040) [-3.35 to 0.74; P=0.212]	-1.30 (1.040) [-3.35 to 0.74; P=0.212]	1.42 (1.04) [-0.63 to 3.46; P=0.174]
IIEF-EF ドメインスコア	9.70 (0.36) n=348	8.68 (0.36) n=350	9.54 (0.36) n=355	-0.85 (0.30) [-1.43 to -0.27; P=0.004]	-0.80 (0.29) [-1.37 to -0.22; P=0.007]	-0.05 (0.29) [-0.62 to 0.53; P=0.866]	-0.85 (0.30) [-1.43 to -0.27; P=0.004]	-0.80 (0.29) [-1.37 to -0.22; P=0.007]	-0.80 (0.29) [-1.37 to -0.22; P=0.007]	-0.05 (0.29) [-0.62 to 0.53; P=0.866]
EDITS スコア	75.68 (1.32) n=348	75.81 (1.31) n=351	79.50 (1.31) n=355	0.12 (1.28) [-2.40 to 2.64; P=0.926]	-3.55 (1.27) [-6.05 to -1.04; P=0.006]	3.66 (1.27) [1.16 to 6.17; P=0.004]	0.12 (1.28) [-2.40 to 2.64; P=0.926]	-3.55 (1.27) [-6.05 to -1.04; P=0.006]	-3.55 (1.27) [-6.05 to -1.04; P=0.006]	3.66 (1.27) [1.16 to 6.17; P=0.004]
朝の勃起頻度	0.11 (0.02) n=347	0.26 (0.02) n=352	0.20 (0.02) n=355	0.15 (0.01) [0.12 to 0.18; P<0.001]	0.06 (0.01) [0.03 to 0.09; P<0.001]	0.09 (0.01) [0.06 to 0.12; P<0.001]	0.15 (0.01) [0.12 to 0.18; P<0.001]	0.06 (0.01) [0.03 to 0.09; P<0.001]	0.06 (0.01) [0.03 to 0.09; P<0.001]	0.09 (0.01) [0.06 to 0.12; P<0.001]

EDITS = 勃起不全治療満足度調査票; IIEF-EF = 国際勃起機能指数 - 勃起機能領域; prn = 必要に応じて; OaD = 1日1回; SEAR = 自尊心と人間関係 \* 平均値 (標準誤差)

† 最小二乗平均 (標準誤差) [95% 信頼区間; P 値]

表7 年齢, 性別, 認知症の重症度を考慮した, 行動スコアおよび感情スコアに対する治療, 時間, 治療と時間の相互作用の効果の発生率比 (IRR) とその 95% 信頼区間

治療と時間 (Treatment and time)	Behaviour 行動		Affect 感情	
	IRR	P value	Positive 陽性	Negative 陰性
治療			IRR	IRR
ラベンダーとコントロールの比較	0.884 (0.778 to 1.004)	0.057	1.072 (0.848 to 1.355)	0.891 (0.504 to 1.573)
時間			P value	P value
曝露前と比較した最初の曝露 30 分後	0.899 (0.793 to 1.020)	0.097	0.393 (0.706 to 1.147)	0.96 (0.550 to 1.675)
曝露前と比較した 2 回目の曝露 30 分後	0.858 (0.755 to 0.974)	0.018	0.248 (0.67 to 1.106)	0.641 (0.348 to 1.179)
治療と時間の相互作用			IRR	IRR
ラベンダー × 最初の曝露 30 分後	0.961 (0.798 to 1.157)	0.672	1.02 (0.726 to 1.433)	0.848 (0.371 to 1.938)
ラベンダー × 2 回目の曝露 30 分後	1.045 (0.869 to 1.259)	0.636	0.79 (0.675 to 1.348)	0.687 (0.269 to 1.750)

表 8 有害事象

有害事象 (Adverse event)	有害事象なし (No of adverse events)	
嘔吐	NSAIDS, プラセボともに有害事象なし	108
	NSAIDS では有害事象が認められなかったが、プラセボでは有害事象が認められた	7
	NSAIDS では有害事象が認められたが、プラセボでは有害事象が認められなかった	13
	NSAIDS, プラセボともに有害事象が認められた	3

NSAID = 非ステロイド性抗炎症薬

奨められるのは、治験責任医師が解析でどのように結果を扱ったかに関係なく、治療と期間の交互作用を理解するのに役立つからである (Liらの表7を参照<sup>30</sup>)。

理想的には、参加者好みのアウトカムも参加者レベルで報告されるべきである。例えば、参加者を介入Aを好む人と介入Bを好む人に分け、McNemar検定、または期間を考慮する場合はMainland-Gart検定またはPrescott検定を用いて解析すべきである。

#### 項目 19：害【Item 19: Harms】

すべての重要な害または意図しない効果を、デザインに配慮した形で記述 (詳細は害に関するCONSORTを参照<sup>32</sup>)。【Describe all important harms or unintended effects in a way that accounts for the design (for specific guidance see CONSORT for harms<sup>32</sup>).】

**標準的な CONSORT 項目** - 各群のすべての重要な害または意図しない効果 (詳細は害に関するCONSORTを参照<sup>32</sup>)。

例-表8を参照 (架空の例)。

**解説**-各介入における有害事象の種類と全体の頻度を記載する必要がある。また、クロスオーバー試験では、一致と不一致の有害事象ペアを提示したり、(群間比較が行われた場合) 効果と精度の推定値を提示したりすることで、被験介入の相対的な安全性を把握することができる。表は、有害事象の集計方法の例を示している。

#### 考察【Discussion】

##### 項目 20：限界【Item 20: Limitations】

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。

持ち越し効果の可能性を考慮。【Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses. Consider potential carry over effects.】

**標準的な CONSORT 項目**-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。

例1-「24時間のウォッシュアウト期間は、刺激の影響を排除するには不十分であった可能性がある。持ち越し効果の可能性については、別の試験デザイン (例えば、並行群間、より長い試験/ウォッシュアウト期間、stepped-wedge デザイン) を用いて対処すべきである<sup>45</sup>。」

例2-「この試験の長所は、治療の盲検化とクロスオーバーデザインであり、患者は同じような健康状態で両方の治療を受けた。これにより、健康状態の違いに左右されない許容度の違いを検出でき、各患者が自分自身の対照として働くことができた。また、2週間のウォッシュアウト期間とランダム割付けにより、治療順序と持ち越し効果の影響を最小限に抑えることができた<sup>39</sup>。」

例3-「最後に、クロスオーバーデザインにより、HCQ [ヒドロキシクロロキン] の投与期間と非投与期間の差が不明瞭になった可能性がある。ウォッシュアウト期間を設けることは、そのような可能性を排除するのに役立つかもしれないが、パイロット試験では、そのようなウォッシュアウト期間は必要ないことが示唆された<sup>60</sup>。」

**解説**-クロスオーバーデザインの限界として、最初

の期間の治療が2番目の期間の結果に影響を及ぼす可能性があることである。これは、反対の治療でアウトカムを改善するか、またはその効果を抑制するかかのいずれかが考えられる。この持ち越し効果によりクロスオーバー試験が無効となる可能性があり、試験結果が無効となるため、このような限界の報告が見つかる可能性は低いと考えられる。報告する必要がある可能性のある限界には、2番目の介入が投与される前に追跡不能になること、複数の介入を混同して参加者がランダム化された順序とは異なる順序で介入を受けたことがある。試験期間中の疾患の安定性という観点から、クロスオーバーデザインの妥当性についても議論できる。

### より複雑な試験デザイン

#### 【More complicated trial designs】

前のセクションでは、各参加者が競合する2つの介入を受ける2つの順序のうち1つにランダムに割り付けられる単純な2x2試験デザインの報告について述べた。クロスオーバーデザインのより複雑なバリエーションには、3つ以上の介入を比較するもの（多群間試験に関するCONSORTの拡張版を参照<sup>61)</sup>）やクラスタークロスオーバーランダム化試験がある<sup>62)</sup>。クラスタークロスオーバーRCTでは、各クラスターはランダム化された順序で複数の介入を受ける。最近のレビューでは、適切な解析が行われるようにする必要があり、報告が改善される必要があることがわかった<sup>63)</sup>。CONSORTのクラスタークロスオーバー試験用拡張版の作成は進行中である（Joanne McKenzieによる私信）。

また、クロスオーバー試験では、反復測定（すなわち、複数の時点での測定）または参加者内の多重性（例えば、参加者内で両目が評価される）の問題が発生する可能性もある。他にあまり使用されないクロスオーバーデザインのバージョンに

は、生物学的同等性試験、Balaam デザイン、追加期間デザイン、n-of-1 デザイン、不完全ブロックデザインなどがある<sup>17)</sup>。

#### コメント【Comment】

RCTの報告には、読者が結果を正確に解釈できるように、方法と結果に関する重要な情報を含めるべきである。この情報は、そのような報告からデータを抽出しようとするメタアナリストにとって特に重要である。CONSORT 2010 声明は、RCTの報告書に含めるべき必須項目に関するCONSORTグループの最新の推奨事項を示している。本論文では、ランダム化クロスオーバー試験の報告に特化したCONSORTチェックリストの拡張版であり、それに対応する最新情報を紹介し、説明する。

2群の並行群間比較試験の報告にCONSORT声明を使用することで、報告の質が向上している<sup>64)</sup>。本CONSORT声明の拡張案を日常的に使用することで、最終的にクロスオーバーデザインの改善につながると考えている。ランダム化クロスオーバー試験を報告する際、著者は、主要なCONSORTガイドラインと本稿を併用し、CONSORTチェックリストの25項目すべてに対応すべきである<sup>3)</sup>。また、他の試験デザインに関するCONSORTの拡張版を参照することも有用である（[www.consortstatement.org/extensions](http://www.consortstatement.org/extensions) で入手可能）。

CONSORT声明は、研究者が将来、臨床試験を計画する際に役立ち、査読者や編集者が原稿を評価する際の指針にもなる。多くのジャーナルは、著者への指示の中で、CONSORTの推奨事項に従うことを推奨している。ジャーナルから、CONSORTの拡張版である本稿や、特定の試験デザインに関する他のCONSORTの拡張版を著者に案内することを推奨している。すべてのCONSORT推奨事項の最新版は、オンラインで入手できる（[www.consortstatement.org](http://www.consortstatement.org)）。

## 参考文献 [REFERENCES]

1. Page MJ, Higgins JP, Clayton G, Sterne JA, Hróbjartsson A, Savović J. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. *PLoS One* 2016; **11**:e0159267. doi:10.1371/journal.pone.0159267
2. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**:408-12. doi:10.1001/jama.1995.03520290060030
3. 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
4. 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
5. Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "Instructions to Authors" in 2014. *Trials* 2016; **17**:301. doi:10.1186/s13063-016-1408-z
6. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; **345**:e5661. doi:10.1136/bmj.e5661
7. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012; **308**:2594604. doi:10.1001/jama.2012.87802
8. Vohra S, Shamseer L, Sampson M, et al., CENT Group. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ* 2015; **350**:h1738. doi:10.1136/bmj.h1738
9. Zwarenstein M, Treweek S, Gagnier JJ, et al., CONSORT Group, Pragmatic Trials in Healthcare (Practihc) Group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008; **337**:a2390. doi:10.1136/bmj.a2390
10. Pandis N, Chung B, Scherer RW, Elbourne D, Altman DG. CONSORT 2010 statement: extension checklist for reporting within person randomised trials. *BMJ* 2017; **357**:j2835. doi:10.1136/bmj.j2835
11. Zeng, L., L.A. Drye, T. Li, Characterizing current registration of phase 3 crossover trials on ClinicalTrials.gov. 2015. <http://jhir.library.jhu.edu/handle/1774.2/38151>.
12. Hui D, Zhukovsky DS, Bruera E. Which treatment is better? Ascertaining patient preferences with crossover randomized controlled trials. *J Pain Symptom Manage* 2015; **49**:625-31. doi:10.1016/j.jpainsymman.2014.11.294
13. Cohlen BJ, te Velde ER, Looman CW, Eijckemans R, Habbema JD. Crossover or parallel design in infertility trials? The discussion continues. *Fertil Steril* 1998; **70**:40-5. doi:10.1016/S00150282(98)00114-9
14. Makubate B, Senn S. Planning and analysis of cross-over trials in infertility. *Stat Med* 2010; **29**:3203-10. doi:10.1002/sim.3981
15. Daya S. Differences between crossover and parallel study designs— debate? *Fertil Steril* 1999; **71**:771-3.
16. Jones B, Kenward MG. Design and Analysis of Cross-over Trials. 2nd edition ed. 2003, London: Chapman&Hall/CRC.
17. Senn S. Crossover-trials in Clinical Research. 2nd ed. Wiley, 2002. doi:10.1002/0470854596.
18. Wellek S, Blettner M. On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2012; **109**:276-81.
19. Gart JJ. An exact test for comparing matched proportions in cross-over designs. *Biometrika* 1969; **56**:75-80. doi:10.1093/biomet/56.1.75
20. Mainland D. Elementary Medical Statistics. W.B. Saunders, 1963.
21. Freeman PR. The performance of the two-stage analysis of twotreatment, two-period crossover trials. *Stat Med* 1989; **8**:1421-32. doi:10.1002/sim.4780081202
22. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005; **365**:1159-62. doi:10.1016/S0140-6736(05)71879-1
23. Mills EJ, Chan AW, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials* 2009; **10**:27. doi:10.1186/1745-6215-10-27
24. Yu LM, Chan AW, Hopewell S, Deeks JJ, Altman DG. Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review. *Trials* 2010; **11**:59. doi:10.1186/1745-6215-11-59
25. Inrig JK, Califf RM, Tasneem A, et al. The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov. *Am J Kidney Dis* 2014; **63**:771-80. doi:10.1053/j.ajkd.2013.10.043
26. 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
27. Gewandter JS, McDermott MP, McKeown A, et al. Reporting of cross-over clinical trials of analgesic treatments for chronic pain: Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks systematic review and recommendations. *Pain* 2016; **157**:2544-51. doi:10.1097/j.pain.0000000000000673
28. Straube S, Werny B, Friede T. A systematic review identifies shortcomings in the reporting of crossover trials in chronic painful conditions. *J Clin Epidemiol* 2015; **68**:1496-503. doi:10.1016/j.jclinepi.2015.04.006
29. Lathyris DN, Trikalinos TA, Ioannidis JP. Evidence from crossover trials: empirical evaluation and comparison against parallel arm trials. *Int J Epidemiol* 2007; **36**:422-30. doi:10.1093/ije/dym001
30. Li T, Yu T, Hawkins BS, Dickersin K. Design, analysis, and reporting of crossover trials for inclusion in a meta-analysis. *PLoS One* 2015; **10**:e0133023. doi:10.1371/journal.pone.0133023
31. Nolan SJ, Hambleton I, Dwan K. The use and reporting of the cross-over study design in clinical trials and systematic reviews: a systematic assessment. *PLoS One* 2016; **11**:e0159014. doi:10.1371/journal.pone.0159014
32. 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**:781-8. doi:10.7326/0003-4819-141-10-200411160-00009
33. Guo X, Kong X, Huang R, et al. Effect of Ginkgo biloba on visual field and contrast sensitivity in Chinese patients with normal tension glaucoma: a randomized, crossover clinical trial. *Invest Ophthalmol Vis Sci* 2014; **55**:110-6. doi:10.1167/iovs.13-13168
34. Johansson DP, Lee I, Risérus U, Langton M, Landberg R. Effects of unfermented and fermented whole grain rye crisp breads served as part of a standardized breakfast, on appetite and postprandial glucose and insulin responses: a randomized cross-over trial. *PLoS One* 2015; **10**:e0122241. doi:10.1371/journal.pone.0122241
35. Hopewell S, Clarke M, Moher D, et al, CONSORT Group. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008; **371**:281-3. doi:10.1016/S01406736(07)61835-2

36. Arruda-Olson AM, Mahoney DW, Nehra A, Leckel M, Pellikka PA. Cardiovascular effects of sildenafil during exercise in men with known or probable coronary artery disease: a randomized crossover trial. *JAMA* 2002; **287**:719-25. doi:10.1001/jama.287.6.719
37. Equi A, Balfour-Lynn IM, Bush A, Rosenthal M. Long term azithromycin in children with cystic fibrosis: a randomised, placebocontrolled crossover trial. *Lancet* 2002; **360**:978-84. doi:10.1016/S0140-6736(02)11081-6
38. Rubio-Aurioles E, Porst H, Kim ED, *et al.* A randomized open-label trial with a crossover comparison of sexual self-confidence and other treatment outcomes following tadalafil once a day vs. tadalafil or sildenafil on-demand in men with erectile dysfunction. *J Sex Med* 2012; **9**:1418-29. doi:10.1111/j.1743-6109.2012.02667.x
39. Escudier B, Porta C, Bono P, *et al.* Randomized, controlled, doubleblind, cross-over trial assessing treatment preference for pazopanib versus sunitinib in patients with metastatic renal cell carcinoma: PISCES Study. *J Clin Oncol* 2014; **32**:1412-8. doi:10.1200/JCO.2013.50.8267
40. Konstas AG, Lake S, Economou AI, Kaltsos K, Jenkins JN, Stewart WC. 24-Hour control with a latanoprost-timolol fixed combination vs timolol alone. *Arch Ophthalmol* 2006; **124**:1553-7. doi:10.1001/archophth.124.11.1553
41. Gelderblom H, Wüstenberg T, McLean T, *et al.* Bupropion for the treatment of apathy in Huntington's disease: a multicenter, randomised, double-blind, placebo-controlled, prospective crossover trial. *PLoS One* 2017; **12**:e0173872. doi:10.1371/journal.pone.0173872
42. Julious SA, Campbell MJ, Altman DG. Estimating sample sizes for continuous, binary, and ordinal outcomes in paired comparisons: practical hints. *J Biopharm Stat* 1999; **9**:241-51. doi:10.1081/BIP100101174
43. Fogel RB, Rosario N, Aristizabal G, *et al.* Effect of montelukast or salmeterol added to inhaled fluticasone on exercise-induced bronchoconstriction in children. *Ann Allergy Asthma Immunol* 2010; **104**:511-7. doi:10.1016/j.anai.2009.12.011
44. Freemantle N, Satram-Hoang S, Tang ET, *et al.* DAPS Investigators. Final results of the DAPS (Denosumab Adherence Preference Satisfaction) study: a 24-month, randomized, crossover comparison with alendronate in postmenopausal women. *Osteoporos Int* 2012; **23**:317-26. doi:10.1007/s00198-011-1780-1
45. Abell TL, Johnson WD, Kedar A, *et al.* A double-masked, randomized, placebo-controlled trial of temporary endoscopic mucosal gastric electrical stimulation for gastroparesis. *Gastrointest Endosc* 2011; **74**:496-503.e3.
46. Hill J, Bird HA, Fenn GC, Lee CE, Woodward M, Wright V. A double-blind crossover study to compare lysine acetyl salicylate (Aspergesic) with ibuprofen in the treatment of rheumatoid arthritis. *J Clin Pharm Ther* 1990; **15**:205-11. doi:10.1111/j.1365-2710.1990.tb00376.x
47. Bonten TN, Snoep JD, Assendelft WJ, *et al.* Time-dependent effects of aspirin on blood pressure and morning platelet reactivity: a randomized cross-over trial. *Hypertension* 2015; **65**:743-50. doi:10.1161/HYPERTENSIONAHA.114.04980
48. Fleiss JL, Wallenstein S, Rosenfeld R. Adjusting for baseline measurements in the two-period crossover study: a cautionary note. *Control Clin Trials* 1985; **6**:192-7. doi:10.1016/01972456(85)90002-9
49. Willan AR, Pater JL. Using baseline measurements in the twoperiod crossover clinical trial. *Control Clin Trials* 1986; **7**:282-9. doi:10.1016/0197-2456(86)90036-X
50. Carpenter JR, Kenward MG. Sensitivity Analysis with Multiple Imputation. In: Molenberghs G, *et al.* eds. Handbook of Missing Data Methodology. CRC Press New York, 2015: 446.
51. Chen CY, Holbrook M, Duess MA, *et al.* Effect of almond consumption on vascular function in patients with coronary artery disease: a randomized, controlled, cross-over trial. *Nutr J* 2015; **14**:61. doi:10.1186/s12937-015-0049-5
52. Marchetti E, Mummolo S, Di Mattia J, *et al.* Efficacy of essential oil mouthwash with and without alcohol: a 3-day plaque accumulation model. *Trials* 2011; **12**:262. doi:10.1186/1745-6215-12-262
53. Markman JD, Frazer ME, Rast SA, *et al.* Double-blind, randomized, controlled, crossover trial of pregabalin for neurogenic claudication. *Neurology* 2015; **84**:265-72. doi:10.1212/WNL.0000000000001168
54. Graff GR, Maguiness K, McNamara J, *et al.* Efficacy and tolerability of a new formulation of pancrelipase delayed-release capsules in children aged 7 to 11 years with exocrine pancreatic insufficiency and cystic fibrosis: a multicenter, randomized, double-blind, placebo-controlled, two-period crossover, superiority study. *Clin Ther* 2010; **32**:89-103. doi:10.1016/j.clinthera.2010.01.012
55. Valentino LA, Rusen L, Elezovic I, Smith LM, Korth-Bradley JM, Rendo P. Multicentre, randomized, open-label study of on-demand treatment with two prophylaxis regimens of recombinant coagulation factor IX in haemophilia B subjects. *Haemophilia* 2014; **20**:398-406. doi:10.1111/hae.12344
56. O'Connor DW, Eppingstall B, Taffe J, van der Ploeg ES. A randomized, controlled cross-over trial of dermally-applied lavender (*Lavandula angustifolia*) oil as a treatment of agitated behaviour in dementia. *BMC Complement Altern Med* 2013; **13**:315. doi:10.1186/14726882-13-315
57. Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2002; **31**:140-9. doi:10.1093/ije/31.1.140
58. Stedman MR, Curtin F, Elbourne DR, Kesselheim AS, Brookhart MA. Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2011; **40**:1732-4. doi:10.1093/ije/dyp345
59. Curtin F, Elbourne D, Altman DG. Meta-analysis combining parallel and cross-over clinical trials. II: Binary outcomes. *Stat Med* 2002; **21**:2145-59. doi:10.1002/sim.1206
60. Solomon DH, Garg R, Lu B, *et al.* Effect of hydroxychloroquine on insulin sensitivity and lipid parameters in rheumatoid arthritis patients without diabetes mellitus: a randomized, blinded crossover trial. *Arthritis Care Res (Hoboken)* 2014; **66**:1246-51. doi:10.1002/acr.22285
61. Juszczak E, Altman DG, Hopewell S, Schulz K. Reporting of MultiArm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA* 2019; **321**:1610-20. doi:10.1001/jama.2019.3087
62. Rietbergen C, Moerbeek M. The Design of Cluster Randomized Crossover Trials. *J Educ Behav Stat* 2011; **36**:472-90. doi:10.3102/1076998610379136.
63. Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE. Appropriate statistical methods were infrequently used in cluster-randomized crossover trials. *J Clin Epidemiol* 2016; **74**:40-50. doi:10.1016/j.jclinepi.2015.11.013
64. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev* 2012; **1**:60. doi:10.1186/20464053-1-60