

翻訳：COVID-19 パンデミック下での臨床試験実施に対する統計的課題と推奨  
日本製薬工業協会 データサイエンス部会 KT-1

原題：R. Daniel Meyer, Bohdana Ratitch, Marcel Wolbers, Olga Marchenko, Hui  
Quan, Daniel Li, Christine Fletcher, Xin Li, David Wright, Yue Shentu, Stefan Englert,  
Wei Shen, Jyotirmoy Dey, Thomas Liu, Ming Zhou, Norman Bohidar, Peng-Liang Zhao  
& Michael Hale (2020): Statistical Issues and Recommendations for Clinical Trials  
Conducted During the COVID-19 Pandemic, *Statistics in Biopharmaceutical Research*,  
DOI: 10.1080/19466315.2020.1779122

注意：本翻訳は、*Statistics in Biopharmaceutical Research* の出版元である Taylor and  
Francis より許諾を得ている (Permission Quote P061720-02/USBR)。

## ABSTRACT

COVID-19 パンデミックは、計画された、もしくは進行中の臨床試験に大きな影響を与えており、その影響は現在も継続している。パンデミックの試験データへの影響は、複合的かつ潜在的な統計学的問題を生み出している。影響の規模は前例のないものであるが、個々に見た場合、問題の多くは明確に定義することができ、対処することが可能である。Estimand、欠測データ、統計解析手法の妥当性と修正、追加解析の必要性、目的を達成する性能、試験全体の解釈可能性に関連する問題を評価し、対処するために、多くの戦略と推奨事項を提案する。

## 1. 序章

2019年12月に中国で発生したCOVID-19アウトブレイクは、2020年3月に世界保健機関(WHO)が宣言した通り、瞬く間に世界的なパンデミックとなった。今日現在、パンデミックからまだ数ヶ月しか経っていないが、この病気と公衆衛生対策は、世界的に臨床試験に非常に大きな影響を与えている。隔離、治験施設の制限、被験者や治験施設のスタッフに影響を与える渡航制限、被験者のCOVID-19感染、治験薬のサプライチェーンの寸断など、治験実施計画書の遵守の困難さを含む運営上の問題が発生している。治験依頼者はこの危機的状況において、被験者の安全を守ることを第一とした迅速な対応をしている。また、被験者の安全のために、いくつかの試験が中止され、登録が中断されている。進行中の試験については、治験依頼者は、被験者の安全を確保し、運営上の問題に対処するために、様々な緩和策を実施してきた。

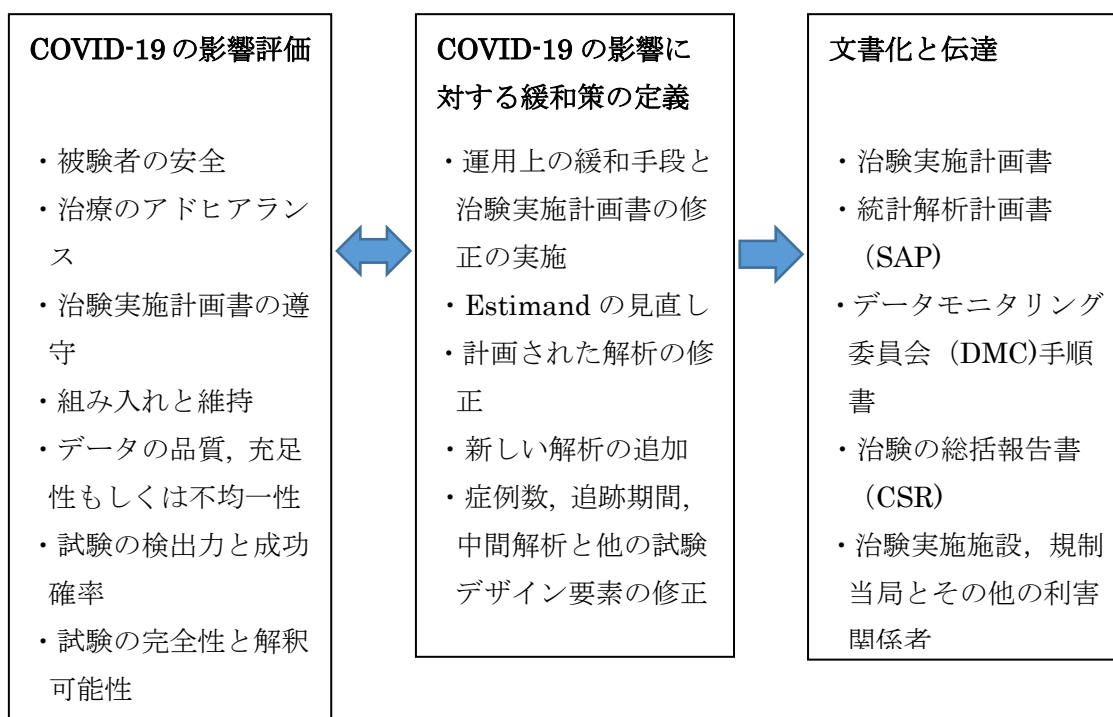
治験実施計画書からの逸脱と試験実施手順の変更は、後に続く臨床試験のデータへ様々な

影響をもたらす。この影響については後の節で詳しく説明するが、臨床試験にとって重要な統計的問題を提起している。極端な場合には、試験の完全性 (integrity), 解釈可能性, 又は試験の目的を達成する性能が低下する可能性がある。一般的には、試験結果の綿密かつ適切な解釈を提供するために、計画された統計解析を修正したり、補足したりする必要があるかもしれない。本論文では、試験の目的、推論、統計解析に関連する問題に対処するための様々な推奨事項を提供する。影響の主な種類と緩和策を図1にまとめた。

ここで議論する問題の多くは、パンデミック前に開始された、COVID-19に関連しない治療に対する進行中の臨床試験に関連するものである。問題点及び推奨事項の多くは、新しい試験にも適用可能であろう。規制当局は、COVID-19 のパンデミックにより起きている問題に対処するための治験依頼者向けのガイダンスを迅速に発表している (FDA 2020;EMA, 2020a, 2020b)。本論文は、これらの重要なガイダンス文書の影響を受けており、また、それらを拡張した内容となる。

本論文は以下のように構成されている。2節では、臨床試験に対する全体的な影響評価について述べる。3節では、estimand のフレームワークを用いた試験への影響評価を行う。4節では、データの欠測のメカニズムや欠測に対処するための推奨される統計的アプローチなど、試験に必要とされる可能性のある解析方法の修正や補足についての推奨事項をまとめている。5節では、追加的に考慮すべき事項を概説している。6節では推奨事項を要約する。

図 1. COVID-19 の影響に対処するための評価, 緩和策, 及び文書化の主要な側面



## 2. パンデミック関連の要因，影響，リスク評価

臨床試験での統計的な問題につながり得るパンデミックに関わる影響の例を表1に示す。影響の中には，パンデミック対策やCOVID-19によって直接的に引き起こされるものもあれば，試験実施への直接的な影響に対処するために適用された運用上の修正に起因するものもある。パンデミックは，被験者がいつ試験に参加するかによって，個々の被験者に異なる影響を与える可能性がある。被験者に対するパンデミックの影響をまとめる方法は，4.3.3節で議論されている。

表 1. COVID-19 関連の要因と臨床試験への潜在的な影響の例

| 要因   | 影響／リスクの例   |
|--|--|
| 隔離，旅行制限，被験者の個人的なパンデミック関連の理由により施設を訪問できない，訪問を望まない，施設の閉鎖または施設スタッフの可用性の低下のために施設に訪問できない | <ul style="list-style-type: none"> <li>• 来院及び評価の未実施又は遅延</li> <li>• 試験治療へのアクセス障害</li> <li>• 試験からの脱落</li> <li>• クエリに対する応答の遅延</li> <li>• 治験担当医師の変更/測定方法の変更</li> <li>• 施設モニタリング活動の遅延</li> <li>• 患者登録の遅延</li> </ul> |
| 試験治療や他の薬剤のサプライチェーンの寸断  | <ul style="list-style-type: none"> <li>• 治験薬の投与ができない</li> <li>• COVID-19に関連しない併用薬の変更</li> </ul>  |
| 薬物の投与方法の変更   | <ul style="list-style-type: none"> <li>• 投与エラーのリスクの増加</li> <li>• 投与方法の同等性の欠如</li> </ul>  |
| 代わりの臨床検体の収集  | <ul style="list-style-type: none"> <li>• 照合と検証の課題</li> </ul>   |
| 代わりのデータの収集   | <ul style="list-style-type: none"> <li>• 測定方法の代替可能性の欠如</li> </ul>  |
| COVID-19感染／関連する治療  | <ul style="list-style-type: none"> <li>• 試験治療又は被験者の一時的/永久的な中断</li> <li>• 有効性評価項目／安全性への潜在的な影響</li> <li>• COVID-19に関連する併用薬と治験薬との相互作用</li> </ul>  |

パンデミックの影響を受けた試験については，被験者に対するベネフィット／リスクの変化を評価することが，意思決定プロセスの最初のステップである（FDA, 2020）。本論文のすべての推奨事項は，被験者の安全を確保するために適切な措置がとられていることを前提としている。

## 2.1. 全体的なリスク評価

治験依頼者は、臨床試験から解釈可能な結果が得られる可能性を評価するために、集約され、かつ盲検化されたデータに基づいてリスク評価を行うことが推奨される。この評価は、まだ見られていないが、起こり得そうな効果を予想して、先を見越して開始されなければならない。感染状況やパンデミック対策の地域差を考慮しながら、状況の変化や蓄積されたデータを考慮して、試験の実施中も継続しなければならない。特に有効性と安全性の主要な評価指標については、運用上のバイアスや統計上のバイアスを最小化するためのあらゆる手段を講じる必要がある。重要となる項目は：

- 被験者の安全を守り、試験の完全性に対するリスクを最小化するために、試験の修正を行うべきかどうかを判断する。
- COVID-19の影響を受けるすべての被験者を特定し、どのように影響を受けているか（例：治療中断、試験中止、未来院）、試験への影響がどのようなものを把握する。
- パンデミック関連の問題を適切に監視し、文書化し、対処するために、どのような追加情報をデータベースに収集する必要があるか、又は治験責任医師からの情報提供の形で収集する必要があるかを決定する（そのような情報を入手するための実現可能性及びその質はさまざまに異なる可能性があり、リスク要因の一部として考慮する必要がある）。
- 治療又は試験中止の理由及び計画された **estimand** と中間事象への影響を理解する。
- データの欠測の程度及び欠測の具体的な理由を評価する。
- 登録者数及び対象集団の経時的変化を評価する。
- 治験実施計画書で設定した仮定と試験が目標を達成できる可能性を評価する。
- 理想的には、代替法（例えば、音声電話やビデオ電話による来院）から得られたデータの有用性を、代替法を実施する前に検証する。そのようなデータは、より多くの変動を生じさせる、又は解釈できない可能性がある。
- 盲検を解除する前に、事前に規定する必要がある計画された解析及び解析対象集団の定義の変更、又は追加の感度分析を決定する。

上記のリスク評価に基づき、多くの治験依頼者は、リスクのモニタリングと評価に対する一貫したアプローチを強化するために、未来院率、治療や試験の中止、治験実施計画書からの逸脱、有害事象 (AE) など、試験の運営状況に関する標準的な指標を開発している。このような指標は、より急速に影響を受けている試験を特定し、試験ポートフォリオの全体的な状態を監視するのに有用である。

### 3. Estimands に対する影響と緩和策

ICH E9(R1) 補遺 (ICH, 2019) は、試験の計画、実施、解析及び解釈のすべての側面が試験の目的と一致することを確実にするための estimand のフレームワークを定義している。また、潜在的なパンデミックと関連する混乱を議論し、それを試験目的と設計要素の状況に当てはめるための厳格な基礎を提供する。

影響を受けた試験では、最初の主要な論点は、主要な目的、すなわち、主要な estimand が COVID-19 の潜在的な交絡の影響を除いた治療効果を対象とすべきかどうかということである。我々は、パンデミック前に開始されたほとんどの試験については、当初の主要な目的を計画通りに維持すること、すなわちパンデミック関連の混乱によって交絡されない治療効果を対象とすることを推奨する（新しい試験では、この治療効果の定義も妥当であろうが、試験計画の多くの側面に依存する）。これは自動的に、すべての可能性のあるパンデミックと関連する中間事象 (ICE) について同じ仮想シナリオを持つ広範な "hypothetical estimand" を意味するものではない。交絡因子は、試験の状況に応じ、異なるタイプの中間事象に対して異なる方法で対処する必要があるかもしれない。これについては 3.1 節で議論する。したがって、パンデミックと関連する混乱を考慮するために、当初の estimand の定義に若干の変更が必要となるかもしれない。また、治験依頼者は COVID-19 パンデミック中に出現した特定の状態又は部分集団における治療効果を特徴づけるために、探索的な目的で補足的 estimand を検討してもよい。

当初の治験実施計画書で estimand が明示的に定義されている試験では、治験依頼者はパンデミックと関連する混乱が estimand の要素に及ぼす影響を検討し、必要に応じて修正すべきである。必要な修正は、計画通りに試験登録を継続するか、一時的に登録を停止するかなど、リスク軽減のシナリオに応じて異なる。Estimand が明示的に定義されていない試験の場合、本節はパンデミックと関連する影響が試験目的、実施、推測及び統計解析に与える影響について考える際の指針として有用である。

以下では、estimand の定義の各要素をレビューし、パンデミックと関連する重要な考慮事項を強調する。ICH E9(R1) ガイドラインを踏まえて、中間事象と欠測データの取り扱いを区別する。中間事象は、estimand を定義するために使用される 5 つの要素の 1 つとして、本節で議論される。試験治療が計画通りに継続されている間に収集されていれば、与えられた estimand に対して意味があったであろうと考えられる評価の欠測は、中間事象として考慮されるべきではない。このような状況は欠測データにつながり、estimand との整合性はあるものの、推定量の側面では欠測データ戦略の一部として考慮されるべきである (4.2 節参照)。

本論では主要な有効性 estimand の検討に焦点を当てているが、他の試験の estimand にも同様のロジックを適用することができる。

### 3.1. 中間事象の取り扱い

パンデミックと関連しない中間事象に対する戦略は変わらないはずである。ここでは、パンデミックと関連する中間事象の取り扱いについてのみ議論する。Estimandのフレームワークは、異なるタイプの中間事象に対して異なる戦略を使用することを可能にしており、そのような estimand が現在の状況では最も適切である可能性が高い。

中間事象のうち、パンデミックと関連する要因の結果として発生し、他のパンデミックと関連しない理由（例えば、有効性の欠如や毒性による治療中止など）に起因しない場合には、パンデミックと関連する中間事象とみなすべきである。パンデミックと関連する重要な中間事象は、まず、試験治療のアドヒアランス（試験治療の中止など）や目標とする結果（死亡など）を確認する能力への影響の観点から分類すべきである。次に、これらの中間事象を取り扱うための適切な戦略を選択するために、中間事象をパンデミックと関連する要因に応じてさらに分類する必要がある。関連する因子は、試験治療の利用可能性、被験者の COVID-19 の感染状態、COVID-19 に感染した場合の併用治療である。パンデミックと関連する各中間事象を説明する重要な要素を表 2 にまとめている。例えば、試験治療の早期中止（表 2 の 1 行目）は、複数のパンデミックと関連する要因（表 2 の 2～4 列目）が原因であり、有効性の欠如や試験治療の毒性などの他の要因に起因しない場合には、パンデミックと関連すると考えられる（このリストは網羅的なものではなく、試験によっては追加のカテゴリーを考慮する必要があるかもしれない）。要因が組み合わせられることによって生じる中間事象の種類は潜在的に多いが、全て又は大部分の中間事象を同様の方法で扱うことができるであろう。

表2.パンデミックと関連する中間事象の要素

|                              |  |
|------------------------------|--|
| <p>被験者の試験治療へのアドヒアランス</p>     | <ul style="list-style-type: none"> <li>• 試験治療が中止され、試験の対象疾患に対する新たな治療は開始されていない</li> <li>• 試験治療が中止され、試験の対象疾患に対する代替療法に切り替え</li> <li>• 試験治療が一時的に中断又は遵守状況が著しく低下、かつ試験の対象疾患に対する併用治療の変更がない</li> <li>• 試験治療が一時的に中断又は遵守状況が著しく低下、かつ試験の対象疾患に対する併用治療の変更あり（例：レスキュー治療の開始）</li> </ul> |
| <p>試験治療へのアクセス</p>            | <ul style="list-style-type: none"> <li>• 試験／地域ごとの医薬品供給の中断</li> <li>• 施設において試験治療の投与／調剤ができない</li> <li>• 施設において試験治療は利用可能だが、パンデミックと関連する被験者の個人的な理由により試験治療を受けることができない／受けたくない</li> </ul>  |
| <p>被験者のCOVID-19の感染状態</p>     | <ul style="list-style-type: none"> <li>• COVID-19陽性かつ生存している被験者</li> <li>• COVID-19*により死亡した被験者</li> <li>• COVID-19の感染の疑いがある被験者（症状の有無によってさらに区別される可能性がある）</li> <li>• 医療へのアクセスの低下による基礎疾患の悪化を伴う被験者</li> </ul>   |
| <p>被験者のCOVID-19に対する併用治療*</p> | <ul style="list-style-type: none"> <li>• COVID-19に対する治療（薬理的治療、酸素治療など）を受けた被験者</li> <li>• 集中治療室以外での入院</li> <li>• 集中治療室での入院</li> </ul>  |

\* COVID-19 と関連する死亡及び COVID-19 の感染に対する治療の開始は、試験治療の完了後、又は他のパンデミックと関連しない中間事象の発現後であって、関心のある評価項目と関連する時点の前に発生した場合にも、中間事象として考慮される場合がある。

試験治療のアドヒアランスがない特定のタイプは、通常は中間事象として考慮されないかもしれないが、パンデミックの状況下では中間事象として考慮する必要があるかもしれない。例えば、著しい遵守状況の低下や一時的な治療中断は、試験計画時には想定されていなかったかもしれないが、これらがパンデミックと関連する混乱に起因する可能性が高いと考えられる場合には、中間事象となりうる。パンデミックと関連する重大な治療の中断を伴う試験では、治療効果が薄められると予想される中断の最小期間を定義することもできるであろう。試験治療の短い期間の中断ではなく、この最小の期間を超える試験治療の中断に対しては、異なるストラテジーを用いることができる。カットオフの選択に対する推測の安定性を評価するために感度分析を使用できる。イベントまでの時間を評価項目

とする場合は、最初の治験薬の投与からイベント又は打ち切りまでの治験薬への曝露状況に応じて、被験者単位での服薬遵守の最小許容レベルを定義し、服薬遵守の最小許容レベルに達しない被験者を除外したいと考えるかもしれない。しかし、このような手法は、不死期間バイアス (immortal time bias) を生じる可能性があるため、避けるべきである (van Walraven et al., 2004)。

COVID-19 に対する試験的治療を受けている被験者については、試験治療を継続するかどうかにかかわらず、特別な配慮が必要であろう。また、死亡の発生が当初予想されていなかった試験では、COVID-19 による死亡は潜在的な中間事象として考慮されるべきである。

パンデミックと関連しない理由に起因する、表 2 の「被験者の試験治療のアドヒアランス」の要素に対して挙げた中間事象 (例：試験治療の中止) の種類の多くは、パンデミック発生以前の主要な estimand において対処されているであろう。パンデミックが原因でこれらの中間事象が発生した場合においても、当初のストラテジーが正当であるかどうかを検討することから始めることを推奨する。当初のストラテジーが正当化できない場合は、別のストラテジーを選択すべきである。この点について、いくつかのハイレベルな検討事項を概説する。

- 治療効果を定義する際に、当該中間事象を問題としないものとする治療方針ストラテジーは、パンデミックが存在しない場合には結論の一般化ができないため、ほとんどのパンデミックと関連する中間事象について、治療方針ストラテジーに対しては一般的に科学的な関心がないであろう。例えば、治療方針ストラテジーの下で推定される治療効果は、パンデミックと関連する混乱に起因する早期の治療中止に対して、パンデミックと関連する要因 (治験薬供給の中断など) に起因する治療の中止や変更が生じるレジメンの影響を反映しており、主要な試験目的に沿わないと考えられる。パンデミックと関連しない理由に起因し、治療方針ストラテジーによる対応が計画されていた中間事象の発現後に COVID-19 に対する治療が開始され、有効性・安全性の評価項目が観察される前である場合、COVID-19 に対する治療開始について慎重に検討する必要がある。当該事象を問題としないと単純にみなすことはできない。治療方針ストラテジーの下では、推定される治療効果は感染症及びそれに対する治療効果を反映している可能性があり、これは主要な目的における関心事ではないであろう。

パンデミックと関連する中間事象に対して治療方針アプローチを使用することの決定は、そのような事象が発現した被験者の割合が低く、パンデミックと関連しない同様の中間事象に対してこのストラテジーが計画されていた場合には、正当化される可能性がある。また、このストラテジーは、比較的短期間の治療中断に対応する中間事象の取り扱いにおいても検討されるであろう。非劣性試験及び同等性試験



における、パンデミックと関連する早期の試験治療中止の中間事象については、そのような事象数の増加に応じて治療群間の類似性が人為的に増す可能性があるため、治療方針ストラテジーの使用は回避されるべきである。複合ストラテジーについても同様の考えが適用される。

- 複合ストラテジーでは、結果変数の定義に中間事象が組み込まれているが、パンデミックと関連する多くの中間事象には適切ではないと思われる。例えば、パンデミックと関連する混乱による試験治療の中止は、有効性の欠如や副作用による中止と同様として治療の失敗としてカウントされるべきではない。呼吸器疾患を対象とした試験では、COVID-19の合併症が好ましくない転帰の一形態として複合ストラテジーで考慮される可能性があるため、より繊細な検討が必要となるかもしれない（後述のCOVID-19と関連する死亡についての議論も参照）。
- COVID-19と関連する事象（例えば、COVID-19による重篤な合併症又は死亡）で層別化する主要層ストラテジーは、将来の実臨床での状況と関連しない要因に基づいて定義された被験者の部分集団に結論を限定することになるため、主要な **estimand** として関心を持たれる可能性は低いと思われる。
- パンデミックと関連しない中間事象に対して当初から計画されている場合、特定の条件下では、治療下ストラテジーが引き続き適切な場合がある。

このストラテジーは、治療効果の確立に治療期間が関係ない場合（例えば、緩和ケアにおける疼痛の治療）には一般的に正当化できるが、試験治療の結果を確実に測定するために必要な最小の治療期間など、特定の条件を考慮する必要があるかもしれない。

- 中間事象が発生しなかった場合の治療効果に関心を持つ仮想ストラテジーは、ほとんどのパンデミックと関連する中間事象に対して自然な選択となる。これは、特にパンデミックを理由として試験治療が中断された場合の中間事象に適用される。このような被験者については、治療へのアクセスが中断されていない類似の被験者と同様の方法で試験を継続したという仮想的な状況を考えることが合理的である。これらの被験者が試験終了まで試験治療を完全に遵守したという仮想的な状況を想定する必要はない。むしろ、仮想的な状況としては、試験治療を遵守する症例と、パンデミックと関連しない理由で試験治療を遵守しない症例が混在している状況を含むであろう。仮想的な状況の詳細について合意を得るために、試験の最終段階で割付を明らかにする前に規制当局との協議を実施することは有用であろう。

推定方法は **estimand** の検討には含まれないが、利用可能なデータに基づいて、仮想ストラテジーの下で安定した方法で治療効果を推定できることは当然と考えるべきではなく、**estimand** の定義を最終化する際に推定手法の性能を評価すべきである。この側面は、全体的なリスクの評価と軽減戦略の選択に関する意思決定の一

部とすべきである。

上記の議論では、表 2 に挙げたようなパンデミックと関連する要因に関連する情報を把握する必要性が強調されている。これは、症例報告書（CRF）の指定された入力欄を介して、又は治験実施計画書からの逸脱を詳細かつ構造化された形式で収集することによって行うことができる。

COVID-19 による死亡の中間事象については慎重に検討する必要がある、適切なストラテジーは試験の対象疾患や臨床評価項目によって異なる。死亡が評価項目の構成要素とならない死亡率の低い疾患領域では、COVID-19 の感染に関連する死亡に対する仮想ストラテジーが推奨される可能性がある。死亡が評価項目の一部である、より重篤な疾患を対象とした試験では、規制の観点から、治療の有益性を評価する際に 2 つ以上の *estimand* に対して関心を持つことは避けられない。COVID-19 と関連する死亡数が少ない場合や、パンデミックの影響を治療効果の推定値に反映させたい場合には、結果に COVID-19 と関連する死亡を含める実用的な方法、すなわち複合ストラテジーを用いる方法が適している（4.2.2 節の競合リスク解析の項も参照）。COVID-19 の感染に関連する死亡に対して仮想ストラテジーを利用することは、COVID-19 が存在しない状況（例えば、当該疾患が根絶された場合や有効な治療法が出現した場合など）での治療の有益性を評価する上で重要である。このような試験には高齢、虚弱、又は免疫不全の被験者がしばしば含まれており、死亡が COVID-19 によって引き起こされたものであるか、又は被験者が COVID-19 に感染した状態で死亡したかを判断することが困難となる可能性があることが認識されている。

治療下ストラテジー、複合ストラテジー、及び主要層ストラテジーは主要な *estimand* としては関心がないかもしれないが、COVID-19 の感染と転帰に基づいて層別化された被験者の部分集団、又は COVID-19 に対して投与される併用治療について検討する科学的根拠がある場合には、補足的な *estimand* として関心を持たれるかもしれない。例えば、これは呼吸器疾患又は COVID-19 の合併症のリスク因子と疑われる病態を対象とする試験において関心があるかもしれない。このような *estimand* の妥当性は、ウイルスが最終的に根絶されるのか、又は季節性インフルエンザのように持続するのかといった、このパンデミックの今後の変遷にも依存するであろう。後者の場合であっても、社会や臨床現場が新しい疾患に対応するにつれて、実際起こることや実臨床は現在の危機管理状況とは異なる可能性がある。

### 3.2. 関心のある治療条件（介入）

一般的に、関心のある治療条件は当初の意図したものと同じであるべきである。しかしながら、運用上、パンデミックと関連する理由により、治療の実施方法を変更する必要があるかもしれない。例えば、医療従事者による診療所での受診ではなく、被験者が自宅で

自己投与する治療など。そのような変更が実行可能な場合には、診療所への受診頻度を減少させ、その結果、感染暴露のリスクを減少させることが検討されてもよい。

試験治療のアドヒアランスや併用薬に伴うパンデミックと関連する合併症は、中間事象として考慮し、適切な戦略で対処すべきである。このような中間事象の程度は、試験中に被験者が受けた治療が意図したものを十分に代表するものであるかどうかという観点から評価されるべきである。これには、治療の中断、コンプライアンスの低下、及び治療方針戦略の対象となることが計画されている基礎療法、レスキュー療法、及びその後の治療へのアクセスが含まれる可能性がある。

### **3.3.対象集団**

主要な試験目的に沿うために、対象集団は当初の計画通りに維持すべきであり、パンデミックのためだけに変更すべきではない。試験の集団をさらに適正にするために、パンデミックとは無関係の試験実施計画書の修正も可能であるべきである。試験の選択/除外基準は、COVID-19 感染患者を除外する可能性を除いて、パンデミックがなかった場合の基準と比較してほぼ変更されないものとすべきである。

### **3.4. 変数 (評価項目)**

臨床評価項目は一般的に当初の計画通りに維持されるべきである。パンデミック中に代替の測定方法が必要となる可能性がある場合、例えば、中央検査と院内検査の比較、外来受診ではなく質問票の遠隔評価など、臨床評価項目の測定の質が損なわれないことを確認し、評価項目の変動性への潜在的な影響を評価する必要がある(4.3.2 節参照)。COVID-19 による死亡などのパンデミックと関連する中間事象を複合戦略で扱う場合は、評価項目の定義を調整する必要があるかもしれない。ランダム化から治療開始までの間に多数の遅延がある場合、評価項目がランダム化日から相対的に定義されている(例えば、イベントまでの時間が評価項目)場合には、治療開始日を評価開始日として再定義することを検討してもよい。しかし、非盲検試験の状況では、これは好ましくないかもしれない。

### **3.5. 治療効果の集団レベルでの要約**

各治療法の結果及び治療法間の比較を記述した集団レベルの要約は、一般的には変更されるべきでない。まれな状況では、例えば、当初計画されていた評価項目が数値であり、COVID-19 の死亡に対して生存した被験者の結果よりも悪いランク付けをする複合戦略が使用される場合などは、要約指標を変更する必要があるかもしれない。この場合、要約指標を平均値から中央値に変更するかもしれない。

もう一つの例は、Cox 比例ハザード回帰からのハザード比 (HR) であり、イベントまでの時間の評価項目を持つ試験で一般的に使用される治療効果の要約指標である。比例ハザードの仮定が満たされていない場合、HR の推定は試験で観察された特定の打ち切りパターンに依存する。これは、被験者の組み入れと脱落のパターンの両方に影響される (Rufibach, 2019)。パンデミック中の打ち切りパターンが、非パンデミック時とは異なるものの影響を受けている場合及び被験者の年齢や併存疾患などの共変量に依存したパンデミックと関連する追加の打ち切りが観察される場合、HR の外的妥当性と解釈可能性を慎重に検討する必要がある。同様に、ログランク検定の妥当性は、生存確率が試験の初期と後期に組み入れられた被験者で同じであるという仮定、及びイベントが記録された時間に起こったという仮定に依存している。このような仮定は評価する必要があるかもしれない。代替の要約指標を用いた感度分析的な estimand が検討されうる (例えば, Boyd et al., 2012; Nguyen and Gillen, 2012; Mao et al., 2018 を参照のこと)。

#### 4. 解析に対する影響とその緩和策

パンデミックの試験への影響により、計画された統計解析を修正する必要があるかもしれない。治療効果を適切に理解し、特徴づけるためには、追加の感度分析や補足的解析が必要となる場合がある。試験に応じて、計画された解析の修正は、例えば、比較的影響が少ない試験の場合のように比較的軽微なものから、パンデミックによって治験薬の投与や来院が著しく影響されるような状況での大規模なものまで、多岐にわたるであろう。解析上の考慮事項の一般的な要約を表 3 に示し、詳細な考察を後続の節で示す。

計画されているすべての変更及び追加解析は、データの盲検化を解除する前の SAP 及び総括報告書に文書化されるべきである。パンデミックの影響を完全に文書化し、この状況下での治療効果を特徴付けるためには、試験の盲検解除後に追加の事後探索的解析が必要となる場合もある。

表 3: 解析に関する考慮事項の概要

|                       |   |
|-----------------------|---|
| <p>計画された解析のレビュー</p>   | <ul style="list-style-type: none"> <li>• 計画された主要な解析及び感度分析をすべてレビューし、改訂されたestimandとの整合性を確保する。</li> <li>• パンデミックと関連する欠測に対応するために、欠測データの取り扱い又は打切りのルールをレビュー／修正する。</li> </ul>  |
| <p>パンデミックの影響の要約</p>   | <ul style="list-style-type: none"> <li>• パンデミックと関連する中間事象と治験実施計画書からの逸脱の発生を要約する。</li> <li>• すべての重要な評価項目に対する欠測又は使用不可能な評価の数を要約する。</li> <li>• 代替的な方法を使用して実施した評価の数を要約する。</li> <li>• パンデミック発生前後の試験集団特性を要約する。</li> </ul>  |
| <p>追加の感度分析及び補足的解析</p> | <ul style="list-style-type: none"> <li>• パンデミックと関連する欠測に対する追加の感度分析を計画する。</li> <li>• 追加の代替的な治療効果の要約指標の必要性を検討する。</li> <li>• 解析方法に追加の補助変数、交互作用、時間依存性の外生的な共変量を含めることを検討する。</li> <li>• パンデミックの影響によって定義される部分集団（パンデミックの発生前後の主要評価項目のための来院など）に基づく部分集団解析を検討する。</li> <li>• 代替的なデータ収集方法の潜在的な影響を評価する必要性を検討する。</li> <li>• 例えば代替的な方法の使用を正当化するために、試験の外部のデータソースを検討する。</li> </ul> |

#### 4.1. 有効性解析のための考慮事項

すべての予定されている有効性解析は、3節のガイダンス及びパンデミックと関連する欠測データの取り扱いを考慮して再評価すべきである（4.2節参照）。中心となる解析手法は変更すべきではない。しかし、estimandの修正、パンデミックと関連する中間事象のために選択されたストラテジー、パンデミックと関連する欠測データの取り扱いによっては、計画されていた解析への変更が必要となる可能性がある。また、パンデミックの混乱の影響を評価するために、追加の解析が必要となることもしばしばである。被験者のアウトカムがパンデミックの影響を直接受ける可能性のある試験や評価項目、例えば呼吸器疾患やQOLの評価項目については、特別な配慮が必要となるかもしれない。

パンデミックのために登録が中断された試験では、治験依頼者は中断前と中断後に登録された被験者集団を比較すべきである。より一般的には、パンデミック期間にわたって登録された被験者集団の変化を評価すべきである。ベースラインの特性（人口統計学的特性、ベースラインの疾患特性、関連する病歴を含む）を登録期間ごとにまとめ、パンデミック

期間と比較して登録された被験者集団に関連する差があるかどうかを評価することができる。パンデミック対策が実施されたり解除されたりすると、登録の開始と中止が国によって異なる可能性があるため、登録率の地域差と関連して、被験者集団に変化が生じる可能性がある。

## 4.2. 欠測データに対する影響とその緩和策

治験依頼者は、COVID-19 パンデミックの間、被験者と試験関係者の安全を損なうことなく、また医療システムに過度の負担をかけることなく、データの欠測を最小限に抑えるためにあらゆる努力をすべきである。被験者及び施設にとって実行可能で安全な場合はいつでも、被験者は試験に参加し続け、評価は継続されるべきである。その際、有効性の主要評価項目と安全性評価項目を優先し、その後、主要な副次評価項目を優先する。治験依頼者は最善の努力をすとしても、欠測データの増加と異なるパターンに備えるべきである。

ICH E9(R1)の枠組みでは、評価又は評価項目の値は、収集が計画されており、特定の estimand に有用であると考えられていたが、解析に利用できなかった場合、欠測と考えられる。仮想ストラテジーによって対処される中間事象の場合、評価項目の値は仮想的なシナリオの下では直接観察されない。このようなデータは、ICH E9(R1)の定義の意味で欠測しているわけではないが、解析では欠測データの取り扱いと同様の方法を用いてモデル化する必要がある。本論文の残りの部分では、欠測データの取り扱いに関する方法について議論し、モデル化の仮定が対応する中間事象に対処するために選択された仮想的なシナリオと一致している場合には、そのような方法が中間事象後の観察されないデータのモデル化に有用であることに留意する。

### 4.2.1. パンデミックと関連する欠測の評価と文書化

治験依頼者は、影響を受けた試験におけるパンデミックと関連する欠測データのパターン（量と理由）を評価し要約するべきである。データが欠測している理由としては、a) 計画された評価が実施できなかった、b) 収集されたデータが解析に使用できないと判断された（例：タイムウィンドウ外）、又は c) 期待される仮想シナリオの下でのデータが中間事象後に観察できなかった（例：打ち切り）、のようなものが考えられる。さらに、パンデミックと関連する各々の欠測事例には特定の理由と状況がある。大別すると、パンデミックと関連する欠測データの理由は、構造的なもの（例えば、政府による強制封鎖や施設の研究活動の停止）や、被験者固有のもの（例えば、個々の COVID-19 疾患、合併症又は COVID-19 に対する個々の懸念）の可能性がある。表 4 には欠測データの影響を評価し、解析上での扱いを計画するための包括的なイメージを提供する各観点が示されている。し

たがって、治験依頼者はこれらの情報を可能な限り臨床試験のデータベースで取得すべきである。表4の最後の2行は、中間事象の文脈で考慮される状況と同様の状況を反映している（表2参照）。欠測データは、中間事象が存在する場合（例えば、仮想ストラテジーによって処理される）と中間事象が存在しない場合（例えば、被験者が試験治療を遵守し続けているが、いくつかの評価を逃している）の両方で起こりうるため、ここではそれらもリストアップする。

表 4. パンデミックに関連する欠測データの属性.

|                                |   |
|--------------------------------|---|
| <p><b>評価項目の未測定</b></p>         | <ul style="list-style-type: none"> <li>• パンデミック関連の理由による試験早期中止により被験者の評価が未実施</li> <li>• 被験者は試験参加中ではあるが、来院しない／手順未実施のために評価が未実施（間欠的な欠測データ）</li> <li>• 評価遅延（タイムウィンドウ外）のため解析に使用されない</li> <li>• 複合スコア（例えば、関節リウマチのACR20）の一部の要素が欠測しているため算出不可</li> <li>• パンデミック関連要因の影響を受ける評価と特定の分析では使用不可能と見なされる</li> <li>• 結果の解釈可能性が影響を受ける可能性があるため（例えば、生活の質、活動／機能尺度、医療利用率の評価など）</li> <li>• 記録されたデータが、COVID-19関連要因により適切に照合又は判定できず、解析に用いるには信頼性が低いと判断</li> <li>• 仮想ストラテジーで処理することを意図した中間事象の後に実施された評価で、当該estimandでは使用不可能と判断された</li> </ul> |
| <p><b>評価の利用可能性</b></p>         | <ul style="list-style-type: none"> <li>• 施設（施設又はスタッフ）が試験関連の評価について利用不可能</li> <li>• 施設／評価手続きは利用可能だが、パンデミックと関連する被験者の個人的な理由により評価を受けることができない／受けたくない</li> </ul>   |
| <p><b>被験者のCOVID-19感染状態</b></p> | <ul style="list-style-type: none"> <li>• COVID-19陽性かつ生存している被験者</li> <li>• COVID-19により死亡した被験者</li> <li>• 既知のCOVID-19感染の無い患者</li> </ul>   |
| <p><b>被験者のCOVID-19併用治療</b></p> | <ul style="list-style-type: none"> <li>• COVID-19（薬理的、酸素吸入など）の治療を受けた被験者</li> <li>• 集中治療室ではないが入院した被験者</li> <li>• 集中治療室に入院した被験者</li> </ul>  |

1行目はデータが欠測している理由を要約し、2～4行目は、これらの理由への関連している条件を要約している

治験依頼者はまた、来院が無かった場合や評価手段の変更（例えば、診察・検査の代わ

りの電話でのフォローアップなど)により、パンデミック期間中の症状やAEが過少報告される可能性も考慮すべきである。(4.4節参照)

治験依頼者はまた、欠測パターンを複数の側面から報告することを検討すべきである。すなわち、経時的(来院毎及びパンデミック前後)に、特定の人口統計学的及びベースラインの疾患特性、COVID-19の感染又は結果に対する潜在的な危険因子と考えられる併存疾患、及び地理的な地域に関して検討すべきである。

試験の盲検解除に先立つ盲検下での欠測データパターンの要約は、欠測データの取り扱い方法の選択に示唆を与える。現在の試験の欠測データパターンを、類似した過去の試験と比較することは、特に部分集団内での欠測状況に関して有用であろう。

試験の盲検化を解除した後(最終解析又は非盲検下でのDMCの解析のため)、欠測データパターンを全被験者及び治療群別で要約すべきである。ほとんどの場合、パンデミックに関連する欠測、特に構造的な欠測は、治療群間で差があるとは想定されないが、このような可能性を排除すべきではない。特別な場合、来院しないことが治療と関連している場合がある。例えば、非盲検試験で、被験治療は施設で実施されなければならない一方対照治療は自宅で実施できる場合、対照治療群でより多くの評価欠測が生じる。緩和策が実施されない場合、治療効果の推定値に偏りが生じる可能性がある。これは、外部対照を用いる単群試験の場合にも当てはまる。

#### 4.2.2. 主とする解析における欠測データの取り扱い

治験依頼者は一般的に、パンデミックと関連しない欠測データの取り扱いについては、治験実施計画書及びSAPで当初計画されていたものと同様の手法を維持すべきである。パンデミックに関連する欠測データについては、各estimandと解析方法の背景に応じて適切な戦略を特定する必要がある。どの戦略が最も適切であるかは、表4に示されているように、基礎にある背景と欠測理由に照らして、また解析の対象となるestimandとの整合性を考慮して検討されるべきである。3つのケースを説明する。

- (1) 被験者が中間事象を持たずにデータが欠測している場合、すなわち、被験者が試験治療を継続しているが、評価項目の値がいくつか欠測している場合：欠測データのモデル化は、被験者が試験治療を継続しているという事実とその被験者観察データを考慮したうえで、欠測値が本来取りうるべき値についての臨床的に尤もらしい仮定に基づくべきである。
- (2) 中間事象が存在する状態でデータをモデル化する場合：中間事象に対処するためにestimand内で定義されたストラテジーを考慮すべきである。
- (3) 中間事象後の結果を仮想ストラテジーによってモデル化する場合：仮定は、estimandで特定された仮想シナリオの下で未観察の結果がどのようなものでありえたかを反映し、



中間事象を取り巻く状況で臨床的に妥当なものでなければならない。例えば、COVID-19 感染症の合併症の有無は異なる全体的な健康状態を示す可能性があるため、試験治療中止の中間事象による未観測データに対して、被験者の COVID-19 感染の合併症の有無に応じた異なるモデルをあてはめることができる。

最近の文献や規制当局によって推奨されている補完法や欠測値を暗黙的に取り扱う方法は、ほとんどの場合に対して、パンデミックと関連する欠測を取り扱う手法の適切な選択方法を与えている（例えば、Molenberghs and Kenward, 2007; NRC, 2010; O'Kelly and Ratitch, 2014; Mallinckrodt et al., 2019 参照）。

欠測データを扱うための方法は、欠測メカニズムに対する仮定の種類に基づいて分類されることが多い。Molenberghs and Kenward (2007)の経時データを伴う試験に対応した欠測メカニズムの分類を用いると、欠測確率がすべての被験者関連因子から独立しているか、又は適切なランダム化前の共変量で条件付けることにより欠測確率が観察された結果と観察されていない結果のどちらにも依存しない場合には、データは完全にランダムに欠測している（missing completely at random: MCAR）と表現する。（Little and Rubin (2002)の枠組みでは、MCAR は観察された要因と観察されていない要因のどちらにも依存しないと定義されていることに注意しなければならない。この定義はその後、ランダム化前の共変量への依存性を包含するように一般化され、共変量に依存した欠測 (covariate-dependent missingness) とも呼ばれている。MCAR は現在、両方のケースにおいて文献で使用されている。）パンデミックに関連する欠測のうちいくつかのタイプは MCAR とみなされるかもしれない。例えば、施設が臨床試験に関連するすべての活動を一時的に停止している場合である。

パンデミックの前後に収集されたデータ量に応じて、そのような被験者を主要な解析対象集団から除外すべきかどうかを検討することができる。例えば、一部の被験者ですべての（大部分の）データが欠測している場合、利用可能なデータを持つ被験者から作成されたモデルに基づいて欠測している被験者のデータを補完しても、推測に新たに追加される情報は無く、そのような被験者を除外してもバイアスが生じる可能性は低い。予想される治療効果の発現時期より前の来院時のみのデータを持ち、かつ残りのデータが MCAR であるような被験者を解析対象集団に含めると、治療効果に関する推測に情報は追加されない一方、欠測データによる不確実性が追加される可能性がある。

適切な（ランダム化前の）共変量と観察された結果（例えば、被験者が試験を中止する前の）を条件付けた下で、欠測確率が未観測の結果に依存しない場合、データは missing at random (MAR) である。欠測に関連する施設特有及び被験者特有の情報が試験中に収集されていれば、パンデミックに関連する欠測の大部分は MCAR 又は MAR と考えられる。

MCAR と MAR のメカニズムの定義は、欠測を説明するための一連の共変量と観察され

た結果を与えた下での欠測データの条件付き独立に基づいている。パンデミックに関連する欠測を説明する因子には、MAR の場合に含まれるランダム化後の結果に加え、追加の共変量が含まれるかもしれない。例えば、パンデミック中の欠測は、COVID-19 感染の合併症などのパンデミックに関連するランダム化後の結果だけでなく、年齢や合併症などの追加のベースラインの特徴にも依存しているかもしれない。

共変量依存性の欠測の場合、適切なベースライン共変量に対する回帰分析は正しい推測には十分であるが、条件付きの治療効果推定値と周辺治療効果推定値が一致しないロジスティック回帰や Cox 回帰のようなモデルでは、解析モデルと治療効果の解釈を複雑にする。

MCAR と MAR の下では、直接尤度、例えば、**mixed models for repeated measures**

(MMRM) のようないくつかのモデル化フレームワークは、結果と欠測のそれぞれの分布のパラメータ空間の分離可能性を利用することができる。この場合、欠測は無視可能とみなされ (Molenberghs and Kenward, 2007)、欠測のみに関連する因子は、結果に対する周辺治療効果についての推測に含める必要はない。しかし、これはすべての推測のフレームワークに適用されるわけではない。多重補完法 (**multiple imputation : MI**) (Rubin, 1987) は、事前に計画した解析モデルを利用しながら、補助変数 (ランダム化前とランダム化後の両方) を補完モデルに含めることができる点が有用でありうる。補助変数を用いた多重補完法は、連続、2 値、計数、及びイベントまでの時間を含む様々なタイプの評価項目に使用可能であり、解析段階で様々な推測手法と組み合わせることができる。複数の補完後データセットからの推測結果を統合するための **Rubin's rule** を用いた多重補完法の使用は、効率減少による検出力への影響を与えうるが、いくつかの代替案が存在する (例えば、Schomaker and Heumann, 2018; Hughes et al., 2016) 。

COVID-19 関連の中間事象に対して仮想ストラテジーを適用するためには、イベントまでの時間を評価項目とした状況の場合、(MCAR 設定の場合には) 適切なベースライン共変量で調整した回帰モデル (例えば、Cox 回帰)、又は (MAR 設定の場合には) 多重補完法が推奨される。結果を完全又は部分的に打ち切るパンデミック関連中間事象、例えば COVID-19 関連の死亡などを競合イベントとして扱う競合リスク解析は、仮想ストラテジーとは両立しない。この状況において競合リスク解析の解釈がさらに複雑になるのは、競合イベントのリスクが各被験者の起点 (例えば、ランダム化又は治療開始) からではなく、試験コホート内で同時期に生じないパンデミックの間だけであるためである。これは、被験者間で共通の競合リスクを仮定する競合リスク解析の妥当性を損なうとともに、そのような解析から得られた結果の一般化可能性を妨げる。

特定の欠測メカニズムを仮定することの意味するところは、観察データで条件づけることにより、他の「類似した」被験者で観察されたランダム化前の共変量又は共変量とランダム化後の結果との関係 (MAR) を用いることにより、欠測した結果をモデル化できることである。

パンデミックに関連する欠測については、たとえ欠測につながる要因が既知で収集されていたとしても、モデル化に十分な「類似した」被験者の観測データがあるかどうかを評価することが重要である。例えば、欠測と評価項目が年齢に依存しており、利用可能な評価項目データを持つ高齢の被験者が少ない場合、主に若い被験者からなる利用可能データから得られたモデルを用いて高齢の被験者の欠測データをモデル化することは適切ではないかもしれない（結果として、信頼性の低い外挿となるかもしれない）。同様に、COVID-19感染による重度の合併症は、これらの被験者の健康状態が全体的に悪いことに起因している可能性があり、合併症のない被験者のデータに基づいて結果をモデル化することは正当化できないかもしれない。データが欠測している被験者と観測されている被験者との比較についての追加の仮定を置いて正当化する必要がある。この仮定はヒストリカルデータに基づくこともありうる。ここでは、表4で述べたような、欠測を取り巻くパンデミックに関連する要因を考慮することが重要となる。

パンデミックに関連する欠測データは、欠測と結果がCOVID-19に関連した危険因子、治療、感染状態に依存するもののそれらのデータが収集されていない場合には、missing not at random (MNAR) と考える必要があるかもしれない。欠測メカニズムも、未観測の結果に依存している場合は、MNARであるかもしれない。パンデミックの状況では、軽症又は治療への反応が低い被験者が試験又は治療を中止し、中止前に結果及び中止理由が記録されていない場合に発生する可能性がある。MNARの下での解析では追加の検証不能な仮定を追加する必要があるが、適切なデータを収集することで回避できる可能性がある。MARの仮定からの逸脱に対する感度分析は、例えば、尤もらしいMNARメカニズムを仮定したモデルを考慮すべきである（例えば、Carpenter et al., 2013; Mallinckrodt et al., 2020を参照）。

パンデミックに関連する欠測データを取り扱うために必要な、計画された主要解析の修正は、試験の盲検化を解除する前にSAPで規定されるべきである。欠測データに関する感度分析の議論については、4.3.1節を参照されたい。

#### 4.3. 感度分析及び補足的解析に関する考察

パンデミックに関連する混乱が試験に与える影響を評価するために、多くの場合で、追加の感度分析及び補足的解析が必要となる。パンデミックに関連しないイベントと欠測データについては、もともと計画していた感度分析を実施すべきであるが、事前に計画していた感度分析をパンデミックと関連する中間事象及び欠測データ、パンデミックと関連しない中間事象及び欠測データの両方に単純に適用することは、3つの理由から問題があるかもしれない。第一に、COVID-19によるパンデミックが発生していなかった場合における治療効果を推定するためには、パンデミックに関連したイベントとパンデミックと関連しないイベントでは異なるストラテジーを必要とする可能性がある。第二に、4.2.2節で議論さ

れているように、COVID-19に関連したイベントではMCAR又はMARの仮定が妥当であることが多いが、この仮定は他の欠測データにはあてはまらないかもしれない。第三に、パンデミックと関連する欠測データが多い場合、欠測データに対する感度分析が過度に保守的になる可能性がある。欠測データが比較的多いということは、通常であれば、試験のデザインや実施方法に問題があるという指標になり得て、試験結果の不確実性を高めることにつながってしまうが、パンデミックに起因する欠測データが多い場合は、この前提は弱くなる。

登録期間又はパンデミック（4.3.3節参照）期間別の主要評価項目及び副次評価項目の部分集団解析が推奨される。部分集団解析で治療効果の不均一性が示唆された場合には、その原因が地域間差に起因する可能性があるか考慮すべきである。ICH（2017）に記載されているような国際協同試験の問題は、パンデミックによって増強される可能性がある。加えて、探索的に、時間依存的に治療効果を評価することもできるかもしれない。例えば、経時データやイベントまでの時間を評価項目としたモデルでは、追跡期間中の患者の動的な状態を記述する外生的な時間依存性共変量と治療の割り当てとの間の交互作用項を含めることができる。しかし、このような分析から得られる結果は、解釈したり一般化したりすることが簡単ではない場合がある。

#### 4.3.1. 評価の遅延や欠測データに対する感度

特定の試験によっては、評価の各来院に対する評価可能な期間幅（タイムウィンドウ）を広げることが適切であるかもしれない。感度分析では、評価スケジュール外のデータを含めたり、含めず欠測データとして取り扱ったりすることで、結果の安定性を評価すべきである。

Tipping point 解析（Ratitch et al., 2013）は、主解析の結論を覆すために、主とする推定量における欠測データに対する仮定からのずれが、どの程度大きくなければならないかを評価するために使用されることがある。Tipping point の調整は、パンデミックに関連した欠測と関連しない欠測の間で、また、欠測の理由によって異なるかもしれない。例えば、パンデミックと関連のない欠測データに対して tipping point の調整を実施するが、パンデミックと関連する欠測データに対しては標準的なMARに基づく補完を使用することができる。ヒストリカルデータが仮定と tipping point の調整に対する妥当性を意味づける上で有用かもしれない。

主解析が補完に基づいた解析である場合、感度分析は、補完モデルの中で、より多くの変数を用いて実施することも可能である。

中間事象に対して仮想ストラテジーを用いている状況で欠測データを取り扱う場合、被験者が試験を通じて治療を継続できたであろう確率とパンデミックと関連しないイベント

を発現するであろう確率を変えて、結果を補完することも可能である。

2 値データのレスポンス解析では、欠測である被験者が少ない場合は、これらの被験者をノンレスポンスとして扱うことは珍しくない。パンデミックと関連した評価の遅延や欠測が、特に中間事象がない状況で発生した場合においては、代わりに MAR の仮定に基づく仮想ストラテジーを用いることが望ましいかもしれない。しかし、補足的解析として、パンデミックと関連しない及び関連する欠測をノンレスポンスとして補完して報告することは可能であるかもしれない。評価項目がイベントまでの時間の場合は、感度分析として、関心のあるイベントが、来院していないもしくは遅延して来院した期間で発現したことは分かっているが、イベントまでの時間が正確には分からないという場合に、区間打ち切り法を用いることを推奨する (Bogaerts, Komarek, and Lesaffre, 2017)。

#### **4.3.2. 治験実施計画書で指定された試験データ収集に対する代替測定に対する感度**

パンデミック期間中は、評価項目の代替測定が必要となる場合がある。これらの代替測定が治験実施計画書で定める標準的な測定方法と交換可能かどうかを判断するためには、試験ごとに慎重に評価をする必要がある。理想的には、交換可能性は実施時に試験の外部情報に基づいて確立される。そうでない場合は、例えば、代替測定値の分布と元の測定値との比較を行うといった盲検下での解析で裏付けることができる。しかし、新しい測定法の妥当性が以前に確立されていない場合、試験のデータのみを用いて厳密に同等性を実証することは困難である。

交換可能性が確立されているか、あるいは想定されている場合、主とする推定量には、元々の方法で収集されたデータ及び代替の方法で収集されたデータの両方を含めることができる。感度分析では、元々の治験実施計画書で定められた方法で収集されたデータのみを含め、それ以外のデータを欠測として取り扱うことになる。交換可能性の仮定の妥当性が不確かな場合は、逆の方法をとることができる。治療と評価法の交互作用をモデル化することは、別の感度分析として実施することができる。

#### **4.3.3. パンデミックが試験結果に与える影響を理解するための課題**

パンデミックが試験結果に与える影響を理解することが重要になる。これをどのように行うことができるかについては、いくつかの考え方があり。前節では、未来院や治療中断といったイベントがパンデミックにどのように起因するかを説明するデータを収集する必要性について述べてきた。これらのデータはパンデミックと関連した中間事象と欠測データの定義、これらを取り扱うストラテジーで考慮される。そうすることで、試験データの統計解析で、パンデミックの影響が適切に調整されることになる。多くの点で、これは、各被験者について分かっている情報を直接解析に取り入れ、非常によく理解されている理

想的なアプローチである。これは、臨床試験で避けられない変動に対して統計解析を調整するための標準的なアプローチである。このアプローチには、パンデミックと関連するデータの詳細を収集する必要性があるという欠点があり、これは状況によっては実施可能ではない場合がある。

別のアプローチでは、試験データベースの外部で定義されたパンデミック期間を使用する（例えば、EMA/CHMP, 2020b に記載されているように、パンデミック前／期間中／パンデミック後のフェーズを定義する）。このアプローチでは、パンデミック期間を正確かつ精度高く定義する必要がある。この方法は、パンデミックの影響と現地で封じ込め対策が被験者間で比較的同質であろう単一国における試験で適用するのがより簡単である。しかし、単一国であっても、パンデミックは緩やかに進展する可能性がある。そのため、パンデミックの開始日と終了日の定義が複雑になり、パンデミックが被験者に与える影響は一律ではない可能性が高い。さらに、感染症のアウトブレイクにはいくつかの波があるかもしれない。グローバル試験では、これらのパンデミック期間の開始日と終了日、及びパンデミックが被験者に与える影響は地域によって大きく異なる可能性があるため、これらの定義は、さらに困難であると示される可能性がある。実際には、地域ごとに報告された COVID-19 の症例数と死亡者数、又は地域ごとの封じ込め対策の開始日と中止日に基づく標準化された実用的な定義が必要となるであろう。パンデミック期間が定義されれば、異なるパンデミックの期間における来院に対する時間依存性の変数を前に議論したように、統計モデルや中間事象の定義に適切に組み込むことができる。このアプローチを使用する場合は、パンデミックの開始日と終了日を定義する根拠を文書化しなければならない。状況は急速に変化しており、現時点では、これらのパンデミック期間の定義、実施、解釈に関する決定的な推奨事項を提供することはできない。

パンデミックの影響を一般的に評価するための第三のアプローチでは、各被験者は、治療や試験データベースに収集される評価に対するパンデミックの影響の程度に応じて分類することができる（治験実施計画書からの逸脱の詳細、中間事象、評価の欠測、パンデミックと関連した中止の理由など）。フォローアップ期間が固定され、追跡不能が少ない試験の場合、この分類は、例えば標準的な部分集団解析における部分集団の定義のような標準的な解析に組み込むことができる。

現在のところ、この問題に対する単一のアプローチを支持するには十分なエビデンスがない。治験依頼者は、少なくとも最初の 2 つのアプローチについては、規制上のガイダンスの対象となっているため、準備を進めている。これらがどのように展開し、パンデミックがどのように発展するかなどを見るまでは、パンデミックの影響をまとめる複数の方法を検討するのが賢明である。

#### 4.4. 安全性解析の考慮事項

標準的な安全性の要約には、通常通りすべての AE が含まれる。しかし、安全性プロファイルを完全に理解するためには、COVID-19 の感染と関連したイベントと関連しないイベントについて、それぞれ別個の追加解析が必要になるかもしれない。AE 及び特に死亡が COVID-19 と関連しているかどうかの判断は、バイアスを避けるために、データの盲検化を解除する前の臨床試験のモニタリング中に行うべきである。

多くの状況では、安全性報告に変更はないであろう。しかし、COVID-19 パンデミックによる混乱から、COVID-19 の感染や死亡者の発生だけでなく、治療の中断、中止、試験からの脱落の増加につながる可能性がある。したがって、3 節で概説された estimand の枠組みは安全性解析にも有用であり、安全性のためのイベントまでの時間に対する解析や競合リスク解析及び estimand の一般的な議論については (Unkel et al, 2019) を参照のこと。介入の安全性を適切に評価するために物理的な来院を必要とする試験では、要件を満たすための物理的な来院スケジュールを維持する必要がある。AE データ収集におけるバイアスの可能性への一般的な対処については、本稿の範囲を超えている。服薬遵守率を考慮した曝露や追跡期間で調整した解析を行うことができる。例えば、パンデミック前とパンデミック中で調整した率を比較する、又はヒストリカルデータを用いて、調整した率を比較する。現時点では他の方法論上の推奨事項はなく、より多くの研究が必要である。

#### 5. 試験レベルの課題と緩和策

前節で議論した欠測データや統計モデルの変更による累積的な影響は、試験全体レベルの影響になりえる。累積的な影響は、試験の目的を達成する性能、あるいは試験結果の解釈可能性を変える可能性がある。

##### 5.1. 欠測データと試験の中断・中止の影響の評価

治験依頼者は、欠測データの影響をいくつかの側面から評価すべきであり、これらの懸念事項のいくつかについては、規制当局との合意を得ることが重要であるかもしれない。

- 収集可能と予想されるデータ量を考慮したうえでの計画された推定方法の実行可能性。
- 治療群間で欠測パターンに重要な違いがある場合、治療効果推定にバイアスがかかる可能性。
- 主要な目的及び重要な副次的目的に対する検出力。
- 試験結果の解釈可能性。

- 薬剤の総曝露時間が減少している可能性があることを踏まえた安全性データベースの妥当性、及び AE の過小報告の可能性.
- 規制当局への申請に求められる地域的なエビデンスの妥当性.

## 5.2. 検出力と成功確率

前節で述べたように、COVID-19に関連する要因は試験データに様々な面で影響を与え、そして結果的に、検出力、成功確率、被験者数、又は試験デザインの他の側面に影響を与える。

様々なパンデミック要因が試験結果に及ぼす潜在的な影響は、臨床試験のシミュレーションにより定量化が可能である。シミュレーションのモデルは当初の試験デザインの他、3節と4節で議論されているように、estimandの修正、欠測データの取り扱いと解析方法に応じた設定する。試験の完全性を維持するために、シミュレーションは、試験からの盲検化されたデータと、外部情報源からのデザインパラメータの想定値のみに基づいて行われるべきである。試験デザインで使用された治療効果の推定値とバラツキを、元の値から変更することも考えられる。検出力や成功確率等の試験の特性もそれに応じて、更新されることになる。シミュレーションの結果によっては、被験者数の再設計が検討されうるであろうし、主要評価項目の変更、解析方法の変更、中間解析の追加とそれに伴う適応など、試験デザインのより大規模な変更も検討されるかもしれない (Posch and Proschan, 2012; Kieser and Friede, 2003; Muller and Schaefer, 2004)。このような変更は困難であり、規制当局と協議すべきであるが、試験の完全性が維持できるのであれば検討の余地があるであろう。

試験によっては、被験者数を増やすことは現実的でなく、登録目標を下回るようになる可能性がある。現在の特殊な状況を考慮して、より柔軟性をもち、対照群のヒストリカルデータやリアルワールドデータの使用を含め、複数の試験や情報源にまたがってエビデンスを定量化する方法を検討することを我々は推奨する。ただし、そのようなデータの入手先や選択方法については、事前に計画し、規制当局と合意する必要がある。データの盲検解除後、意味のある治療効果が観測されているものの、COVID-19の影響により統計学的基準を満たさない場合、治験依頼者は、上述の方法から蓄積されたエビデンスに基づいて、その試験結果が規制上許容されるものであるかどうかを評価することもできる。又は、後続試験のための推論的事前情報として試験結果を使用することができるかどうかを評価することができる (Viele et al., 2014)。



### 5.3. DMC 及び中間解析に関する考慮事項

DMC を有する試験については、被験者の安全を保護するため、及び運営上の問題に対処するために取られたすべての手段について、DMC に十分に情報提供されていることを治験依頼者は確認すべきである。データに関して既知の、又は潜在的な問題点がある場合は、それを伝えるべきである。事前に計画されていた定期的な安全性中間評価のタイミングは、再検討する必要があるかもしれない。さらに、提示するデータの変更や追加が必要になるかもしれない。本節で議論されているような中間解析の状況によっては、既に存在していない場合には、DMC を設立する必要があるかもしれない。また、被験者の安全性に関連して、盲検化されていないデータを緊急に検討する内部の DMC（試験の他のメンバーから適切にファイアウォールで隔離されていることが推奨される）を設ける必要が生じる状況もあり得る（例えば、DMC を設置しておらず既存の外部 DMC を編成するのに何ヶ月もかかる可能性がある場合）。

有効性の中間解析は、治験実施計画書に記載されている情報レベル（例えば、主要評価項目の被験者数や特定の情報分数）に応じて計画通りに実施すべきであり、そのために期待されていたタイムラインより遅れが生じる可能性がある。計画していない試験途中での有効性又は無益性の中間解析は、安全性及び倫理的な理由がなければ一般に推奨されない。しかし、計画された情報量のレベルに到達することが現実的でない場合には、中間解析の計画を変更することを検討する必要がある。例えば、カレンダータイムを基準としたタイミングとすることなどが考えられる。計画外の中間解析を行うことに強い科学的根拠がある場合には、DMC に情報を提供し、解析のタイミングと実施計画について協議すべきである。パンデミック関連の混乱やデータの欠測に対処するため、estimand、計画された解析方法、又は決定基準が変更された場合、その変更は DMC に伝えられ、DMC の手順書に明記されるべきである。

臨床試験実施の決定に盲検化されていないデータが影響を及ぼすことを防ぐため、運営上のリスクを評価/低減するプロセスとして DMC を利用することは、一般的に推奨されない (EMA/CHMP, 2005; FDA, 2006)。依頼者が DMC を関与させることを決定した場合は、詳細を明確に定義し、DMC メンバーの追加的な責任とバイアスを防止するための対策を含め、DMC の手順書に明記されるべきである。

## 6. 結論

これまで述べてきたように、COVID-19 のアウトブレイクは、計画され、進行中である臨床試験に大きな影響を与え続けている。試験データへのこの影響は複数の意味合いを持っている。多くの場合、これらの影響は個々の臨床試験を超えており、そのような結果が他の試験結果とあわせて有効性や安全性の統合サマリーとして含まれる場合でも考慮する

必要がある。私たちの目的は、COVID-19 が現在進行中の臨床試験に与える潜在的な影響から生じる統計的な問題の性質を説明し、この問題に対処するための解決策として一般的な提言を行うことであった。

以下は最も重要な知見と推奨事項である。

- 試験の実施、データ収集、及び解析に関するリスク評価、緩和策、及びすべての変更は、統計解析計画書、及び必要に応じて総括報告書に明記されなければならない。変更によっては、治験実施計画書の修正や規制当局との協議が必要となる場合がある (FDA, 2020; EMA, 2020a, 2020b)
- 試験データの統計解析に対する運営上の緩和策、特に主要な有効性と安全性の評価項目に対するものは、それらを実施する前に影響を考慮すべきである。別のデータ収集方法の妥当性と交換可能性については、慎重な検討が必要である。
- 5つの重要な要素で構成される estimand のフレームワークは、パンデミックが主要な試験目的に及ぼす影響を体系的かつ構造化された方法で評価するための筋道を提供するものであり、estimand が治験実施計画書に正式に定義されているかどうかに関わらず有用である。
- 可能な限り、試験の当初の目的を維持することを推奨するが、計画された estimand への一定の影響は避けられないかもしれない。予測できないパンデミックの影響を適切かつ厳密に説明するためには、パンデミックに関連した中間事象を定義する必要があるであろう。
- 計画されている有効性及び安全性解析は、推定値及び欠測データを取り扱うストラテジーが更新された estimand と整合するように、必要な変更がないか慎重に見直されるべきである。パンデミックに関連した混乱の影響を完全に説明するためには、追加の感度分析及び補足的解析が必要であるかもしれない。
- 治験依頼者は、被験者と研究従事者の安全性を損なうことなく、また医療制度に過度の負担をかけることなく、欠測データを最小限にするためにあらゆる努力をすべきである。この際、主要評価項目、重要な安全性評価項目、次に重要な副次的有効性評価項目に関する評価を優先すべきである。
- パンデミックのために欠測しているデータの多くは、特にそれらが構造的な理由で欠測している場合には、MCAR 又は MAR であると主張されるかもしれないが、特に特定の疾患や被験者に特有の欠測については、追加的な考慮が必要となる場合がある。
- 治験依頼者は、試験とデータの完全性に関する厳密かつ体系的なリスク評価を実施し、定期的に更新すべきである。臨床試験が目的を達成する能力をどれだけ保持しているかは、臨床試験の estimand への影響、欠測データ及び欠測データの取り扱い、及び必要となる解析方法の修正内容を考慮して、定量的に評価されるべきであ

る.

## 謝辞

現在進行中の臨床試験において、これらの問題を解決するために多くの時間を割いてくれた各企業の同僚のサポートに感謝しています。彼らは寛大にアイデアを共有していただきました。この原稿はこのような幅広い意見から恩恵を受けています。また、著者チームをまとめ、貴重な意見を提供して下さった”**Biopharmaceutical Statistics Leaders Consortium**”のメンバー、タイトなスケジュールの中で広範で有益な意見を提供して下さった編集者と査読者にも感謝します。

## 引用文献

省略