



# Estimand の治験実施計画書への実装

日本製薬工業協会 医薬品評価委員会 データサイエンス部会

2021年度 TF4/2022年度 KT4

プロトコルサブチーム

## 目次

1. はじめに.....	6
Part 1: 臨床試験に携わるすべての方へ.....	9
2. ICH E9(R1)と estimand.....	9
2.1 治療効果と中間事象の定義.....	9
2.2 Estimand のフレームワーク.....	12
2.3 Estimand の 5 つの要素.....	13
2.4 中間事象を取り扱うストラテジー.....	15
2.4.1 治療方針 (Treatment Policy) ストラテジー.....	15
2.4.2 仮想 (Hypothetical) ストラテジー.....	16
2.4.3 複合変数 (Composite Variable) ストラテジー.....	17
2.4.4 治療下 (While-on-Treatment) ストラテジー.....	18
2.4.5 主要層 (Principal Stratum) ストラテジー.....	19
2.5 感度分析と補足的解析.....	21
2.5.1 感度分析 (Sensitivity Analysis).....	21
2.5.2 補足的解析 (Supplementary Analysis).....	22
2.6 Estimand 実装のためのプロセス.....	22
2.6.1 ステップ 1: 疾患と関心のある治療の設定に基づく試験目的の決定 ...	22
2.6.2 ステップ 2: 中間事象の特定.....	23
2.6.3 ステップ 3: 中間事象に対応するストラテジーの検討.....	23
2.6.4 ステップ 4: estimand の設定.....	24
2.6.5 ステップ 5: 試験デザイン, データ収集, 解析手法の選択.....	24
2.6.6 ステップ 6: 主解析における仮定の特定及び適切な感度分析の設定 ...	25
2.6.7 ステップ 7: estimand の文書化.....	25
2.7 小括.....	25
3. 試験実施計画書における estimand の記載.....	26
3.1 ICH E9(R1)ガイドライン上の方針.....	26
3.2 EIWG からの推奨事項.....	27
3.3 データサイエンス部会としての推奨事項.....	28
3.3.1 Estimand の設定に関する推奨事項.....	28
3.3.2 Estimand の記載に関する推奨事項.....	28
3.3.2.1 試験実施計画書「目的 (Objectives) 」の章.....	28
3.3.2.2 試験実施計画書「統計解析 (Statistical Analyses) 」の章.....	30
3.3.2.3 その他の章における記載.....	30
3.4 安全性評価の estimand の記載.....	31

3.5	Estimand 記載における留意事項.....	31
3.5.1	治験実施計画書に記載する estimand の程度.....	31
3.5.2	試験開始後の estimand の変更に関する対応.....	32
3.5.3	規制当局間で異なる estimand の記載.....	33
3.5.4	治験総括報告書で求められる中間事象の発現状況に関連する記載.....	33
3.6	小括.....	34
4.	仮想事例.....	35
4.1	プライマリーケア領域: 糖尿病の仮想事例.....	35
4.1.1	典型的な試験デザイン.....	35
4.1.2	Estimand, 中間事象及び対応するストラテジーの設定例.....	35
4.1.2.1	中間事象.....	35
4.1.2.2	Estimand.....	36
4.1.2.3	中間事象のストラテジー.....	37
4.1.3	治験実施計画書への記載.....	39
4.1.3.1	目的 (Objectives) の章.....	39
4.1.3.2	統計解析 (Statistical Analyses) の章.....	40
4.2	オンコロジー領域: 悪性リンパ腫の仮想事例.....	42
4.2.1	典型的な試験デザイン.....	42
4.2.2	Estimand, 中間事象及び対応するストラテジーの設定例.....	43
4.2.2.1	中間事象.....	43
4.2.2.2	Estimand.....	44
Part 2:	さらに理解を深めたい方へ.....	46
5.	Estimand に関する議論.....	46
5.1	治療の設定及び試験目的の決定.....	46
5.1.1	Tripartite Framework.....	46
5.1.2	関心のある治療の設定.....	47
5.1.3	試験目的の決定.....	48
5.1.4	主たる目的の検討.....	48
5.2	中間事象の特定.....	49
5.3	中間事象に対応するストラテジーの検討.....	49
5.3.1	ストラテジーの選択.....	49
5.3.2	治療方針 (Treatment policy) ストラテジー.....	50
5.3.3	仮想 (Hypothetical) ストラテジー.....	52
5.3.4	複合変数 (Composite Variable) ストラテジー.....	53
5.3.5	治療下 (While on treatment) ストラテジー.....	54
5.3.6	主要層 (Principal stratification) ストラテジー.....	55

5.3.7	ガイドラインに記載されていないストラテジー .....	56
5.4	Estimand の設定.....	57
5.5	試験デザイン及びデータ収集.....	58
5.6	仮定の特定及び感度分析の設定.....	58
5.7	欠測データ.....	59
5.8	試験の目的に応じた estimand .....	60
5.8.1	PK/PD 解析における estimand .....	60
5.8.2	早期相における estimand.....	61
5.8.3	非劣性／同等性試験における estimand.....	62
5.8.4	Per Protocol Set と estimand.....	62
5.8.5	Per-protocol 解析による治療効果に対応する estimand.....	63
5.8.6	安全性評価における estimand.....	64
5.9	パンデミックの影響評価.....	65
5.10	Estimand に言及している主な疾患ガイドライン.....	66
6.	オンコロジー領域における estimand の意義.....	67
6.1	はじめに.....	67
6.2	Estimand の要素.....	68
6.2.1	関心のある治療.....	68
6.2.2	対象集団.....	68
6.2.3	変数（評価項目） .....	69
6.2.4	集団レベルの変数要約.....	69
6.3	打ち切りと中間事象.....	69
6.4	治療変更.....	71
6.5	Estimand の記述について.....	74
6.6	小括.....	74
7.	まとめ.....	75
8.	参考文献.....	76

## 図の目次

図 2-1	中間事象有無別の“投与開始24週後の HbA1c” .....	11
図 2-2	仮想臨床試験の集団（各群20例） .....	11
図 2-3	中間事象の取り扱い方別の治療効果の違い .....	12
図 2-4	試験の目的に対する、推定の対象、推定の方法及び感度分析の整合化.....	13
図 2-5	中間事象に対して治療方針ストラテジーを適用した例.....	16
図 2-6	中間事象に対して仮想ストラテジーを適用した例.....	17
図 2-7	中間事象に対して複合変数ストラテジーを適用した例.....	18

図 2-8	中間事象に対して治療下ストラテジーを適用した例.....	19
図 2-9	中間事象に対して主要層ストラテジーを適用する例.....	20
図 4-1	試験デザイン概要（糖尿病の仮想事例） .....	35
図 4-2	各 estimand での Patients' Journey（糖尿病の仮想事例） .....	39
図 4-3	試験デザイン概要（悪性リンパ腫の仮想事例） .....	42
図 6-1	OS の評価が困難になる仮想事例 .....	71

### 表の目次

表 4-1	Primary estimand（糖尿病の仮想事例） .....	36
表 4-2	補足的解析①及び②の estimand（糖尿病の仮想事例） .....	37
表 4-3	中間事象とそのストラテジーのまとめ（糖尿病の仮想事例） .....	38
表 4-4	Primary estimand（悪性リンパ腫の仮想事例） .....	44
表 4-5	Secondary estimand（悪性リンパ腫の仮想事例） .....	45

## 1. はじめに

本報告書は臨床統計家のみならず、臨床的疑問や試験の目的を協議し共に治験実施計画書を作成するすべての臨床担当者（メディカルライターや臨床薬理担当者も含む）を対象に作成した。

2022年12月現在、ICH E9『Statistical Principles for Clinical Trials 臨床試験のための統計的原則』<sup>1)</sup>の補遺である ICH E9(R1)『Addendum on Estimands and Sensitivity Analysis in Clinical Trials to The Guideline on Statistical Principles for Clinical Trials 臨床試験のための統計的原則補遺 臨床試験における estimand と感度分析』（邦題はステップ2公開時のもの）<sup>2)</sup>は ICH 調和ガイドラインとして最終合意されている（ステップ4）。この ICH E9(R1)で定義・導入された「estimand」は、試験の目的と対応する試験デザインや解析方法を繋ぐフレームワークを提示し、ICH E8(R1)『General Considerations for Clinical Studies 臨床試験の一般指針』でも触れられており、今日では臨床研究の各分野で広まってきている。特に医薬品開発の分野では、関心のある臨床的疑問に応じて何を推定すべきかを明確にするため、estimand の利用に関する研究が進み、欧米を中心に実試験へ適用されている。

医薬品開発において重要視されるランダム化比較試験において、ランダム化によってベースライン時点の因子の治療群間での交絡の回避が期待できたとしても、治療の中止、レスキュー薬の使用、治療の切り替えといった治療開始後に起こる事象に起因するバイアスはランダム化では防ぐことができない。ICH E9(R1)の中で「中間事象」と呼ばれるこのような事象は治療効果の推定やその解釈を複雑にしてしまう。治療効果に影響する中間事象を考慮した上で、期待される治療効果を適切に説明する必要があるという考えの下で ICH E9(R1)の検討が始まった。ICH 改革と ICH ガイドライン解説<sup>3)</sup>では、次のとおり estimand を明確に定義する必要性が説明されている。

これまでも estimand という言葉自体は用いていなかったが、試験目的に合わせた試験デザインや解析計画を検討してきた。後述する中間事象も、中止例や欠測データの取り扱いとして一部は議論してきた。（中略）その一方で、『推定されるべきもの』を十分考慮することなく欠測データに対する解析方法を決定することで、推定される治療効果が曖昧であったり、推定される治療効果と試験目的との潜在的な不整合に繋がっていたりした恐れもあった。これを回避するためには、推定に影響する中間事象を考慮し、推定の対象とする治療効果、すなわち estimand を試験計画時に明示的に検討することが重要になる。

古くから、臨床研究では PICO/PECO といった概念があり、これは Patients/Intervention (Exposure)/Comparison/Outcome という4項目を用いて試験の意味合いを整理するために用いられていた。したがって estimand はこれまでに存在しなかった全くの新しい概念ではなく、既存のフレームワークを拡張したと捉えることも可能である。ただし、estimand の設定は解析手法やデータハンドリングの検討に先んじて行われるべきものであり、臨床試

験を計画する際、関心のある臨床的疑問を明確にしたうえで試験デザイン、データ収集、統計解析等を検討することが重要となる。

ICH E9(R1) Expert Working Group Step 4 Training Material (以降、Training Material)<sup>4)</sup> では、estimand を「試験目的によって提起される『臨床的疑問を反映した治療効果』の正確な叙述である」と説明している。適切な estimand を設定することで、試験の目的に関心のある重要な臨床的疑問に変換し、試験目的と対応する試験デザインや解析方法を繋ぐフレームワークを提示している。このフレームワークは治験依頼者における試験計画時及び結果解釈時だけでなく、規制当局のそれらのレビューにも役立つ。事前に規制当局と estimand を議論・合意することにより、治験依頼者と規制当局間の認識の乖離を防ぐことができるだろう。また、医薬品の承認審査において試験デザインの適切性や試験結果の解釈について議論する際に、estimand に基づいて協議することで治験依頼者と規制当局間の意思疎通がよりよいものとなることが期待される。これまで検証的試験で示すべき事項については規制当局と試験開始前に議論してきた。ICH E9(R1)で estimand のフレームワークを用いて、異なるステークホルダー間で関心のある臨床的疑問を議論するための方法が整理されたため、今後はこのフレームワークに沿って estimand と解析方法について規制当局と合意しておくことが望ましい。

以上の背景から、試験計画段階で臨床試験に関わるすべての関係者が estimand のフレームワークを用いて推定すべき治療効果に対し事前に合意し、共通の理解・認識を持った上で、それが正確に規定された文書により臨床試験を実施することは重要である。また、このフレームワークを用いることにより関心のある治療効果の定義に対する透明性が上がり、議論及び合意をスムーズに行えることも期待できる。そのため臨床試験においては、設定した estimand を治験実施計画書や統計解析計画書に明記することが求められてきている。TransCelerate が発行している Common Protocol Template<sup>5)</sup> でも、ICH E9(R1)と整合させるために、目的の章 (3. Objectives, Endpoints, and Estimands) 及び統計解析の章 (9. Statistical Considerations) に estimand が追記されている。また現在治験実施計画書テンプレートに関するガイドラインを ICH M11 CeSHarP (Clinical electronic Structured Harmonised Protocol) の Expert Working Group にて検討中であり、公開されているステップ2ドラフトテンプレート<sup>6)</sup> では、目的の章や統計解析の章に estimand や中間事象の取り扱い等に関する記載項目が設けられている。今後 ICH M11の検討過程で治験実施計画書への estimand の記載項目がさらに整理される可能性があるが、本報告書では現状入手可能な文献等を踏まえて治験実施計画書に estimand を記載する際の考慮事項を整理し、治験実施計画書への記載方針について提言を行う。

本報告書は2つのパートで構成している。Part 1では、臨床試験に携わるすべての方に向けて、estimand を解説した上で例示を用いながら治験実施計画書へ実装する方法を提示した。2章で ICH E9(R1)に記載されている estimand の概要及びその実装のプロセスに関して Training Material Module 3に記載されている“thinking process”を紹介する。3章にて治験実

施計画書における estimand の記載方針やいくつかの課題に対し提言を試みる。4章ではプライマリーケア領域とオンコロジー領域から1例ずつ仮想事例を挙げ、治験実施計画書への記載案を掲載する。Part 2は estimand に関してさらに理解を深めたい方向けに構成した。5章では関連する文献を紹介しながら考慮すべき事項を整理した。また試験目的に応じた estimand の紹介では、検証的な試験以外での estimand の設定についても言及した。加えて6章では ASA (American Statistical Association) science group 及び European special interest group “Estimands in oncology”, sponsored by PSI and EFSPI による Oncology estimand working group にて検討が進められているオンコロジー領域での estimand に関する話題についても要約した。

冒頭に記載したように本報告書は、臨床統計家に限らず広く臨床試験に関連する業務に従事する者を対象としたため、具体的な解析手法等、統計学的に高度な知識が求められる議論は省略した。なお、ICH E9(R1)をより深く理解するには、因果推論の知識が役に立つが、本報告書では触れず、因果推論を用いた estimand の解釈については、日本製薬工業協会医薬品評価委員会データサイエンス部会2020年度継続タスクフォース1による報告書「ICH E9(R1)の理解に役立つ因果推論」<sup>7)</sup> や2022年度継続タスクフォース4サブチームによる報告書「ICH E9(R1)の理解に役立つ因果推論～時間依存性治療編～」<sup>8)</sup> を参照いただきたい。

なお、本報告書では ICH E9(R1) や Training Material, 英語文献等に基づいて作成しているが、特筆しない限り和訳は筆者による参考訳であり、規制当局等による正式な和訳ではないことをご留意いただきたい。



# Part 1: 臨床試験に携わるすべての方へ

## 2. ICH E9(R1)と estimand

本章では ICH E9(R1)の概要を俯瞰する。ICH E9(R1)で鍵となるのは estimand であるが、estimand について理解するためには、それらを構成する試験における“治療効果”、“中間事象”という概念を理解することが必要である。

本章の構成として、まず治療効果及び中間事象の意味を説明した後、estimand のフレームワーク、estimand の要素、中間事象に対するストラテジーについて、主に ICH E9(R1)の内容、及び必要に応じて Training Material Module 2の内容を用いて説明を行う。

### 2.1 治療効果と中間事象の定義

医薬品は、有効性（治療効果）について統計的に検証され、更に治療効果の大きさ（ベネフィット）と治療による安全性（リスク）の評価を持って承認される。通常、**治療効果は臨床試験において推定されるものであり、治療効果を推定するための臨床試験の立案段階で、各規制当局とともに関心のある治療効果を明確にし、合意することが望まれている。** 実際、ICH E9(R1)でも“agreed estimand”という表現が用いられており、治験依頼者と規制当局間で合意された estimand の存在を前提としているであろう記載も見受けられる。治療効果の定量化として、試験治療後のアウトカムと、異なる治療条件下（例えば、試験治療を受けなかった場合、異なる治療を受けた場合）で同一被験者に起きたかもしれない潜在的なアウトカムとの比較を考えることが理想であろう。その比較のために、多くはランダム化比較試験を実施し、治療条件以外の条件（人口統計学特性やベースライン時の疾患状態、治療状況等）の分布を比較群間で確率的に同じにすることで、観察されたアウトカムの群間での対比（群間差や群間比等）を治療条件のみの違いによる結果として解釈している。しかし、中間事象と呼ばれる、治療開始後に発現する事象の取り扱いによって、結果及びその解釈が異なる場合が生じる。そのため、**関心のある治療効果を明確に定義するには、治療条件、評価指標、評価時点等だけでなく、中間事象の取り扱いも併せて決めることが重要となる。** 具体例とともに、その重要性を確認してみる。

ここでは、Training Material Module 2と同様に、2型糖尿病患者を対象としたプラセボ対照ランダム化比較試験を例として考えていく。治験実施計画書において、“投与開始24週後の HbA1c の群間差”のように評価時点や評価指標のみを用いて関心のある治療効果が記載されていることがあるが、これだけでは関心のある治療効果の“明確”な定義とは言えない。通常、臨床試験では、割り当てられた治療を規定通りに遵守できる患者もいれば、副作用等によって割り当てられた治療を継続できなくなる患者がいることもある。その

他、効果不十分のために併用薬の追加が必要となる患者が出てくる可能性もある。ICH E9(R1)は、中間事象を「治療開始後に発現し、関心のある臨床的疑問に関連する測定値の解釈又は存在に影響を及ぼす事象」と定義している。測定値とは、この事例では HbA1c が相当する。ここで、関心のある治療条件を「ベースライン時の治療+プラセボ又は試験治療」とすると、この例で取り上げた「副作用による治療中止」や「効果不十分による併用薬追加」といったものが、まさに中間事象に相当し、これらの取り扱いによって推定される治療効果が変わってくることを、確認していく。

まず、実際には不可能だが、ここでは同一の患者が試験治療を受けた場合とプラセボを受けた場合の結果を観察できたとしたときの、(A) いずれの治療でも規定通り完遂できる患者、(B) 試験治療で副作用が発現し治療継続できない患者、(C) 試験治療では効果不十分で併用薬追加が必要な患者、の3パターンの患者における“投与開始24週後の HbA1c”を [図 2-1](#)に示す。患者ごとにこれらのパターンを特定することも実際には不可能であることに注意頂きたい。論点の明確化のために、いずれの患者も治療開始時の Hb1Ac は8.0%とし、プラセボ群に割り付けられた時には (B) の患者では中間事象が発生せず投与開始24週後の HbA1は8.5%に悪化、(C) の患者では中間事象が発現し投与開始24週後の HbA1は8.5%に悪化するものとする。まず、(A) の24週時までいずれの治療でも治療継続可能な患者では、試験治療を受けると Hb1Ac が6.0%に低下したことから、この患者における治療効果（試験治療を受けたときとプラセボ治療を受けたときの投与開始24週後の HbA1の差）は - 2.5%であった。次に、(B) の治療継続できなくなる患者では、試験治療中止時点での Hb1Ac は7.5%で投与開始24週では8.0%であったことから、もし試験治療の中止有無によらず投与開始24週後の比較に関心があるのであれば、この患者における治療効果は - 0.5%となる。一方、24週間治療を継続できていたらという仮想的な状況での比較に関心があるのであれば、観測値を用いるのではなく、その仮想的な状況下での結果を推定する必要がある。ここでは推定方法の説明は割愛するが、その推定値が6.5%だったとすると、(B) の患者における24週間治療を継続したときの治療効果は - 2.0%となる。最後に、(C) 併用薬追加が必要となる患者では、併用薬を追加した時点での Hb1Ac は8.5%で投与開始24週では7.5%であったことから、もし併用薬の追加有無によらない投与開始24週後の比較に関心があるのであれば、この患者における治療効果は - 1.0%となる。一方、併用薬を追加しなかったらという仮想的な状況での比較に関心があるのであれば、同様にその仮想的な状況下での結果を推定する必要がある、その推定値が8.0%だったとすると、(C) の患者における併用薬を追加せずに24週間治療を継続したときの治療効果は - 0.5%となる。このように、(A) ~ (C) のそれぞれの患者で“投与開始24週後の HbA1c”が異なることは自然ではあるものの、特に (B) や (C) の患者においては中間事象の発現有無によらない投与開始24週後の Hb1Ac に関心があるのか、中間事象が発現しなかったらという仮想的な状況での投与開始24週後の Hb1Ac に関心があるのかによって治療効果が変わってしまうことが分かる。

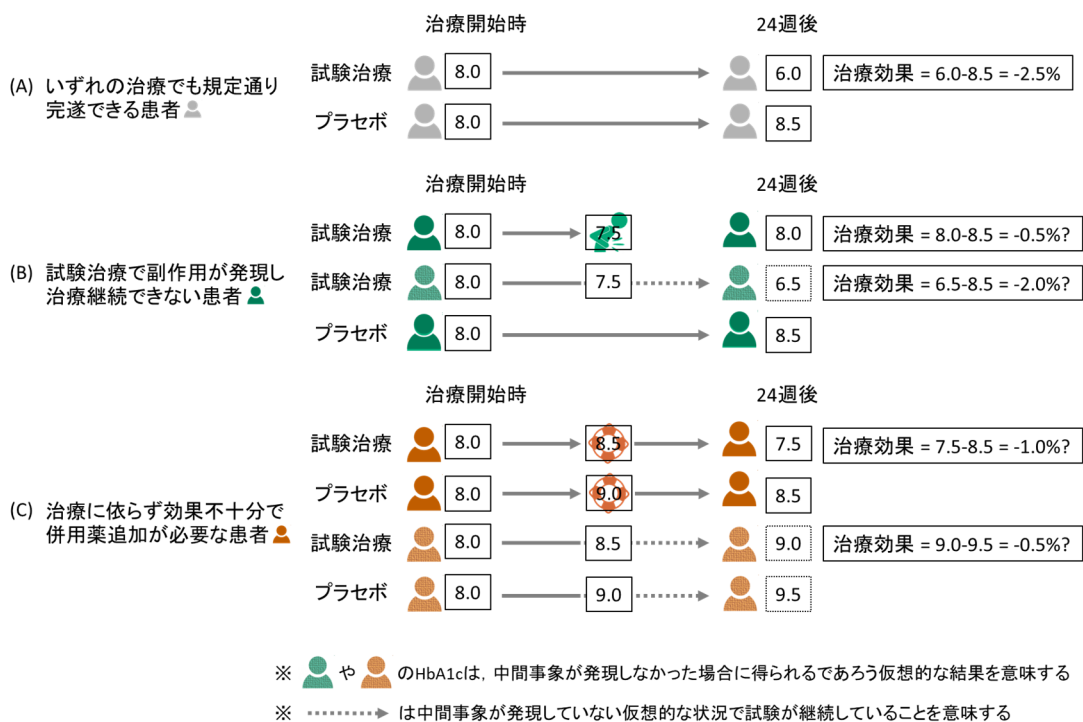


図 2-1 中間事象有無別の“投与開始24週後の HbA1c”

次に、通常の臨床試験のように集団として考えてみる。患者背景が群間で均一な各群20例の仮想臨床試験を考え、(A) 24週後まで試験治療の継続が可能な患者が各群11例、(B) 副作用が発現し試験治療を継続できなくなる患者が各群6例、(C) 試験治療中に併用薬追加が必要となる患者が各群3例の集団を考える (図 2-2)。

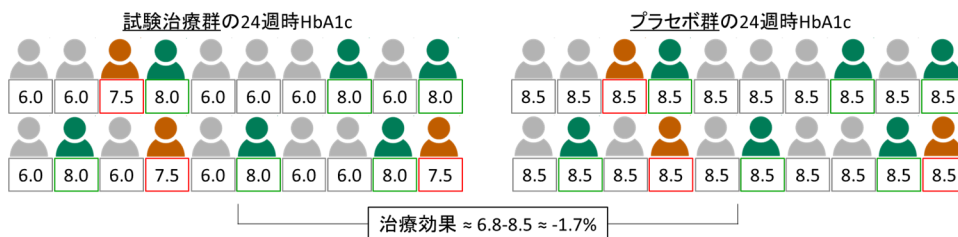


図 2-2 仮想臨床試験の集団 (各群20例)

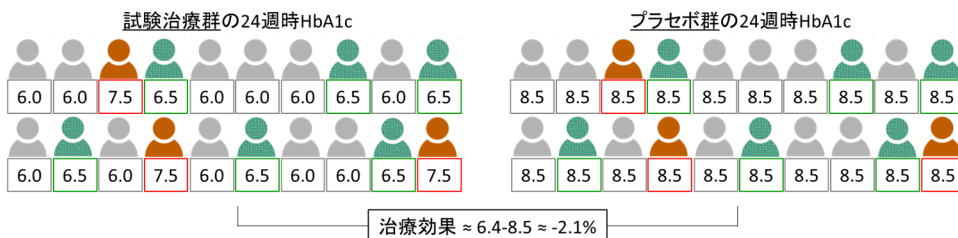
このとき、「副作用による治療中止」、「効果不十分による併用薬追加」の2つの中間事象に対して、3通りの取り扱い方による、“投与開始24週後の HbA1c の群間差”として推定される治療効果の違いを図 2-3に示した。この結果のように、単に“投与開始24週後の HbA1c の群間差”と記載しても、中間事象の取り扱いによって臨床試験で推定される治療効果は異なってくる。中間事象の取り扱い方はこの3つに限定されるものではなく、例えば、“治療遵守有無によらず併用薬の追加がなかったという仮想的状況下”等、その他

の中間事象の取り扱い方をしたときの治療効果に関心があるかもしれない。

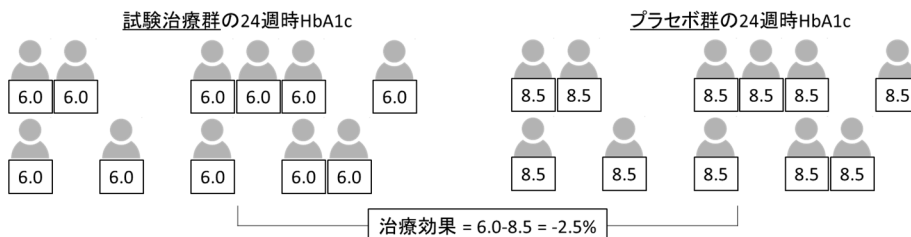
定義①: 治療遵守有無や併用薬追加有無を考慮しない, 投与開始24週後のHbA1cの群間差



定義②: 併用薬有無は考慮せず, 治療不遵守がなかったという仮想的状況下の, 投与開始24週後のHbA1cの群間差



定義③: 試験治療を不遵守なく併用薬追加なしで継続できる患者層※での, 投与開始24週後のHbA1cの群間差



※ 試験治療群は中間事象が発現しなかった患者のみ, プラセボ群は仮に試験治療を受けた時に中間事象が発現しなかったであろう患者のみを対象としており, 観察された中間事象発現有無に基づく部分集団ではないことに注意が必要。なお, プラセボ群における仮に試験治療を受けた時に中間事象が発現しなかったであろう患者層は実際には直接特定できないものである。

図 2-3 中間事象の取り扱い方別の治療効果の違い

以上のことから, 関心のある治療効果を定義する際には, 中間事象の取り扱いの定義も重要であることは明らかであり, ICH E9(R1)では estimand と呼ばれる, 試験の目的と対応する試験デザインや解析方法を繋ぐフレームワークを提示しており, 中間事象の取り扱いもその中での検討事項となっている。

## 2.2 Estimand のフレームワーク

ICH E9(R1)では, 試験計画は図 2-4 に示す手順で策定することを求めており, 明確な試験の目的を, 適切な estimand を設定することにより, 関心のある重要な臨床的疑問に変換すべきとしている。そのため, まずは2.3節で後述する estimand を設定し, 次に定義した estimand に対して適切な推定方法, すなわち主とする推定量 (解析手法) を選択する。最後に, 主とする推定量より得られた推定値について, 解析手法の仮定からの逸脱に対す

る頑健性を探索するために、主とする推定量と同一の estimand を対象として1つ以上の感度分析を実施する。

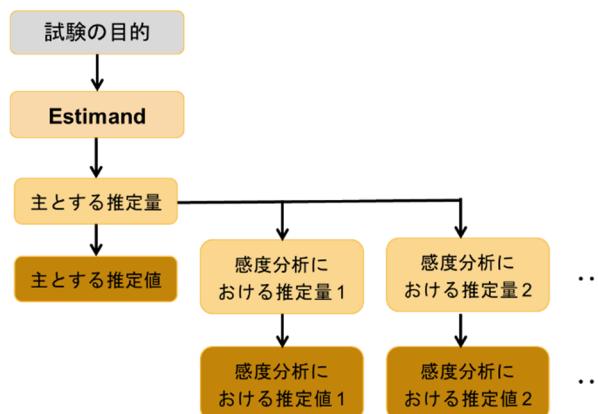


図 2-4 試験の目的に対する、推定の対象、推定の方法及び感度分析の整合化

このフレームワークにより、推定の対象（試験の目的、estimand）、推定方法（推定量）、数値結果（推定値）、及び感度分析を明確に区別する適切な試験計画の立案が可能となる。これは、治験依頼者による試験計画時及び結果解釈時だけでなく、規制当局のレビューの際も有用と考えられる。特に、医薬品の承認審査において試験デザインの適切性や試験結果の解釈について議論する際、治験依頼者と規制当局間のコミュニケーションをより円滑にすることが期待されている。

## 2.3 Estimand の5つの要素

Estimand は以下の5つの要素 (attributes) から構成され、関心のある治療の効果を定義することになる。

- **関心のある治療 (treatment)**

関心のある治療条件 (treatment condition) は、**試験において患者が受ける治療 (介入)** を指す。また、比較試験の場合は比較の対象となる治療条件も含まれる。関心のある治療条件は異なる介入が個別に行われる場合や同時に行われる複数の介入の組み合わせ (例えば、標準治療への上乗せ、時点ごとの複雑な介入を包含した一連の治療レジメン) となる場合がある。治験実施計画書で許容されている併用薬変更やレスキュー薬使用、治療の変更等の中間事象も含めて治療レジメンとして定義することで、被験者が受けると想定される治験薬以外の治療についても関心のある治療条件を定義する際に考慮することができる。なお、以降では明確に区別が必要な場合を除き治療条件は単に治療と記載する。

- **対象集団 (population)**



臨床的疑問に答える際に、対象となる患者集団を指す。代表的なものとしては、試験に登録された全体集団、ベースライン時に測定された特性によって定義される部分集団、特定の中間事象の発現可能性によって定義される集団が挙げられる。

- **変数（エンドポイント）（variable/endpoint）**

変数とは、試験において臨床的な疑問に答えるにあたり必要となる患者ごとに得られる測定値や評価尺度の値を指す。また、解析時点を含んだ評価項目として定義されることもある。また、特定の中間事象の発現状況が含まれることがある（例えば、治療の成功又は失敗という2値変数に対して、試験治療中止を治療失敗とみなす等）。

- **その他の中間事象（other intercurrent events）**

中間事象とは治療開始後に発現し、関心のある臨床的疑問に関連した変数の測定を不可能とする事象や変数を試験治療の効果として解釈する際に影響を与える事象を指す。関心のある変数の存在に影響を与える中間事象は、死亡や下肢切断等の終末事象（ただし糖尿病壊疽の症状評価時等、終末事象自体が主要評価項目あるいはその構成要素ではない場合）が代表例である。これらは被験者の状態の変化に起因するものであり、中間事象発現後のデータは存在しない又は意味がないことが多い。一方、関心のある変数の解釈に影響を与える中間事象は、主に治療の中止、代替治療の開始、レスキュー薬の開始、治療法の変更等が挙げられる。関心のある治療効果を正確に定義するため、起こりうる中間事象を事前に特定し、それらをどのように考慮するかを規定しておく必要がある。

また、中間事象発現後の存在しない又は意味がないデータと欠測データ（規定した estimand の解析に対して意味があると考えられるが、収集されなかったデータ）は明確に区別される。欠測値との違いという観点では、欠測値はできる限り発生しないよう対処する必要がある一方、中間事象は必ずしも問題と捉えるべきではないと言える。中間事象は例えば実臨床でも同じように起こる事象であったり、倫理的観点から設定している場合（治験の中止、代替治療等）があったりし、これらに対し発現を防ごうと対処するのではなく、設定した中間事象の発現の影響も考慮した上で関心のある治療効果を定義すべきものである。例えば、関心のある治療効果を求めるために中間事象発現以降のデータが必要な状況で、それらのデータが観察されていなければ、それは欠測値の問題として解析で処理される。

関心のある治療効果を正確に定義しようと、関心のある治療、対象集団、及び変数の定義について議論をする中で、多くの中間事象はそれらの要素の中で取り扱われるであろう。それらの要素で取り扱われなかった「その他の中間事象」については、2.4節で紹介するようなストラテジーを規定することによって関心のある治療効果を反映する必要がある。

- **集団レベルの変数要約 (population-level summary)**

試験において治療を比較するための指標（どのように比較するか）であり、**集団レベル**での**変数の要約方法**（例えば、平均値の群間差）を明示するものである。

## 2.4 中間事象を取り扱うストラテジー

前述の通り、中間事象が発現すると変数の測定が不可能になったり、治療効果の解釈に影響を与えたりする。したがって関心のある臨床的疑問を表現する estimand を設定する際には中間事象に対して何らかの対処が必要となり、その方針をストラテジーと呼ぶ。ICH E9(R1)には例として5つのストラテジーが明示されている。Estimand を設定する際には、関心のある臨床的疑問を反映できるようにストラテジーを定めることになる。なお、1つの estimand に複数の中間事象が存在する場合、中間事象ごとにストラテジーを検討する必要がある。また、試験で複数の estimand を設定する場合に中間事象に対するストラテジーを検討した結果、estimand の治療、変数、対象集団の要素に含まれず「その他の中間事象」として明記される中間事象も、各 estimand によって異なる可能性がある。

ICH E9(R1)において明示されたストラテジーを以下に記載する。事例は Training Material から引用した。

### 2.4.1 治療方針 (Treatment Policy) ストラテジー

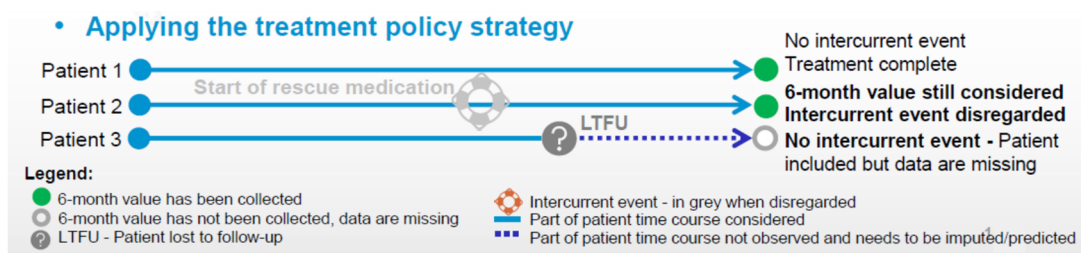
治療方針ストラテジーを適用する場合には、関心のある治療効果の定義において中間事象の発現の有無を考慮しない。すなわち、関心のある変数の値は、中間事象が発現したか否かに関わらず解析に使用される。例えば、レスキュー薬の使用を中間事象とし、これに治療方針ストラテジーを適用する場合、レスキュー薬の使用後の変数の値も解析に含める。

治療条件に関わる中間事象（例えば、治療の中止、レスキュー薬や併用薬による治療の開始・変更等）について治療方針ストラテジーを適用する際、中間事象を治療条件の一部とみなすことができる。これは、estimand の要素「関心のある治療」においてその中間事象を含めて定義した場合と同義であり、ICH E9で定義されている Intention-to-Treat (ITT) 下での治療方針を反映している。なお、一般に、終末事象を中間事象とする場合は、中間事象発現後の変数の値が存在しないため、治療方針ストラテジーを用いて解析することはできない。

以下の事例では、中間事象としてレスキュー薬の使用を設定して治療方針ストラテジーを適用した場合を考える（[図 2-5](#)）。

Patient 1はレスキュー薬を使用していない（中間事象が発現していない）ため、評価時点の値を変数の値とする。Patient 2は治験期間中にレスキュー薬が使用され（中間事象が発現し）、その後もデータ収集を継続し、評価時点の値を変数の値とする。Patient 3はレスキュー薬を使用していない（中間事象が発現していない）が lost to follow-up となった

ため、評価時点での変数の値は欠測値として扱われる。



引用元: Training Material Module 2.3 p.102

図 2-5 中間事象に対して治療方針ストラテジーを適用した例

## 2.4.2 仮想 (Hypothetical) ストラテジー

中間事象に対して仮想ストラテジーを適用する場合、もしその中間事象が起きなかったら、という仮想的な状況を想定して治療効果を推定する。仮想的な状況下での試験治療の効果を考慮することは、臨床的にも規制上にも意義がある場合がある。ただし、仮想シナリオをいくつも想定できたとしても、規制上の意思決定において関心のあるシナリオは限定的であろう。例えば、試験に参加する患者の安全を確保するためには、レスキュー薬の使用を制限することは倫理的に認められないことがある。このとき、レスキュー薬を使用しなかった場合の治療効果は企業側の関心の対象となることが多い一方、レスキュー薬が利用可能でないために使用できない状況下での治療効果は規制上の関心の対象となることもある。また想定する状況が実臨床で起こることが稀な場合は、临床上及び規制上ともにその仮想的な状況下での治療効果への関心は限定的である。そのようなシナリオに関心がある場合には、関心をもつ理由や、実臨床で想定される状況で得られる結果との差異を明確にする必要がある。

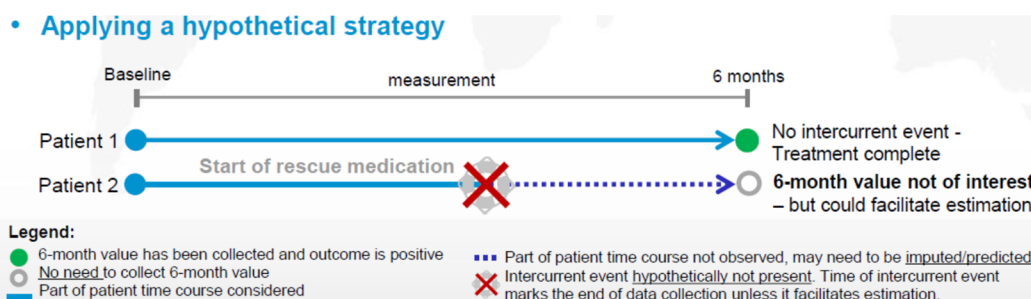
仮想ストラテジーの適用を考える場合には、仮想的なシナリオにおける具体的な想定及びそのシナリオに臨床的な関心がある理由を明確にすべきである。例えば「患者がレスキュー薬を使用しない状況」という仮想的なシナリオの記載は、「レスキュー薬が利用可能でないために使用できない状況」なのか「特定の患者がレスキュー薬を必要としない状況」なのか複数の解釈が可能で、関係者間での誤解を招く可能性があるため、仮想的なシナリオの想定は明確にすべきである。

図 2-6では、レスキュー薬を使用するという中間事象に対して、仮想ストラテジーとして「レスキュー薬を使用しないシナリオ (特定の患者がレスキュー薬を必要としない状況)」を適用した場合を考える。ここでは中間事象発現前までのデータに対して何らかの統計モデルを用いて、このシナリオの下での治療効果を推定する。仮想ストラテジーを適用する場合は、関心のあるデータが仮想的なシナリオの下で得られたであろう結果であることから、中間事象発現後の実際に観測された値は関心のあるデータではなく、解析に含



めずに中間事象が起こらなかった場合の治療効果を推定する必要がある。図2-6の Patient 1はレスキュー薬を使用しなかった（中間事象が発現しなかった）ため、評価時点の値を変数の値とする。Patient 2では治験期間中にレスキュー薬を使用した（中間事象が発現した）ため、中間事象発現前までのデータを使用する。なお、仮想ストラテジーの適用には、中間事象によって観測できなかったデータを補完する際、補完に用いるモデルやデータの限界から、得られる推定値が不安定となる場合があり、信頼における推定値を得ることができているか、という観点からの検討も必要である。

なお、特定の中間事象に対して仮想ストラテジーを適用することによって、その中間事象が発現した後のデータの収集は必ずしも不要と結論づけられないことに注意が必要である。仮想ストラテジーとして想定されるシナリオや（別の）estimandに基づいて設定する解析手法によって、中間事象発現後のデータも利用して関心のある変数の値を推定することができる。一例として、6章に掲載したオンコロジー領域における全生存時間を調整した解析手法が考えられる（詳細は6.4節に示した RPSFT モデルを参照の事）。



引用元: Training Material Module 2.3 p.105

図 2-6 中間事象に対して仮想ストラテジーを適用した例

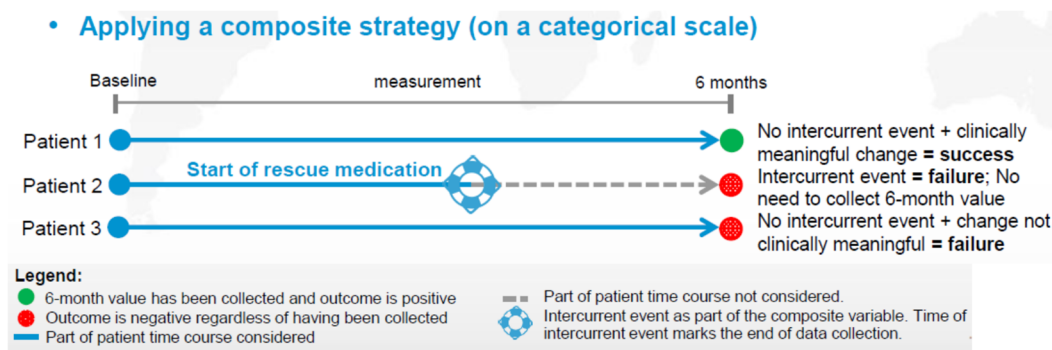
### 2.4.3 複合変数（Composite Variable）ストラテジー

複合変数ストラテジーは、中間事象の発現自体を「変数（エンドポイント）」の一部に組み込むことで対応するストラテジーである。複合変数ストラテジーが適用される中間事象は、その発現が臨床的に試験治療の失敗に相当すると解釈できるような事象であり、また変数（エンドポイント）も治療の成否自体であることが多い。例えば、毒性のために治療を中止した患者は、その治療が失敗したと考えると、その患者の変数は失敗と取り扱う。また致命的な疾患に対して試験を実施する際、実際に患者が死亡した場合に患者の状態に関する測定値は存在しないか、意味がない値になることがある。この時に中間事象として患者の死亡を設定し複合変数ストラテジーを用いることで、中間事象の発現に伴う患者の状態を変数として意味があるものとして記述することができる。なお複合変数ストラテジーを用いる場合、変数は二値変数（例：治療の成功・失敗）に限定されることはない。例えば、身体機能を測定する試験における変数が連続尺度による値の場合、死亡した被験者

は身体機能の完全な欠如を意味する値（例えば連続尺度での最悪値）を代入する等の対応も考えられる。

死亡に代表される終末事象が中間事象となる状況は、複合変数ストラテジーが適用される最たる例と考えられる。試験治療によって生存期間の延長が期待できるならば、多くの場合、生存者に対する様々な測定値の変数だけでなく、それらの変数に加えて生存しているという事実も関心のある変数となるだろう。複合変数ストラテジーにより中間事象としての死亡を変数に組み込んだ例として、がんの臨床試験における無増悪生存期間（Progression Free Survival: PFS）がある。これは死亡という中間事象を、画像判定による腫瘍の増悪というイベントと複合した変数として扱い、死亡までの時間と増悪までの時間を複合させた治療効果としていずれか早い時点を変数の値としている。

図 2-7ではレスキュー薬の使用という中間事象に対して複合変数ストラテジーを適用している。変数（エンドポイント）は治療の成功・失敗の二値変数であり、レスキュー薬の使用（中間事象の発現）は治療の失敗として取り扱う。Patient 1はレスキュー薬を使用せず（中間事象が発現していない）、かつ臨床的に意義のある変化が認められたため、変数の値は治療成功となる。Patient 2はレスキュー薬を使用した（中間事象が発現した）ため、変数の値は治療失敗となり、Patient 2の中間事象発現後のデータを収集する必要はない。Patient 3はレスキュー薬を使用しなかった（中間事象が発現しなかった）が、臨床的に意義のある変化が認められなかったため、変数の値は治療失敗となる。



引用元: Training Material Module 2.3 p.107

図 2-7 中間事象に対して複合変数ストラテジーを適用した例

#### 2.4.4 治療下（While-on-Treatment）ストラテジー

治療下ストラテジーは中間事象の発現前までの治療に対する反応に関心がある場合に用いられる。例えば患者が生存している期間における反応が関心の対象となり、死亡を中間事象として取り扱うと、治療下ストラテジーを用いて中間事象に対処していることとなる。

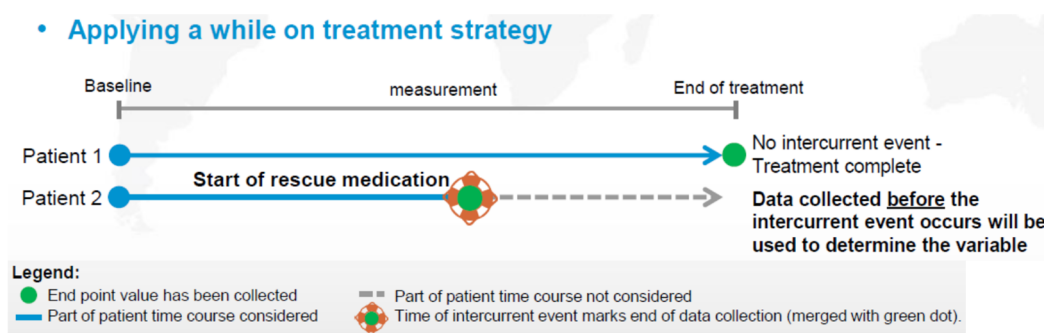
同じ変数が繰り返し測定される場合、全被験者共通の規定されたすべての時点における

測定値ではなく、中間事象発現までの測定値に臨床的な関心があるときにこのストラテジーが適用される。

中間事象発現時までの関心の対象が二値変数であっても同様である。例えば、末期の疾患を有する患者を対象とした対症療法による緩和効果に関心がある場合、治療の成否は、患者が死亡するまでの症状に対する緩和効果を評価した変数の値に関心があり、患者の生死自体は関心の対象ではない。つまり、患者の死亡を中間事象とすると、中間事象発現までの変数の値を用いて評価する治療下ストラテジーを適用しているとみなすことができる。あるいは、治療計画を遵守している間の薬剤の副作用のリスクを評価することに関心がある場合、試験治療の中止は中間事象であり、治療下ストラテジーを適用して対処するといえる。

治療下ストラテジーは、関心のある期間を中間事象発現前に限定しており、これは複合変数ストラテジーと同様に変数の定義に影響を与えると考えることもできる。ただし、複合ストラテジーでは主に評価指標そのものに影響がある一方、治療下ストラテジーでは評価期間（評価時点）に影響があると言える。また、比較する治療群間で中間事象の発現頻度や時期が異なる場合には、解釈に注意が必要となる。例えば、先に示した対症療法による緩和効果に関心がある場合で、治療群間で生存時間が大きく異なるようなときには、治療下ストラテジー適用時の結果のみならず、生存時間への影響も考察が必要かもしれない。

図 2-8 では中間事象をレスキュー薬使用として治療下ストラテジーを適用した治療効果を示す。Patient 1 はレスキュー薬を使用しなかった（中間事象が発現しなかった）ため、最終観察時点までの変数の値を利用する。Patient 2 はレスキュー薬を使用した（中間事象が発現した）ため、レスキュー薬使用時点までの値を用いて変数の値を決定する事になる。



引用元: Training Material Module 2.3 p.111

図 2-8 中間事象に対して治療下ストラテジーを適用した例

#### 2.4.5 主要層（Principal Stratum）ストラテジー

主要層ストラテジーは estimand のうち対象集団の定義に影響を与えるストラテジーである。主要層ストラテジーは、規定された中間事象によって定義される特定の患者層の治

療効果に関心がある場合に用いる。

図 2-9 では中間事象のレスキュー薬使用に対して主要層ストラテジーを適用している。この例では実際に割付けられた治療群に関わらず、レスキュー薬を使用しない患者層 ( $S_{11}$ ) に対する治療効果に関心がある。この場合、 $S_{11}$  を主要層とし、 $S_{11}$  に層別された患者の変数の値から治療効果を推定する。なお、並行群間試験においてはどちらか一方の治療群における中間事象の発現の有無しか観察できず、被験者がどの層に属するかを観測できないため、推定には強い仮定が必要である。

		Treatment	
		Rescue medication	No rescue medication
Placebo	Rescue medication	$S_{00}$	$S_{01}$
	No rescue medication	$S_{10}$	$S_{11}$

引用元: Training Material Module 2.3 p.114

図 2-9 中間事象に対して主要層ストラテジーを適用する例

( $S_{00}$ : 試験治療群と対照群のどちらに割付けられたとしてもレスキュー薬を使用する患者層,  $S_{01}$ : 対照群に割付けられた時にレスキュー薬を使用し、試験治療群に割付けられた時にレスキュー薬を使用しない患者層,  $S_{10}$ : 試験治療群に割付けられた時にレスキュー薬を使用し、対照群に割付けられた時にレスキュー薬を使用しない患者層,  $S_{11}$ : 試験治療群と対照群のどちらに割付けられたとしてもレスキュー薬を使用しない患者層)

主要層は、治療介入ごとの中間事象の潜在的な発現の有無に基づく層別（例えば、被験薬に割付けられた場合には中間事象が発現し、対照薬に割付けられた場合には中間事象が発現しない患者層）であり、実際の中間事象の発現の有無に基づく部分集団（例えば、実際に中間事象が発現した患者集団）ではないことに注意が必要である。主要層と部分集団は定義上異なる集団であり、実際に発現した中間事象の有無に基づく部分集団で解析したとしても、主要層における関心のある推定値とは異なる可能性がある。

ここで、主要層と部分集団の違いを説明するために、部分集団として中間事象が発現した患者の集団を考える。この集団には、試験治療群、対照群いずれに割付けられても（被験薬、対照薬いずれの介入を受けても）中間事象が発現する患者と、試験治療群に割付けられたことで（被験薬の介入を受けると）中間事象が発現する（対照薬の介入を受けても中間事象が発現しない）患者と、対照群に割付けられたことで（対照薬の介入を受けると）中間事象が発現する（被験薬の介入を受けても中間事象が発現しない）患者という3つの異なる背景をもつ（異なる主要層に属する）患者から構成される。他方主要層では、こ

の3つの異なる背景を持つ患者は別々の集団（主要層）として考える。すなわち、いずれの介入をうけても中間事象が発現する集団、対照薬の介入では中間事象が発現せず被験薬の介入でのみ中間事象が発現する集団、被験薬の介入では中間事象が発現せず対照薬の介入でのみ中間事象が発現する集団にそれぞれ属することになる。

上述のとおり、主要層は部分集団とは異なり、例えば、主要層解析を目的としている際に、中間事象が実際に発現したか否かで患者を分類して検討してしまうと、興味関心のある対象集団である主要層とは異なる集団となるため、関心のある治療効果の推定に影響を与える可能性がある。

このストラテジーが適用される状況では、関心のある臨床的疑問は特定の主要層内での試験治療の効果となる。例えば、感染症に対するワクチン接種を試験治療とした比較では、ワクチンの接種有無によらず感染する患者層（実際にワクチンを接種し感染した部分集団又は実際にワクチンを接種せずに感染した部分集団とは異なる）に対し、ワクチンによる重症化抑制効果に関心があることがある。これは感染を中間事象としたときに、治療介入によらず中間事象が発現する患者層に対して主要層ストラテジーを適用したと考えることができる。別の例として、複数回治験薬を投与する試験で、中間事象を毒性による治療中止と規定し、主要層を治療介入によらず試験投与に耐えられる患者層とした主要層ストラテジーによる治療効果に関心があることも考えられる。

## 2.5 感度分析と補足的解析

ICH E9では「安定性（ロバストネス）とは、データ、仮定及び解析方針についての様々な制限に対して全体の結論がどの程度変わり易いかに関連した概念」とされている。ICH E9(R1)では、この安定性の概念について拡張した議論をしており、**選択した特定の解析での仮定に対する推測の感度と、より広い意味での解析手法の選択に対する感度を区別している**。前者を感度分析、後者を補足的解析としてまとめているので以下に要約する。

分析の種類	Primary estimand との関係	例
感度分析	同一 estimand	解析手法における仮定の変更
補足的解析	異なる estimand	「変数」の変更: 連続変数から二値変数への変更

### 2.5.1 感度分析（Sensitivity Analysis）

ICH E9(R1) 用語集においては、**感度分析とは「主とする推定量の、モデル化における仮定からのずれとデータの限界に対する推測の安定性を調べるために実施される一連の解析」とある**。特定の estimand に基づく推測は、データの限界と、主とする推定量に対する統計モデルで用いる仮定からのずれに対して安定しているべきであり、この安定性は感度分析によって評価される。感度分析は、試験で定義されるすべての primary estimand に対して計画すべきであり、治験依頼者と規制当局の間で議論し事前に合意する事項となりうる。



## 2.5.2 補足的解析 (Supplementary Analysis)

ICH E9(R1)では、主とする推定量及び感度分析に加えて、試験データをより十分に評価し追加的な考察を得るために実施する解析を「補足的解析」と呼ぶ。感度分析の定義より、主とする推定量を導く **estimand** と異なる **estimand** を用いて実施される解析は感度分析ではなく補足的解析となりうる。ICH E9で定義された安定性を確認するための解析として従来から「感度分析」と呼称された解析には、試験の主目的に紐づく **primary estimand** とは異なる **estimand** 下で実施され、補足的解析と解釈すべきものが含まれていることには注意すべきである。Primary estimand が治験依頼者と規制当局の間で合意され、主とする推定量が事前に明確に規定され、主とする推定値の安定性が感度分析により確認できれば、一般的に補足的解析は感度分析よりも試験結果の解釈に用いる結果としては優先度が低い位置づけとなる。ただし、規制当局とのコミュニケーションにおいて解釈の優先度が常に低く扱われるとは限らないため、補足的解析の必要性和有用性は、各試験で考慮されるべきである。補足的解析の具体例は、4章を参照されたい。

## 2.6 Estimand 実装のためのプロセス

これまで述べてきた **estimand** のフレームワークを踏まえ、本節では臨床試験における **estimand** 実装のためのプロセスとして、ICH E9(R1)及び Training Material Module 3の Generic Example p.197-237に記載された “**thinking process**” を紹介する。

Thinking process は下記の7つのステップから構成される。

- 1) 疾患及び関心のある治療の設定に基づく試験目的の決定
- 2) 中間事象の特定
- 3) 中間事象に対応するストラテジーの検討
- 4) **estimand** の設定
- 5) **estimand** に沿う試験デザイン、データ収集、解析手法の選択
- 6) 主解析における仮定の特定及び仮定を検証するための適切な感度分析の設定
- 7) **estimand** の文書化

以下の節では、最終ステップ7の「**estimand** の文書化」を目指し、ステップ1~7の各段階において検討すべきこと・留意すべきことを要約する。

### 2.6.1 ステップ1: 疾患と関心のある治療の設定に基づく試験目的の決定

Training Material の **thinking process** では、対象とする疾患と関心のある治療を設定し、試験目的を決定することが最初のステップとして定められている。

この段階では疾患領域に関する治療の臨床的な意味合いを理解することが必要となる。

これには、疾病、試験で対象とする患者状態、治療の臨床的な位置づけも含まれる。診断、治療、予防ではそれぞれ関心のある治療レジメンが異なるため、異なる試験目的が設定される。これらをもとに estimand の要素を設定することになる。

関心のある治療を定義すると、estimand の対象集団や変数が検討可能となる。また治療のレジメンとして割付けられる治療や基礎治療、併用薬・レスキュー薬等を詳細に規定することが中間事象を特定する一助となる。

## 2.6.2 ステップ2: 中間事象の特定

ICHE9(R1)において中間事象は、治療開始後に発現し、関心のある変数の解釈又は存在に影響を与える事象と定義される。また、ICHE9(R1)では関心のある臨床的疑問に対する叙述において、推定される治療効果を正確に定義するため、中間事象をあらかじめ特定することを求めている。中間事象の検討には、試験における疾患領域との関連性や estimand の要素の1つである関心のある治療への正確な理解が必要になる。

Training Material Module 3には、中間事象には一般的な中間事象とその他の臨床事象から特定される中間事象の2種類があるとされている。

一般的な中間事象は、測定項目の解釈に影響を与えるものであり、例えば以下の事象が想定される。なお、頻繁に発生する中間事象だけでなく、試験の結果解釈に影響を与えるそれぞれの中間事象に対処する必要があると記載されている。

- 治療中止（有効性の欠如、毒性による中止等、理由によって複数の異なる中間事象として区別される場合がある）
- 追加又は代替治療の使用（基礎療法又は併用療法の変更、治療のスイッチ等、治療の種類や理由により区別される場合がある。また試験においては、これらはレスキュー治療又は禁止治療と表示することもある。）
- 死亡及びその他の終末事象

その他の臨床事象としては、例えば運動機能を測定する試験における脳卒中が挙げられる。これらは測定項目の評価や推定値の解釈に影響を与える場合に中間事象となる。

## 2.6.3 ステップ3: 中間事象に対応するストラテジーの検討

2.6.2節で特定された中間事象に対応するストラテジーを検討する。また、その中間事象を estimand のどの要素を用いて考慮するか検討する。

関心のある臨床的疑問を反映するストラテジーを選択する。ストラテジーの選択は、試験の目的に対応させるべきであり、(主とする)推定量に基づいて選択するものではない。検討する際のポイントは以下のとおりである。

- 臨床に関連する意思決定ができる推定すべきものは何か
- 信頼できる推定値が得られるか
- 各中間事象に対するストラテジーの選択は、治験依頼者のチーム内、規制当局内、及び治験依頼者と規制当局間等、立場の異なるステークホルダーでの議論の主題とすべきである

なお、Training Material Module 3では、例として収縮期血圧（SBP）への慢性治療の効果を定量化する目的の試験において、治療中止という中間事象を5つのストラテジーそれぞれで取り扱った場合、どのような意味をもつか検討されている。

#### 2.6.4 ステップ4: estimand の設定

本ステップは例示のみとなっている。例示された内容では、特定された中間事象とそのストラテジーを、estimand のどの要素に含めるか整理した後、記載方法として estimand の5つの要素を明示的に記載する方法と、一つの文章として記載する方法を示しているので引用する。

< Estimand の5つの要素を明示的に記載する方法 >

- **関心のある治療 (treatment):** 追加又は代替治療を6か月後まで使用しなかったとし、被験者が継続したかどうかに関わらない Drug X の治療条件、もしくは、追加又は代替治療を6か月後まで使用しなかったとし、被験者が継続したかどうかに関わらないプラセボの治療条件
- **対象集団 (population):** コントロール不能な高血圧患者
- **変数 (variable):** SBP のベースラインから6か月後までの変化量
- **その他の中間事象 (other intercurrent events):** 該当なし（試験治療の中止や治療変更といった中間事象は治療方針ストラテジーとして治療の要素に含まれているため）
- **集団レベルの変数要約 (population-level summary):** 治療条件間での平均値の差

< Estimand を一つの文章で記載する方法 >

コントロール不能な高血圧患者における、SBP のベースラインから6か月後までの変化量について、治療条件（追加又は代替治療を使用しなかったとした、治療中止に関わらない Drug X と治療中止に関わらないプラセボ）間での平均値の差

#### 2.6.5 ステップ5: 試験デザイン、データ収集、解析手法の選択

関心のある臨床的疑問を反映するためのストラテジーの選択に加えて、適切な試験デザイン及び統計解析が実施可能であることを判断する必要がある。特に以下の内容が述べられている。



- 収集する必要のあるデータを検討する
  - － 中間事象発現後のデータ収集が困難であることは、必ずしも estimand そのものや中間事象に対するストラテジーを変更する強い根拠とはならない
- 意思決定において信頼できる治療効果の推定値が得られるかを検討する

推定値の妥当性を示すことが難しい仮定に基づく解析手法を用いているが臨床的関連性は高い estimand と、より正当化できる仮定に基づく解析手法を用いているが臨床的関連性は低い estimand の間で、バランスを取る必要があるかもしれない。また、これは感度分析によって結果の信頼性をどの程度担保できるかにも依存する。

### 2.6.6 ステップ6: 主解析における仮定の特定及び適切な感度分析の設定

本ステップは感度分析と補足的解析の定義が示されている。すなわち感度分析は、主とする推定値を導出する際の仮定に関する検討を目的とした解析で、補足的解析は、それ以外の追加的な解析が該当すると記載されている。主とする推定量（主解析）における仮定を明示し、その仮定と異なる仮定を置いて推定の頑健性を探索することが一連の感度分析の目的となる。

### 2.6.7 ステップ7: estimand の文書化

Estimand を設定した上で、明示的に治験実施計画書に規定する必要があるとしている。また estimand の設定は対象集団、変数、試験デザイン、データ収集及び解析計画に影響するため、テクニカルな内容が別の場所に記載されていたとしても estimand は試験の目的と合わせて規定することが推奨されている。計画段階では予見していなかったが、試験途中で中間事象が特定され estimand を修正する必要がある場合には、治験実施計画書の改訂を計画すべきとされている。

また総括報告書では、estimand を参照しながら主解析・感度解析・補足的解析の結果を体系的に提示し議論する必要があるとされている。あわせて各解析が事前規定か否か、中間事象の発現回数と時期、感度解析の結果に基づいた推定の頑健性等を報告すべきと記載されている。

## 2.7 小括

本章では ICH E9(R1)及び Training Material に基づいて estimand とその要素、中間事象とそれに対応するストラテジー、及び感度分析ならびに補足的解析について概観した。また、estimand 実装のためのプロセスとして Training Material Module 3 “thinking process”を紹介した。

### 3. 治験実施計画書における estimand の記載

本章では治験実施計画書における estimand の記載についてデータサイエンス部会としての推奨事項を提言する。

#### 3.1 ICH E9(R1)ガイドライン上の方針

ICH E9(R1)の A.6（一部抜粋）では、治験実施計画書上で記載すべき estimand について次のように示されている。

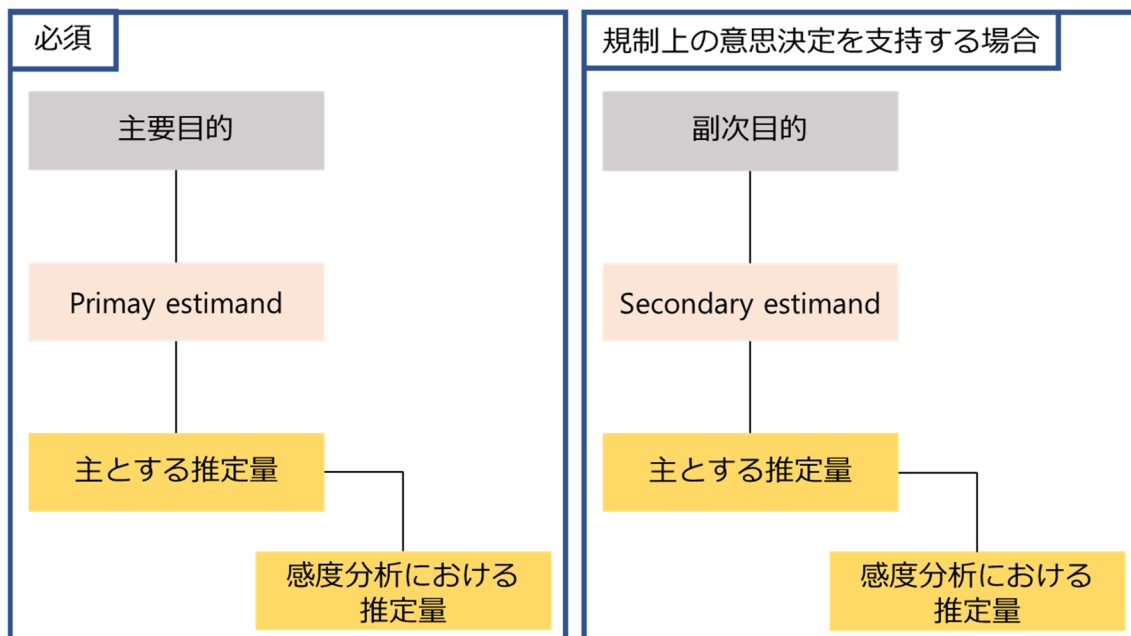
治験実施計画書では、試験の主要目的に対応する **primary estimand** を設定し、明確に規定すべきである。治験実施計画書と統計解析計画書には、**primary estimand** に整合し、主要解析を導く主とする推定量を、仮定からの逸脱に対する頑健性を検討するための適切な感度分析とともに事前に規定すべきである。また、規制上の意思決定を支持する可能性がある副次的な試験目的（例えば、副次評価項目に関連するもの）に対する **estimand** についても、それぞれに対応する主とする推定量及び適切な感度分析とともに定義し、明確に規定すべきである。

**Primary estimand** の選択は、多くの場合、試験デザイン、実施及び解析の特徴の主要な決定要因となる。通常行われているように、これらの特徴は治験実施計画書に十分に記載すべきである。また、**secondary estimand** に重要な関心がある場合には、試験デザイン、実施及び解析の特徴の検討は、必要に応じてそれらの **estimand** も支持するように拡張される可能性があり、同様に記載すべきである。

個々の探索的な目的に対する **estimand** を記載することは治験依頼者には役立つが、規制上の要件ではない。

**Estimand** の変更は、通常、治験実施計画書の改訂を通じて反映されるべきである。

上記の治験実施計画書における記載要件を図に整理すると次のようになるだろう。



### 3.2 EIWG からの推奨事項

EUROPEAN Federation of Statisticians in the Pharmaceutical Industry (EFSPI) 及び European Federation of Pharmaceutical Industries and Associations (EFPIA) がスポンサーとなって立ち上がった Estimands Implementation Working Group (EIWG) は estimand のフレームワークを実装させる際の推奨事項を以下のとおり提示している<sup>9)</sup>。

- 1) 試験の目的, estimand (治療効果), 試験デザインの選択, 試験の実施及び統計解析を繋ぐためのツールとして estimand の thinking process の利用を推進する
- 2) 可能であれば, 非専門家にも通ずる表現を用いて部門間の協力及び estimand についての議論を促し, estimand の考え方を臨床開発のルーチンの一部とする
- 3) 欠測データを最小限にするために, primary estimand (及び重要な副次的 estimand) の評価に不可欠なすべてのデータを収集する必要性を臨床試験チーム, 治験責任医師及び患者が認識していることを確認する
- 4) 各 estimand の解析で用いられるデータに焦点を当てる. このデータは, 患者と含めるべき観察の両方を反映するものである
- 5) 治験実施計画書及び統計解析計画書に estimand を組み込む方法を, 並びに総括報告書及び公表文献に estimand 及び結果を伝達する方法を示す事例研究を共有する
- 6) チームが専門家からの助言にタイムリーにアクセスできるように, 気軽な (Drop in) 相談セッションを提供する
- 7) 規制当局及びその他の主要なステークホルダーから, 提案された estimand 及び推定方法のストラテジー (妥当性を含む) に関するフィードバックをできるだけ早く入手する・チームでこのフィードバックを共有する

- 8) 広範な科学コミュニティにおいて estimand に関する議論を促進するために、多様かつ部門横断でのファシリテーターを含む研修及び主催セミナーを提供する

### 3.3 データサイエンス部会としての推奨事項

本節では、前章までに展開した estimand に関する議論をふまえ、estimand を設定し記載する際のデータサイエンス部会としての推奨事項を記載する。

#### 3.3.1 Estimand の設定に関する推奨事項

データサイエンス部会としては、Training Material Module 3に estimand の設定プロセスとして明記されている thinking process の順序に従い、estimand の検討を進めることを推奨する。以下に thinking process の順序について再掲する。

- 1) 疾患及び関心のある治療の設定と試験目的の決定
- 2) 中間事象の特定
- 3) 中間事象に対応するストラテジーの検討
- 4) estimand の設定
- 5) estimand に沿う試験デザイン、データ収集、解析手法の選択
- 6) 主解析における仮定の特定及び仮定を検証するための適切な感度分析の設定
- 7) estimand の文書化

特に estimand の検討時には統計担当者だけでなくステークホルダーの間で試験の目的・関心のある臨床的疑問について認識をそろえる必要がある。3.2節に引用した EIWG の推奨事項では、「非専門的な表現を用いて部門間の協力及び estimand についての議論を促し、estimand の思考を臨床開発のルーチンの一部とする」という記載がある。データサイエンス部会としても、estimand の各要素の規定や解析手法の選択によって、データ収集の範囲や治療効果を推定するために必要となるサンプルサイズ等にも影響があるため、ステークホルダー間で共同して理解を深めるべきと考える。

#### 3.3.2 Estimand の記載に関する推奨事項

前述したとおり ICH E9(R1)では estimand について、治験実施計画書での具体的な記載箇所、記載方法については明確に定められていない。しかしながら、Training Material 及び Transcelarate の eCPT template 等をもとに検討した結果、データサイエンス部会としては以下に記載した内容が目安となりうると考えている。

##### 3.3.2.1 治験実施計画書「目的 (Objectives)」の章

試験の目的の章について Transcelarate eCPT template (2021 release)では「Objectives,

Endpoints, Estimands」という見出しとし、estimand の設定についても記載できるように構成されている。また、ICH M11 CeSHaP (Clinical electronic Structured Harmonised Protocol) の試験実施計画書テンプレートに関するガイドラインのステップ2では、目的の章や統計解析の章に estimand や中間事象の取り扱い等に関する記載項目が設けられている。従来、目的の章では試験の目的と評価項目（エンドポイント）の対応が記載されていたが、今後 estimand が設定される試験においては、試験の目的と設定する estimand の対応についても当該章にて記載し、primary estimand 及び規制上の意思決定を支持する可能性がある secondary estimand を明示することを推奨する。具体的には、治験実施計画書に明記すべき estimand を、ICH E9(R1)に従い、次の5つの要素が読み取れるように記載する。

- Treatment（関心のある治療）
- Population（対象集団）
- Variable（変数）
- Other intercurrent events（その他の中間事象）
- Population-level summary（集団レベルの変数要約）

なお、Fletcher ら<sup>9)</sup>も同様の主張をしている。Estimand のフレームワークが導入される前は、試験目的は評価項目（エンドポイント）と対応していたため、“endpoint-driven” protocol が用いられていた。しかし、estimand のフレームワークを適用する場合は、評価項目は estimand の1つの要素であり、“estimand-driven” protocol となるため、治験実施計画書テンプレートについては移行期間が必要である。

Estimand の記載方法においては、目的ごとに estimand の各要素を明示して箇条書きや表形式で記載する方法のほか、5つの要素を含めた総体的な文言で estimand を表現する方法も考えられる。また、一つの文章として関心のある臨床的疑問と estimand を記述することで、当局や関係者間で試験に関する理解を深める上で役立つかもしれない。試験の目的や estimand ごとに節を分ける等、レビュー者が理解しやすい工夫がなされるとよい。記載例は4.1.3節を参照されたい。その他、Bell ら<sup>10)</sup>やLawrance ら<sup>11)</sup>は、特に中間事象の取り扱いに関する記載において統計担当者以外でも理解でき、部門横断で議論を可能にすることを意図した Detailed Clinical Objective (DCO)アプローチを文書化の方針として提案している。

なお、開発計画上での当該臨床試験の位置付けや開発相、試験目的等に応じて、「目的 (Objectives)」の章と「統計解析 (Statistical Analysis)」の章における estimand の記載の程度を変えることも可能かもしれない。また、今後、疾患ガイドライン等において検証的試験で設定する estimand として推奨されるものが明確になれば、ガイドラインを参照して試験ごとに薬剤の特性等を鑑みた上で慎重に検討し、estimand を記載していくことになる。

### 3.3.2.2 治験実施計画書「統計解析 (Statistical Analyses)」の章

「統計解析 (Statistical Analyses)」の章には、「目的 (Objectives)」の章で規定した primary estimand に整合した、主とする推定量、対応する主要解析及び感度分析の解析手法等を記載する。また、規制上の意思決定を支持する可能性のある secondary estimand や重要な副次目的に対応する estimand が設定されている場合は、primary estimand と同様に主とする推定量、対応する主要解析及び感度分析の解析手法等を記載する。その他の estimand については、治験実施計画書への記載は必須ではなく、必要に応じて統計解析計画書に詳細を記載することで対応可能と考える。

主とする推定量 (解析手法) がどのような仮定に基づくのかを「統計解析 (Statistical Analyses)」の章に記載すべきと考える。欠測値に対する取り扱い、感度分析も「統計解析 (Statistical Analyses)」の章に記載することを推奨する。また、記載にあたっては、その感度分析が主要解析におけるどのような仮定に対する頑健性を確認しているかに明記すべきである。加えて、補足的解析が primary estimand 又は規制上の意思決定を支持する可能性がある secondary estimand に対応する解析結果の解釈に役立つと考えられる場合には、補足的解析についても明記することを推奨する。

他方で、解析手法に関するプログラムのコードに対するオプションの設定に代表される、詳細な技術的設定については統計解析計画書や各種仕様書への記載のみで対応可能と考える。また、技術的詳細に関する記載については、従前の検討 (成果物としては2019年度 CDISC チーム「総括報告書の解析帳票レイアウトと使用する ADaM データの事例」<sup>12)</sup>) から大きな変更にはならないと考える。

### 3.3.2.3 その他の章における記載

3.3.2.1節, 3.3.2.2節に基づき記載された estimand 及び主要解析, 感度分析等から, 推定値を求めることができるように必要な方策を適切な個所に記載することが望ましい。例えばデータの収集範囲であれば, データ収集の章への記載を, 患者の試験治療の中止理由が必要であれば治療中止に関する章への記載等が考えられる。薬剤の追加を許容する場合は併用薬の章について記載されることも考えられる。

また, 中間事象に関しては, ICH E9(R1)及び Training Material Module 3では総括報告書において各治療群における各中間事象の発現件数及び発現時期の要約を記載することが求められている。中間事象は, 計画時に推定したい治療効果に影響を及ぼすと考えられた事象のため, 中間事象の発現件数が多かったり, 治療群別で発現状況が異なる場合には, 中間事象について一覧表を出力したり, 集計したり, estimand にもとづいて推定された治療効果に与える影響を分析することが適切かもしれない。必要に応じて治験総括報告書に中間事象の発現状況が記載できるように, 治験実施計画書や統計解析計画書への記載を検討することは有用であろう。

### 3.4 安全性評価の estimand の記載

本節では、安全性評価に関して estimand を利用することが有用な状況と、治験実施計画書への記載方針に関して述べる。

Estimand のフレームワークを用いた安全性の評価について具体的な状況を考えると以下が想定される。

- 当該評価項目が、安全性の評価において重要な場合
  - － 糖尿病治療薬の低血糖状態等、治療効果と副作用に密接な関係がある
  - － 高血圧治療薬投与後、高血清カリウム値となるイベント発現
- 評価項目に関する関心のある治療に対してベネフィット・リスク評価を行う際に安全性に関する仮説の評価が重要となる場合
  - － データからは検証不可能な仮定を含むモデルを用いた解析
  - － 仮定の頑健性を示す感度分析／補足的解析を実施する予定がある場合

すなわち、薬剤の特性から、事前に安全性における懸念が明確に特定されており、リスク評価を明示的に行う場合に、estimand のフレームワークを利用し、その構成要素を明記することが望ましいと考えられる。

治験実施計画書への記載にあたっては、安全性評価が試験の主要目的か否かによって異なると考えられる。検証的試験においては有効性の検証が主たる目的であり、primary estimand が有効性の主要目的に対応することが一般的であると想定されるため、安全性の estimand が規制上の意思決定を支持する可能性がある secondary estimand に該当する場合は「目的 (Objectives)」及び「統計解析 (Statistical Analysis)」の章に明示することを推奨する。

### 3.5 Estimand 記載における留意事項

本節では、estimand を治験実施計画書に記載するうえで留意すべき点について、文献をもとに整理する。

#### 3.5.1 治験実施計画書に記載する estimand の程度

ICH E9(R1)では3.1節で触れたように、主たる目的に対応する primary estimand とその感度分析、及び重要な副次的目的に対応する estimand について治験実施計画書への記載を求めているが、具体的な記載方法については規定されていない。記載の程度についても同様に明確に示されていない。

なお、Mallinckrodt ら<sup>13)</sup>には、以下の記載がある。

『Estimand の詳細についてどこまで治験実施計画書に含み、どこからは統計解析計画書での記載でよいかは実用上重要な問題として存在する。各ステークホルダーのニーズに対応するために設定された多数の estimand に対して詳細な定義や中間事象の取り扱いを記



載するのは非常に困難である。臨床試験は幅広い性質をもつため治験実施計画書に詳細に含めるべき内容と、詳細は後続する統計解析計画書等に記載すればよい内容について、一般的な推奨事項として定めることは実用的ではないと考えられる。しかし、原則として、データ収集方針や試験デザインへの影響から、primary estimand の詳細及びその解析方法については治験実施計画書に記載すべきである。』

また、primary estimand とは異なる主たる目的に対応する補足的解析 (Mallinckrodt らはこれらを secondary estimand と呼称) の取り扱いについては以下のように記載されている。

主たる目的に対応する secondary estimand については、治験実施計画書では要約 (ハイレベルサマリー) を記載すればよいが、必要とされるデータ収集の範囲は primary estimand と同様に治験実施計画書に記載すべきである。例えば、primary estimand では割付けられた治療による効果について関心があり、治療中止やレスキュー薬の使用という中間事象に対して、その後のデータは利用しないストラテジーを適用する場合を考える。この場合、前述の中間事象後のデータは収集する必要はない。しかしながら、規制上の意思決定を支持する可能性がある secondary estimand において治療方針ストラテジー等の中間事象後のデータを必要とするストラテジーを適用する場合には、データ収集の範囲には治療中止後のデータも含まれることを治験実施計画書に規定する必要がある。

ただし、他の評価項目にて試験デザインやデータ収集に関して考慮されていれば、すべての estimand について中間事象の取り扱いと解析方法の詳細を治験実施計画書に規定する必要はなく、盲検解除前に統計解析計画書に記載することで対応可能かもしれない。

Mallinckrodt らは中間事象後のデータがすべての合理的な estimand に関連しなくとも、有用な情報を与える場合があるので、少なくとも主要評価項目については中間事象後のデータもすべて収集することが望ましいという立場をとっている。

### 3.5.2 試験開始後の estimand の変更に関する対応

ICH E9(R1) A.6では試験途中での estimand の変更は、試験の信頼性を低下させる可能性があること、また、estimand の変更にあたっては治験実施計画書の改訂を伴うべきと記載されている。Training Material Module 3においても、『立案段階で特定されず、試験段階で特定された中間事象によって estimand が変更になる場合は治験実施計画書の改訂を計画すべきである』と記載されている。

これについて Mallinckrodt ら<sup>13)</sup>には以下のとおり記載されている。

試験開始後に新たに中間事象として取り扱うべき事象が判明することもある。この場合、中間事象に対して、試験目的を適切に反映するストラテジーを選択することはもちろんだが、試験途中の評価の変更となるため、結果への影響についても十分に評価することが望ましい。特に治験実施計画書に記載された primary estimand を変更する場合には、規制要件上、治験実施計画書を改訂することが求められる。また、primary estimand は試験の主要な目的と強く連関するため、中間事象の定義漏れ等は可能な限り早期に特定し治験



実施計画書を改訂することが望ましい。

なお、Training Material Module 2.6には、計画段階で予見していなかった、又は試験実施中に特定された中間事象に対し治験実施計画書の改訂が難しい場合には、予見していなかった中間事象の解析での取り扱い及び影響の評価を行うべきと述べている。また、Cro<sup>14</sup>も例えば COVID-19関連等、想定していなかった中間事象が発生した場合について同様な主張を述べている。

### 3.5.3 規制当局間で異なる estimand の記載

国際共同治験においては規制当局間で関心のある臨床的疑問に対する考え方の違いにより、単一の治験実施計画書内で規制当局ごとに異なる estimand を設定する可能性がある。実際に、2型糖尿病に対する臨床試験において、FDA 以外の規制当局に対しては primary estimand に仮想ストラテジーを適用し、FDA に対しては治療方針ストラテジーを適用したという事例が存在し、治験実施計画書にも明記されている（ClinicalTrials.gov NCT03989232）。ICH E17 2.2.4節を参照すると当局間で estimand が異なる場合には多重性の調整は必要ないと考えられる。ただし、主要評価項目や重要な副次評価項目で多重性の調整を行っている場合には、estimand の違いにより当局間で多重性の調整方法を変える必要があるか検討し、各当局と estimand の設定について協議し治験実施計画書に記載する必要があるだろう。

【ICH E17 2.2.4】十分に正当化された科学的又は規制上の理由により合意できない場合は、複数の規制当局による異なる要件を満たすように、該当する評価項目の項で対応を記載した上で単一の治験実施計画書を作成すべきである。この場合、規制当局ごとに異なる主要評価項目に基づき試験の成功が判断されるため、規制上の意思決定のために複数の評価項目について多重性の調整を行う必要はない。

### 3.5.4 治験総括報告書で求められる中間事象の発現状況に関連する記載

治験総括報告書（CSR）での考察を行うにあたって、解析のために準備すべき必要な情報を盛り込むことが挙げられる。Training Material Module 2.6には以下のとおり記載されている。

中間事象は、推定したい治療効果に影響を及ぼす事象として計画時に考えられた事象であるから、各 estimand におけるすべての中間事象の発現状況について、発現件数と発現時期が分かるよう、帳票として作成することが望ましい。また、主解析、感度分析、補足的解析それぞれについて、試験開始前に規定されていたのか、試験開始後盲検下の期間に規定されたか、事後解析として行われたか、報告に盛り込むべきことが望ましい。また、試験開始後に追加された中間事象については、治療効果の推定に当たって解析でどのように考慮されたか、事前規定された estimand への影響と、中間事象追加にともなう治療効

果の推定値への影響について説明できるよう、解析を計画すべきである。

Fletcher ら<sup>9)</sup> は、中間事象の治験総括報告書への記載に関して、EIWG での議論を踏まえて以下の方法を推奨している。

Estimand の設定の1つとして中間事象は重要であることを考慮し、(中略) 中間事象の発現件数と試験内での患者の治療経過の多様性がすぐに理解できるよう、要約表、又は図による提示が治験総括報告書内にあるとよい。これを記載として盛り込むことは比較的容易と考えるが、レスキュー薬の使用を治療方針ストラテジーで取り扱うような状況では、試験結果の解釈に大きく影響するかもしれない。

なお、上記解析計画を記載する具体的な文書については、Training Material でも明記されていない。治験実施計画書に記載する場合の記載箇所に関しては、例えば Lynggaard ら<sup>15)</sup> は、『Protocol section for intercurrent events and associated handling strategies section』に含めることを提案している。

### 3.6 小括

本章では前章までの議論に基づき、治験実施計画書等への estimand の文書化にあたってデータサイエンス部会にて推奨する事項と、関連した文献等の議論状況について整理した。

なお、本章に記載した指針は一つの例であり、記載事例をすべて満たすことが必須要件ではない。ICH E9(R1)では estimand に関する治験実施計画書への記載箇所や分量については規定されておらず (Training Material p.185)、今後治験依頼者及び規制当局間で実際の臨床試験での estimand を用いた議論を深めていくことにより、本報告書で示した方針とは多少異なるかたちで定着することもありうるだろう。

## 4. 仮想事例

本章では、仮想事例での estimand 及び中間事象の設定例、治験実施計画書の記載例を示す。仮想事例の対象疾患は、糖尿病（プライマリーケア領域）と悪性リンパ腫（オンコロジー領域）とし、いずれも被験薬の優越性を検証する第 III 相臨床試験を想定した。

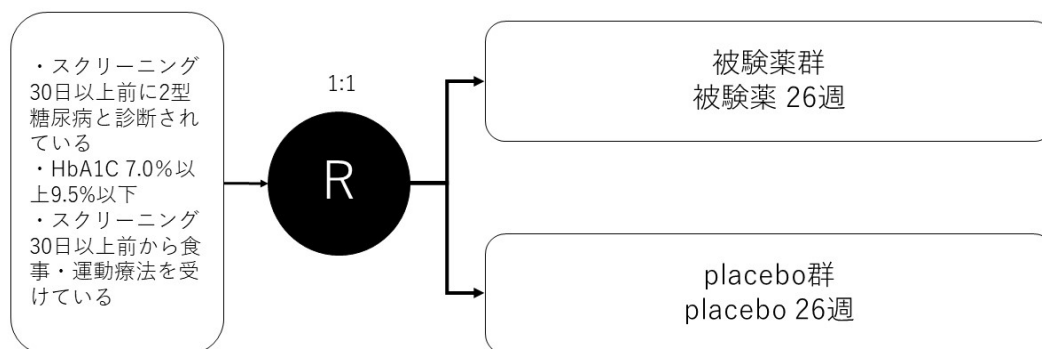
なお、本章はあくまで仮想事例及び記載例を紹介することを目的としており、データサイエンス部会として当該 estimand の設定及び記載に対する公式な見解を推奨するものではないことにご留意頂きたい。

### 4.1 プライマリーケア領域: 糖尿病の仮想事例

#### 4.1.1 典型的な試験デザイン

##### 1) 試験概要

一つめの仮想事例では、食事・運動療法のみでは効果不十分である2型糖尿病患者を対象とし、プラセボに対する被験薬の有効性及び安全性を検討する、第 III 相ランダム化二重盲検プラセボ対照多施設共同治験を想定した。試験デザインの概要を図 4-1 に示す。



R: ランダム化

図 4-1 試験デザイン概要（糖尿病の仮想事例）

##### 2) 目的と主要評価項目

本試験の主要目的は、「食事・運動療法のみでは効果不十分である2型糖尿病患者を対象として、被験薬1日1回投与の血糖コントロールに対する効果を、プラセボの1日1回投与と比較検討すること」であった。

主要評価項目は、HbA1c(%)のベースラインから26週までの変化量とした。

### 4.1.2 Estimand, 中間事象及び対応するストラテジーの設定例

#### 4.1.2.1 中間事象

本試験では、以下2つの事象を中間事象と特定した。

- 治験実施計画書で規定したレスキュー薬/追加の糖尿病薬の使用（この仮想事例では便宜上、治験薬投与を継続した状態で開始する治験薬以外の糖尿病薬と、治験薬早期中止後に使用される治験薬以外の糖尿病薬をまとめて「(治験実施計画書で規定した) レスキュー薬/追加の糖尿病薬」とする)
- 理由を問わない治験薬中止

#### 4.1.2.2 Estimand

##### 1) Primary estimand

本試験では、primary estimand を「食事・運動療法のみで効果不十分の2型糖尿病患者に対し、治験薬の投与を早期中止したかどうかに関わらず、すべてのランダム化割り付けされた被験者におけるレスキュー薬/追加の糖尿病薬の効果を含む治療効果(被験薬 vs プラセボ)をHbA1c(%)のベースラインから26週までの平均変化量の群間差を用いて評価する」と定めた。本試験の primary estimand を5つの要素を用いて、表 4-1に整理する。

表 4-1 Primary estimand (糖尿病の仮想事例)

関心のある治療 Treatment	どのような治療効果を	プラセボ (+必要に応じてレスキュー薬/追加の糖尿病薬*) に対する被験薬 (+必要に応じてレスキュー薬/追加の糖尿病薬*) の効果を,
対象集団 Population	どのような患者に対し	治験実施計画書に規定した選択除外基準を満たす、食事・運動療法のみでは効果不十分である2型糖尿病患者 (集団) に対し,
変数 Variable	どのような評価尺度で	HbA1c(%)のベースラインから26週までの変化量で,
その他の中間事象 Other intercurrent event	どのようにその他の中間事象を扱い	被験薬/プラセボ投与の早期中止*の有無に関わらず (治療方針ストラテジー)
集団レベルの変数要約 Population-level summary	どのように要約するか	プラセボ群と被験薬群の平均値の群間差で要約する。
備考 Stakeholder 等	治験薬投与の早期中止やレスキュー薬/追加の糖尿病薬の使用を含めた上でも、被験薬を使用した集団の治療効果がプラセボを使用した集団を上回ることが、規制当局等、対象患者集団の治療方針の決定に関わる人にとって重要であると考えられた。	

\*: 中間事象

##### 2) 補足的解析

補足的解析として、以下の2つの estimand も設定した。

- ① 食事・運動療法のみでは効果不十分である2型糖尿病患者に、レスキュー薬/追加の糖尿病薬を使用せず、すべての被験者が治験薬の投与を継続したと仮定した場合

の治療効果（被験薬 vs プラセボ）を HbA1c(%)のベースラインから26週までの平均変化量の群間差を用いて評価する。

- ② 食事・運動療法のみでは効果不十分である2型糖尿病患者に対し、レスキュー薬/追加の糖尿病薬を使用しなかったと仮定した場合の、治験薬の投与を早期中止したかどうかに関わらない治療効果（被験薬 vs プラセボ）を HbA1c(%)のベースラインから26週までの平均変化量の群間差を用いて評価する。

補足的解析①及び②の estimand を表 4-2に整理する。

表 4-2 補足的解析①及び②の estimand（糖尿病の仮想事例）

関心のある治療 Treatment	どのような治療効果を	プラセボに対する被験薬の効果を,		
対象集団 Population	どのような患者に対し	治験実施計画書に規定した選択除外基準を満たす、食事・運動療法のみでは効果不十分である2型糖尿病患者（集団）に対し、 ※primary estimand と同じ		
変数 Variable	どのような評価尺度で	HbA1c(%)のベースラインから26週までの変化量で、 ※primary estimand と同じ		
その他の中間事象 Other intercurrent event	どのようにその他の中間事象を扱い		補足的解析①	補足的解析②
		レスキュー薬/追加の糖尿病薬の使用	なかったと仮定して（仮想ストラテジー）	
		治験薬投与の早期中止	なかったと仮定して（仮想ストラテジー）	有無に関わらず（治療方針ストラテジー）
集団レベルの変数要約 Population-level summary	どのように要約するか	プラセボ群と被験薬群の平均値の群間差で要約する。 ※primary estimand と同じ		
備考 Stakeholder 等	<p>補助的解析① 薬剤の中止やレスキュー薬の影響を排除した被験薬の純粋な薬効を考慮に入れたうえで、個々の患者に対し臨床上の判断を行う医師にとって重要と考えられたことから設定した。</p> <p>補助的解析② （EMA 糖尿病ガイドライン<sup>18</sup>）に従い）レスキュー薬/追加の糖尿病薬の影響を排除した効果を評価したい一方、ある一定の割合で発生する、忍容性の問題による治験中止は考慮に入れた評価をする必要がある規制当局にとって重要と考えられたことから設定した。</p>			

#### 4.1.2.3 中間事象のストラテジー

この試験の primary estimand 及び補足的解析での、中間事象とそのストラテジーのまとめを表 4-3に示す。

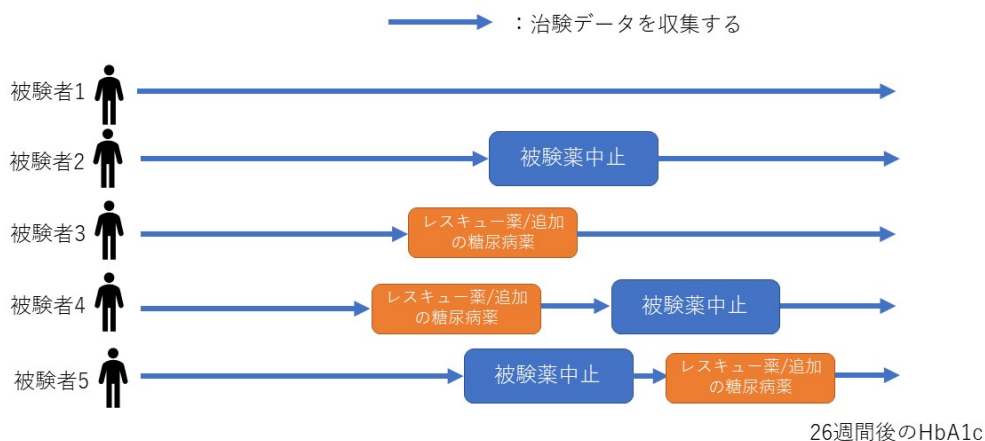
表 4-3 中間事象とそのストラテジーのまとめ（糖尿病の仮想事例）

	Primary estimand	補足的解析①	補足的解析②
治験実施計画書で規定したレスキュー薬/追加の糖尿病薬の使用	治療方針ストラテジー	仮想ストラテジー	仮想ストラテジー
理由を問わない治験薬中止	治療方針ストラテジー	仮想ストラテジー	治療方針ストラテジー

中間事象及び対応するストラテジーについては初めに発現した事象を採用することとした。

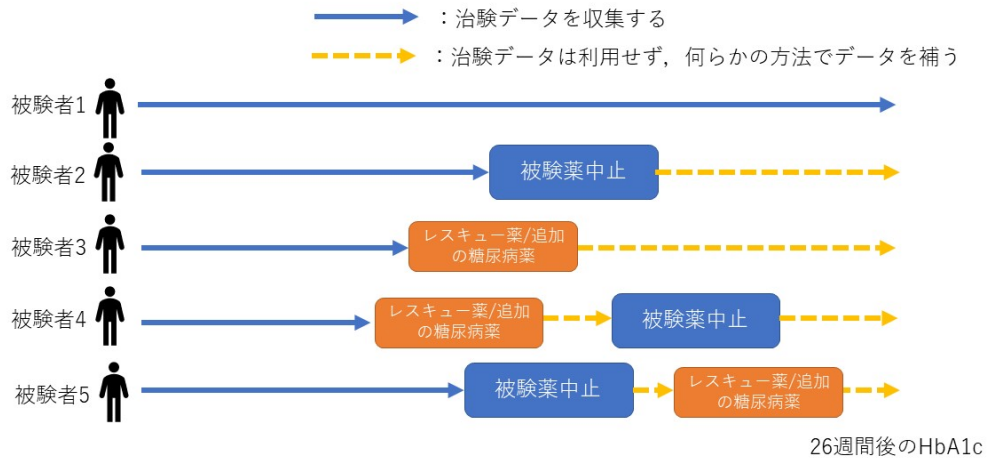
Primary estimand 及び補足的解析で、中間事象が発現した後の個々の患者データの扱い方を示した模式図（Patients' Journey）を図 4-2に示す。

Primary estimand：治療方針ストラテジー





補足的解析①：仮想ストラテジー



補助的解析②：Hybrid（治療方針ストラテジー及び仮想ストラテジー）

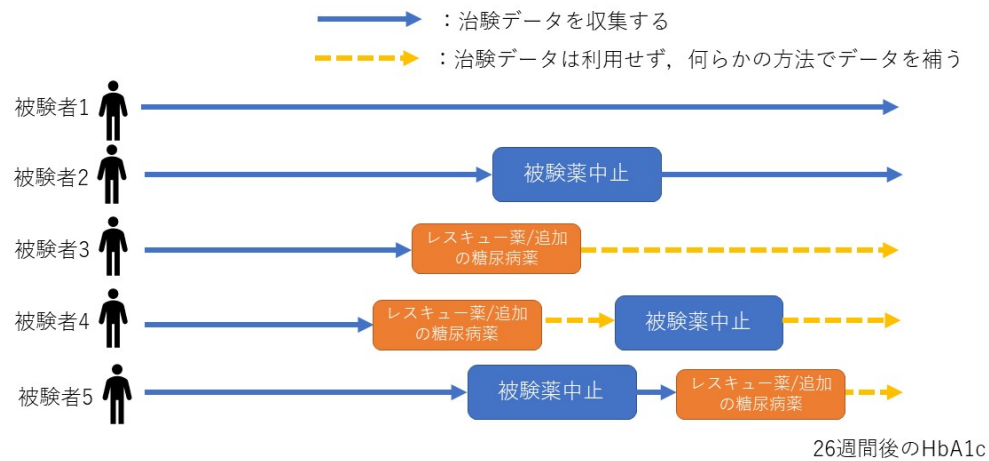


図 4-2 各 estimand での Patients' Journey（糖尿病の仮想事例）

上記で述べた中間事象の他にも、実際には検討が必要な中間事象が存在すると考えられる。例えば、理由を考慮した治験薬中止（有効性欠如，有害事象，それ以外），併用禁止薬の使用，死亡等である。これらの事象について，中間事象として設定するかの判断は，想定している試験の状況や治療効果への影響等から判断すべきである。

4.1.3 治験実施計画書への記載

4.1.3.1 目的（Objectives）の章

以下に「目的（Objectives）」の章における主要目的，主要評価項目及び estimand の記載例を示す。Estimand の5つの要素を含めた，文章形式（平叙文，疑問文いずれも可）を記

載すると共に、各要素に対する説明（表形式、箇条書き等）の記載を例示する。ただし、必ずしも併記する必要はない。

#### 目的及び評価項目

目的	評価項目
主要評価	
食事・運動療法のみで効果不十分な2型糖尿病患者を対象として、被験薬の1日1回投与の血糖コントロールに対する効果をプラセボの1日1回投与と比較検討すること	HbA1c(%)のベースラインから26週までの変化量

#### Estimand

食事・運動療法のみで効果不十分な2型糖尿病患者に対し、治験薬の投与を早期中止したかどうかに関わらず、すべてのランダム化割り付けされた被験者におけるレスキュー薬/追加の糖尿病薬の効果を含む治療効果（被験薬 vs プラセボ）を HbA1c(%)のベースラインから26週までの変化量の平均値の群間差を用いて評価する。

Estimand を構成する5つの要素を以下に示す。

- 関心のある治療：プラセボ（+必要に応じてレスキュー薬/追加の糖尿病薬） vs 被験薬（+必要に応じてレスキュー薬/追加の糖尿病薬）
- 対象集団：食事・運動療法のみで効果不十分な2型糖尿病患者集団（※必要に応じて治験実施計画書の該当項を参照させる）
- 変数：HbA1c(%)のベースラインから26週までの変化量
- その他の中間事象：治験薬投与の早期中止。治療方針ストラテジーを採用する（※必要に応じて治験実施計画書の該当項を参照させる）
- 集団レベルの変数要約：平均値の群間差

Estimand の設定根拠：治験薬投与の早期中止やレスキュー薬/追加の糖尿病薬の使用を含めた上でも、被験薬を使用した集団の治療効果がプラセボを使用した集団を上回ることが、規制当局等、対象患者集団の治療方針の決定に関わる人にとって重要であると考えられた。

#### 4.1.3.2 統計解析（Statistical Analyses）の章

以下に「統計解析（Statistical Analyses）」の章における primary estimand 及び主解析の記載例を示す。「目的(Objectives)」の章に記載した内容を再度「統計解析(Statistical Analyses)」の章に記載する場合には、治験実施計画書の編集方針(同一の内容を再掲する, 要約する,



章や項を引用する、等)に従い適切な段落に記載する。その際は primary estimand の解析方法を記載する箇所の近傍が望ましいだろう。

### Primary estimand

Primary estimand を構成する5つの要素を以下に示す。

- 関心のある治療：プラセボ（+必要に応じてレスキュー薬/追加の糖尿病薬） vs 被験薬（+必要に応じてレスキュー薬/追加の糖尿病薬）
- 対象集団：食事・運動療法のみで効果不十分な 2 型糖尿病患者集団（※必要に応じて治験実施計画書の該当項を参照させる）
- 変数：HbA1c(%)のベースラインから 26 週までの変化量
- その他の中間事象：治験薬投与の早期中止、治療方針ストラテジーを採用する（※必要に応じて治験実施計画書の該当項を参照させる）
- 集団レベルの変数要約：平均値の群間差

### Primary estimand における解析

- 解析方法

HbA1c(%)のベースラインからの変化量に対して mixed model for repeated measures (MMRM) を用いて解析する。モデルには治験薬群、来院時期、及び治験薬群と来院時期の交互作用を含める。分散共分散行列には無構造を仮定する。適切な対比を用いて 26 週時点における被験薬群とプラセボ群の平均値の群間差を求める。

- 中間事象の取り扱い

レスキュー薬/追加の糖尿病薬の服用中に収集されたデータ、及び治験薬中止後に収集されたデータも解析に用いる。

- 欠測値の取り扱い

HbA1c のデータが間欠的に欠測している場合、Missing at Random (MAR) の仮定に基づいて多重補完する。また、治験薬中止後のデータがない場合、補完方法の詳細は解析計画書に記載する。

### 感度解析

Primary estimand の統計的推論を覆すのに必要なペナルティの程度を評価するため、MAR の仮定からのずれに対する感度解析として tipping point analysis を行う。欠測データの補完の不確実性を考慮して、多重代入法により治験薬群毎に補完する。なお、詳細な方法は SAP に記載する。

### 補足的解析

補足的解析① 被験薬とプラセボとの薬効に対し、レスキュー薬/追加の糖尿病薬の影響を排除し、更に全員が服薬コンプライアンスを遵守したと仮定した場合（仮想ストラテジー）の薬効を定量化するために、primary estimand に対し、2つの中間事象であるレスキュー薬/追加の糖尿病薬の使用及び治験薬中止に対し仮想ストラテジーとして扱った評価を行う。具体的には、中間事象発現後に収集されたデータを除き、主解析と同様に解析する。

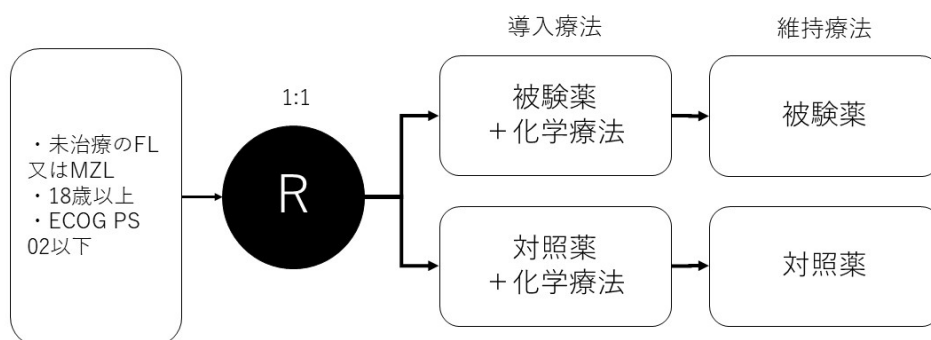
補足的解析② レスキュー薬/追加の糖尿病薬の影響を排除した効果を評価したい（仮想ストラテジー）一方、ある一定の割合で発生する、忍容性の問題等による治験薬中止によりベネフィットが得られない場合（治療方針ストラテジー）を治療効果として考慮する。具体的には、レスキュー薬使用後に収集されたデータを除き、主解析と同様に解析する。

## 4.2 オンコロジー領域: 悪性リンパ腫の仮想事例

### 4.2.1 典型的な試験デザイン

#### 1) 試験の概要

二つめの仮想事例では、未治療の進行期低悪性度非ホジキンリンパ腫患者を対象とし、被験薬 + 化学療法併用後の奏効例に対して被験薬維持療法を施行する群と、対照薬 + 化学療法併用後の奏効例に対して対照薬維持療法を施行する群の有用性を評価する、第 III 相多施設共同非盲検ランダム化試験を想定した。試験デザインの概要を図 4-3 に示す。なお、本仮想事例における中間事象や適用するストラテジーの考え方については Sun ら<sup>22)</sup> を参考にした。



FL: 濾胞性リンパ腫, MZL: 辺縁帯リンパ腫, R: ランダム化

図 4-3 試験デザイン概要（悪性リンパ腫の仮想事例）

## 2) 目的と評価項目

本試験の主要目的は、「未治療の進行期濾胞性リンパ腫患者を対象として、対照薬 + 化学療法併用後に対照薬維持療法を行う治療に対する、被験薬 + 化学療法併用後に被験薬維持療法を行う治療の有効性を主治医評価の無増悪生存期間（PFS：病勢進行又は死亡までの期間）を用いて検証すること」であった。

主要評価項目は、濾胞性リンパ腫被験者集団の PFS（主治医評価）とした。全生存期間（OS）は副次的評価項目の1つとした。

悪性腫瘍は死亡に直結する疾患であることから、多くの抗悪性腫瘍薬の治療目標は「患者の生存期間を延長すること」であり、悪性腫瘍を対象疾患とする臨床試験の主要評価項目は OS が依然としてゴールドスタンダードである。一方、悪性腫瘍の種類により、また対象疾患の病期により、臨床試験の期間内に、評価に必要な OS イベントを十分に集積することが困難な場合もあり、PFS は OS のサロゲートエンドポイント（代替評価項目）として主要評価項目に採用されることもある<sup>19) -21)</sup>。

### 4.2.2 Estimand, 中間事象及び対応するストラテジーの設定例

#### 4.2.2.1 中間事象

本試験では、以下2つの事象を中間事象と特定した。

- 治療中止（治療方針ストラテジー：中止有無に関わらず PFS イベント発現までデータ収集\*する。なお、治療中止の理由により異なるストラテジーを採用することも想定されるが、今回の仮想事例の primary estimand では中止の理由を問わないこととした。）
- PFS イベント発現前の新たな抗がん剤の開始（治療方針ストラテジー：治療の変更によらず、PFS イベントまでデータ収集\*する）

\*: なお、今回の仮想事例の試験では副次的評価項目の1つに OS を設定しているため、試験としてのデータ収集は OS イベント発現までである。

治療中止した場合は中止時点で病勢進行（PD）として扱う解析や、PFS イベント発現前の新たな抗がん剤の開始は打ち切りとして扱う解析は、従前から PFS（主治医評価）に対して、割付けられた治療にもとづく治療効果を確認することを目的に行われてきた。ICH E9(R1)の観点からは、治療中止や新たな抗がん剤の開始を中間事象として取り扱い、これらについて治療方針ストラテジーだけではなく仮想ストラテジーを用いることで推定できる治療効果が、ステークホルダーが判断を行う上で有意義である場合には、primary estimand に対する補足的解析として取り上げてもよいと考えられる。

また、腫瘍評価を中央判定（IRC）でも行う場合、主治医評価による PFS と IRC による PFS の2つの変数が存在する。一般に IRC での評価は主治医評価より後に行われることから、IRC による評価結果が出る前に、主治医評価による PD に基づいて、腫瘍評価が打ち

切られてしまう可能性がある。Sun ら<sup>22)</sup> は、IRC による評価はこの情報がある打ち切りによるバイアスが生まれ、特に盲検試験では主治医評価による潜在的なバイアスよりも結果に悪影響を及ぼす可能性がある」と述べられている。

さて、estimand のフレームワークにおいてこの2つの変数は同一として扱うか（すなわち感度分析として対応させるか）、別個のものとして扱うか（補足的解析として対応させるか）という異なる立場が考えられるが、Sun らは、IRC による PFS を主治医評価による PFS の感度分析として位置づけている。（筆者注：いずれの PFS も評価方法は異なるが「増悪、もしくは理由を問わない死亡までの期間（無増悪生存期間）」を表す変数としては同一であり、治療効果の推定量としても、「無増悪生存期間」に差があるかどうか、という疑問に答えるという点で同一であるからと考えられる。）

#### 4.2.2.2 Estimand

##### 1) Primary estimand

本試験では、primary estimand を『未治療の進行期濾胞性リンパ腫患者を対象として、治療中止及び PFS イベント発現前の新たな抗がん剤の開始に関わらない主治医評価の PFS における治療効果を「治験薬-化学療法→治験薬維持療法」の「対照薬-化学療法→対照薬維持療法」に対するハザード比を用いて評価する』とした。本試験の primary estimand は5つの要素を用いて、表 4-4に整理する。

表 4-4 Primary estimand（悪性リンパ腫の仮想事例）

関心のある治療 Treatment	どのような治療効果を	「対照薬-化学療法→対照薬維持療法」に対する、「被験薬-化学療法→被験薬維持療法」の治療効果を
対象集団 Population	どのような患者に対し	治験実施計画書に規定した選択除外基準を満たす被験者の内、濾胞性リンパ腫の患者集団で、
変数 Variable	どのような評価尺度で	PFS（主治医評価）を、 （ランダム化から改変版悪性リンパ腫効果判定規準に基づき主治医により評価される最初の病勢進行又は再発まで、又は原因を問わない死亡までの期間）
その他の中間事象 Other Intercurrent event	どのようにその他の中間事象を扱い	治療中止*の有無に関わらず、 PFS イベント発現前の新たな抗悪性腫瘍薬の使用*に関わらず
集団レベルの変数要約 Population-level summary	どのように要約するか	層別 Cox 比例ハザードモデル（95%信頼区間を含む）で推定されるハザード比で要約する。

\*: 中間事象

また、本試験では、副次的評価項目の一つに「ランダム化されたすべての被験者（＝濾

胞性リンパ腫患者及び辺縁帯リンパ腫患者) の PFS」を規定している。これは、secondary estimand (重要な副次評価項目の主要な estimand) として、表 4-5 のように整理することもできる。

表 4-5 Secondary estimand (悪性リンパ腫の仮想事例)

関心のある治療 Treatment	どのような治療効果を	「対照薬-化学療法→対照薬維持療法」に対する、「被験薬-化学療法→被験薬維持療法」の治療効果を ※primary estimand と同じ
対象集団 Population	どのような患者に対し	治験実施計画書に規定した選択除外基準を満たす濾胞性リンパ腫患者、辺縁帯リンパ腫患者全体の集団に対し、
変数 Variable	どのような評価尺度で	PFS (主治医評価) を、 (ランダム化から改変版悪性リンパ腫効果判定規準に基づき主治医により評価される最初の病勢進行又は再発まで、又は原因を問わない死亡までの期間) ※primary estimand と同じ
その他の中間事象 Other Intercurrent event	どのようにその他の中間事象を扱い	治療中止*の有無に関わらず、 PFS イベント発現前の新たな抗悪性腫瘍薬の使用*に関わらず ※primary estimand と同じ
集団レベルの変数要約 Population-level summary	どのように要約するか	層別 Cox 比例ハザードモデル (95%信頼区間を含む) で推定されるハザード比で要約する。 ※primary estimand と同じ

\*: 中間事象

繰り返しになるため、オンコロジー領域における仮想事例での「治験実施計画書への記載」は割愛する。治験実施計画書各項への記載内容・分量の目安は、4.1.3節参照のこと。

## Part 2: さらに理解を深めたい方へ

### 5. Estimand に関する議論

本章では、さらに estimand への理解を深めたい方向けに、関連すると考えられる文献や著書を紹介し、Training Material Module 3の“thinking process”に沿って、estimand に関する考慮事項について各著者の見識を整理する。また、優越性の検証以外を目的とする試験に対する estimand や COVID-19に対する対応等、特別な考慮事項については最後にまとめて示す。なお、オンコロジー領域に関する最近の話題については6章を参照いただきたい。

#### 5.1 治療の設定及び試験目的の決定

##### 5.1.1 Tripartite Framework

Estimand のフレームワークを用いて関心のある治療を設定する際に言及されるものとして Akacha ら<sup>23)</sup>において提唱された“Tripartite framework”がある。Akacha らは ICH E9(R1)が制定された背景として言及される2011年の糖尿病治療薬に関する製薬企業と規制当局 (FDA) との治療効果の推定に関する議論にて、製薬企業と規制当局との間で臨床試験において関心のある臨床的疑問が異なったことから議論を始めている。Akacha らは ICH E9(R1)において、ITT の原則等 ICH E9が制定された段階では曖昧な定義であったものを estimand のフレームワークを利用することで整理できること、そして臨床試験で推定された治療効果を基に議論を進めるステークホルダーとして、患者、医師、規制当局、保険者 (支払者) を挙げ、4つの視点から整理を行った。以下に4つの視点について記載する。

患者と医師がもつ疑問としては、患者における忍容性、安全性 (安全性によるアドヒアランスの低下)、及び有効性の効果として、効果が判定できるまでの期間や投与量の増加 (効果不十分による治療中止の割合)、治療期間中に発生する有効性反応及び副作用 (有害事象への影響) がある。

規制当局がもつ疑問としては、予想される副作用、副作用の発現時期及び発現期間、副作用の転帰 (安全性として忍容性がなく治療中止した患者の割合)、治療中止や用量変更の決断を下すまでに治療を受ける期間 (効果不十分による治療中止の割合)、治療で患者に期待される有効性の効果や安全性の懸念があるか (患者への効果) があげられている。

保険者 (支払者) がもつ疑問としては、治療を遵守できない患者数、治療完了、中止によらず副作用が発現する患者数、費用対効果の観点から肯定的な転帰となる患者数等が挙げられている。

(筆者注: 日本では公定薬価制度が導入されていることや、欧米とは異なるシステムで



治療へアクセスするため、本議論における保険者（支払者）に相当する存在があるかは不明瞭である。しかしながら、医薬品の使用指針となるフォーミュラリーの議論が進んでおり、仮に保険者（支払者）の視点による費用対効果に基づく治療薬の選択が行われることになれば、上記の疑問を持つ可能性がある。）

これらの視点から Akacha らは、背景や動機は異なるが、estimand に対して共通の関心を持っていることを述べ、3つの観点（Tripartite）として提言している。

- 安全性による服薬不遵守：治療に耐えられず服薬を中止した患者の割合
- 効果不十分による服薬不遵守：効果不十分により服薬を中止した患者の割合
- アドヒアランスにおける効果：計画した期間にわたって治療を実際に遵守することができる患者について、治療の予想される有効性及び安全性プロファイルはどのようなものか

Akacha らは、試験治療の安全性と有効性の全体像を把握するためには、上記3つの観点を検討すること、またアドヒアランスの差異による治療のリスクやベネフィットについても臨床的に意味のある評価が可能だと述べている。またこれに続けて、ITT の原則による治療効果の実態や、3つの論点に対応する統計学的手法について述べているが、本報告書では割愛する。

### 5.1.2 関心のある治療の設定

Mallinckrodt ら<sup>13)</sup> も同様に、関心のある治療を設定するために、ステークホルダーの特定と意思決定内容、そして、意思決定の根拠となる試験における「臨床的疑問」について具体化することの必要性を述べている。関心のある治療を定めるためには、試験で評価したい治療が何であるかを明確に定義することが重要であるとし、試験の目的、試験の相、疾患領域、実臨床で想定される治療方針、試験治療の特性等に応じて関心のある治療は変わりうる可能性があるとして述べている。

Ratitch ら<sup>24)</sup> は、推定の対象とする治療効果について論じている。以下に推定の対象ごとに整理した。

- 薬剤を服用することによる治療効果を評価する：  
治療の変更や他の治療の追加を行うことなく、一定期間の試験治療を受けた状態の治療効果や安全性に関心があるかもしれない。
- 有効性と忍容性の両方を加味した効果、又は複数の薬剤・治療法の組み合わせによる治療効果を評価する（評価したい治療条件によって2つの場合に分けられる）：
  - － 治療の状態が複数の時期によって規定される場合（multi-period regimen）  
例：有害事象や効果不十分といった理由により割付けられた治療を中止したが、その後の代替治療が存在しない又は適切ではなく無治療となる状況

- 複数の治療法によって規定される場合 (multi-treatment regimen)  
例：割付けられた治療の開始後に一定の改善がみられない被験者に対し、レスキュー薬の使用又は追加投与を行う状況

Mallinckrodt らはステークホルダーごとに興味のある目的と治療の定義が異なる可能性があることにも触れた上で、適切なサンプルサイズを設定するためにも治療効果を明確に定義することが重要であると述べている。

### 5.1.3 試験目的の決定

Ratitch ら<sup>24)</sup> は、関心のある治療と試験の目的との対応について、ステークホルダーの多様性という観点から検討し以下のように考察した。

第 I/II 相試験では Proof of Concept の確認や以後の試験での用量の特定を目的とし、主に製薬企業での意思決定の支援を主たる目的とすることが多い。第 III 相試験（検証的試験）は一般的に多様なステークホルダーが関与する試験であり、多様な目的に対処する必要があると考えられる。臨床試験のこのような多面的な性質は、目的及び estimand を選択する際に重要な考慮事項の一つである。複数の目的に対応するためには、複数の estimand が必要になることが多い。

Training Material Module 2.1にも記載されているとおり、関心のある臨床的疑問は、試験で意思決定したいことに対して、どのような治療効果を推定すればよいのかを整理したものである。試験目的は、介入、対照、患者集団、評価項目、期間等の要素を考慮して、関心のある臨床的疑問を表現することになる<sup>13)</sup>。ICH E9(R1)でも試験目的の記載があいまいであると、試験デザインの設計や解析手法の選択が、目的に適合したものではなくなる可能性があること、結果的に治療効果の推定値を正確に解釈することができず、ステークホルダーの意思決定をミスリードする可能性があることを指摘している。Training Material Module 2.2 p.72にあるとおり、ここで定義される明確な試験目的は、適切な estimand を設定することで関心のある臨床的疑問に変換される。

### 5.1.4 主たる目的の検討

Ratitch ら<sup>24)</sup> は、試験に基づいて判断したいことは複数あることが多いが、たとえステークホルダーが単一であっても、一つの試験から検討したい目的が複数存在し、それゆえ estimand も複数設定する必要があるときも存在すると述べている。しかしながら、Ratitch らは同時に、試験において一つの目的とそれに対応する estimand を主たる目的とし、他の estimand を主たる目的に対応する secondary estimand や副次的な目的に対応する estimand として設定することが適切であると述べている。

Training Material Module 1でも同様に、臨床担当者と統計担当者、あるいは規制当局と製薬企業といった立場の異なる複数のステークホルダーの視点を持つことや、primary

estimand によって試験デザインの主要部分が定義された場合でも、規制当局との検討をサポートするような secondary estimand も特定した上で、正確に規定すべきだと述べている。

## 5.2 中間事象の特定

中間事象は「関心のある治療」の定義に直接的に関連すると考えられ、関心のある治療効果を明確に規定するためには、中間事象の発現後の結果の解釈への影響を十分に検討する必要がある。代替治療やレスキュー薬を例にとっても、その種類、理由、期間、用量、タイミング等は多岐にわたることが想定され、試験の目的に応じて異なるストラテジーを適用するといった対応も考えられるであろう。複雑な例として、治療中止後にレスキュー薬を使用した場合と、治療中止前にレスキュー薬を使用した場合を区別して取り扱う例が考えられる。“このような”一人の被験者に複数の中間事象が発現する場合にどのように取り扱うか事前に“規制当局等と”検討し規定する必要がある<sup>25)</sup>。

中間事象と治験実施計画書からの逸脱との関連については、Fletcher ら<sup>9)</sup>において ICH E9(R1)の作成時に以下の議論があったと説明されている。

(中間事象と治験実施計画書からの逸脱として定義される事象の) 一部は重複している。例えば、重要な治験実施計画書からの逸脱は、治験薬投与中の予期しない投与中断である可能性があり、これは中間事象とみなすこともできる。重要な治験実施計画書からの逸脱の定義、発生及び報告は継続すべきであり、重要な治験実施計画書からの逸脱の数と程度は、試験が高い質で実施されたかどうか及び試験の完全性が維持されたかどうかの代替指標になりうることから、逸脱の発生を最小限に抑えるために医薬品の臨床試験の実施に関する基準を遵守すべきである。

## 5.3 中間事象に対応するストラテジーの検討

### 5.3.1 ストラテジーの選択

本節では中間事象に対応するストラテジーの選択にあたっての議論をまとめる。ストラテジー固有の議論に関しては、次の5.3.2節を参照いただきたい。

関心のある臨床的疑問と中間事象との関係を考えるにあたり、中間事象に対して2つの異なるアプローチが存在する<sup>24)</sup>。

- 1) 中間事象そのものが関心のある治療の状態の一部となりうる場合
- 2) 中間事象が関心のある治療の状態からの逸脱（乖離）とみなされる場合

前者の代表的な例は、中間事象を治療方針ストラテジーで対応する場合であり、中間事象の発現を許容した治療の状態に関心がある。これに該当する中間事象については、対応するストラテジーについて深掘する必要はない<sup>13)</sup> が、治療の一部として治療の要素に組み込むことになる<sup>9)</sup>。後者は、評価の際に適切な考慮が必要であり、関心のある治療の状態に応じて、適切に中間事象に対応するためのストラテジーを検討し、解析手法に反映さ

せることが重要となる。

Mallinckrodt らの報告<sup>13)</sup> には以下の記載がある。

中間事象に対応するための戦略は評価の対象となる治療レジメンが試験の目的と整合するように選択されるべきものであり、選択したい推定量（推定方法）によって定められるべきではない。戦略の選択の際には、推定しようとする治療効果がステークホルダーの意思決定のために妥当であるか、また、信頼できる推定値を得ることが可能であるかという点を考慮に入れることが推奨される。

なお、中間事象として取り扱うことが適切であり、かつその事象が試験あるいは実臨床においてある程度の蓋然性で起こりうると考えられるが、一般的な戦略が定まっていなかったり、あるいは戦略を複数取りうる可能性があることも想定しておくべきである。特に、その選択がステークホルダー間で重要な関心事になりうる場合には、事前に戦略の選択について合意しておくことが望ましい。

Mallinckrodt らは、関連するデータが欠測している場合、中間事象に対応するすべての戦略で欠測に対する仮定が必要であると述べている。詳細は5.7節を参照されたい。また、Ratitch ら<sup>24)</sup> は、推定される治療効果への影響という観点から、歴史的に推定における仮定は結果が保守的となること、すなわち試験治療に有利に働くバイアスがないよう保証することを重視してきたと述べている。彼らは、多様なステークホルダーの意思決定に当たって試験で得られた結論が有意義であり役立つように、生物学的にもっとももらしい仮定を置くことの必要性を主張している。

なお、より厳密で数学的な議論は本報告書では取り扱わず、戦略ごと一般論を記載する程度に留める。

Ionan ら<sup>25)</sup> によれば、試験の経験が増加することで、例えば最近発行された慢性副鼻腔炎の疾患ガイドラインのように、規制当局及び治験依頼者が、特定の中間事象に対する疾患特有の戦略を提案することもありうると述べられている。2022年7月時点における FDA・EMA で estimand の概念を取り入れた主な疾患ガイドラインの一覧を5.10節に参考情報として記載した。

### 5.3.2 治療方針（Treatment policy）戦略

Mallinckrodt らの報告<sup>13)</sup> には以下の記載がある。

一般的に、治療方針戦略を用いて推定された治療効果は中間事象による影響を含んだものとなる。割付けられた治療開始後の他の治療の影響等を加味した上で、割付けられた治療を開始することの治療効果に関心がある場合には、それらの中間事象に治療方針戦略を適用することは有用と考えられる。また、実際の中間事象の発現状況も踏まえて得られた治療効果を解釈することが重要である。

試験治療群の被験者において有効性が減弱するような中間事象（例：忍容性の問題によ

る治療中止)が発現し、対照群の被験者において有効性が補強されるような中間事象(例:効果不十分によるレスキュー薬の開始)が発現されやすいような状況では、実際の中間事象発現の結果として、治療効果が薄まるような推定値が得られることになる。また、治験薬の有効性が不十分かつ忍容性が低いことによってレスキュー薬が開始されるような状況では、試験治療単独での有効性が不十分という事実が見えづらくなることが起きうる。

レスキュー薬の使用を治療方針ストラテジーで対応する場合には、可能性のある薬剤の種類、用量、開始時期を考慮に入れた上で、その主要変数に対する効果の程度、ならびに想定される使用頻度をあらかじめ検討しておくことが重要である。その他の中間事象(試験治療の中止や併用治療の変更等)についても、それぞれについて考えるシナリオをあらかじめ検討しておくことが望ましく、場合によっては同一被験者が複数の中間事象を発現するシナリオを想定することもある。

治療方針ストラテジーに関しては ICH E9における ITT 原則の観点で議論されることがある。一つの議論としては、「ITT の原則によるこれまでの解析とすべての中間事象に治療方針ストラテジーをあてはめたときの解析は等価であるか?」がある。

Ratitch ら<sup>24)</sup> は、“Broad treatment policy strategy”として起こりうるすべての中間事象に治療方針ストラテジーを適用する状況を考察した。ここでは Mallinckrodt, Molenberghs and Rathmann<sup>27)</sup> の引用として broad treatment policy strategy は「純粋な ITT アプローチ」として“ITT の原則”を厳格に適用した治療効果が推測されるためしばしば実用的に有用性を評価できる方法と主張している。

ITT の原則を適用した試験の治療効果については、例えば Lachin ら<sup>28)</sup> のように『ITT 解析は、臨床的有用性に関連する疑問に対して最も現実的で偏りの少ない答えを与える』といった理解をされることがある。これに対して、Ratitch ら<sup>24)</sup> は前述した Broad treatment policy strategy というアプローチを用いて次のように主張している。

すべての中間事象に治療方針ストラテジーを使う方法は単純と思われる一方、試験の実施、統計解析、サンプルサイズ設計、費用、及び結果の解釈が複雑であることを覆い隠してしまう。(中略)すべての中間事象に適用した場合、受けた(又は受けなかった)あらゆる治療が関心のある治療レジメンに含まれるため、関心のある治療レジメンは幅広くなる。このような漠然とした定義では、推定の対象となる治療効果が明確でなく、試験治療との因果関係が弱まり、単にランダム化から結果観察までの因果関係を示すのみとなり、意味のある比較とならない可能性がある。(中略)Broad treatment policy strategy を用いた試験デザインは実臨床を反映した結果が得られると期待されることが多いが、リアルワールドデータを用いた市販後の試験等の例外を除き、治療条件、対象集団は実臨床と差があると考えられる。(中略)実臨床では、盲検化とプラセボは利用されないが、試験では被験者はプラセボ群へランダム化される可能性がある。プラセボ群では服薬率を低下させる可能性があるため、試験治療群の見かけの奏効率が上昇し、プラセボ対照の盲検試験で治療方針ストラテジーを用いた estimand による結果の一般化可能性には疑問がある。可能



な限り実臨床に近くなるよう試験がデザインされた場合でも、ランダム化は必ずしも薬剤の投与、処方、又は支払いといったステークホルダーにとって関心のある真の疑問と同等であるとは限らない。この estimand による評価を有用なものとするためには、試験で患者が従う治療レジメンが臨床的に意味のある治療評価に関連し、「リアルワールド」と同様の方法で追跡されるという仮定が必要である。一般に、broad treatment policy strategy によるアプローチは、アドヒアランス及びレスキュー治療開始の決定が実臨床へと一般化できる、pragmatic real-world デザインで最もよく評価される。治療方針に関連する試験目的は、承認後の段階、特に医療技術評価で重要性が増す可能性がある。

Akacha ら<sup>23)</sup>でも『本当に治療方針ストラテジーで中間事象に対処する estimand に関心があるのであれば、有用性に関する試験、すなわち pragmatic study のようなリアルワールド試験デザインが適している。』と同様の主張がされている。

### 5.3.3 仮想 (Hypothetical) ストラテジー

Mallinckrodt らの報告<sup>13)</sup>では次の内容が記載されている。仮想ストラテジーは、ある一定の理想的な条件下、例えば被験者が治療を遵守した場合に期待される治療効果に主たる関心がある場合等に適用される。言い換えると、関心のある薬剤の有効性を、忍容性又は治療の遵守状況と区別して評価したいような場合には仮想ストラテジーの適用が有用である。実際には、すべての被験者が理想とする条件を完全に満たすように臨床試験を計画及び実施することは不可能である。したがって、仮想ストラテジーでは、統計モデルや観測値からのサンプリング等に基づき、仮想している状況での治療効果を推測する必要がある。適切な統計モデルを選択するためには、関心のある仮想条件を明確かつ詳細に規定すべきであり、そのモデルに用いられる仮定とその妥当性を整理する必要がある。

関心のある仮想条件を詳細に定義する際には、通常、複数の中間事象に関して様々なシナリオを考慮する必要がある。例えば、最初に割り付けられた治療の早期中止とレスキュー薬の使用を考える場合であっても、実際には、割付けられた治療を中止した後に無治療となるケース、割付けられた治療を中止した後にレスキュー薬のみで治療するケース、割付けられた治療にレスキュー薬を上乗せするケース、割付けられた治療にレスキュー薬を上乗せし、その後割付けられた治療を中止するケース等、様々なシナリオが考えられる。想定される中間事象の組み合わせの中で、関心のある疑問に即した仮想条件を規定する必要がある。

Mallinckrodt らによれば仮想ストラテジーでは、想定しているシナリオの下での推定値が得られていることに注意する必要があると述べている。特にそれは治療レジメンの設定に依存するところがあるとも記載されている。Training Material Module 2.3でも同様の注意があり、臨床的疑問を反映するようなシナリオであることの必要性が述べられている。このストラテジーでは、多くのシナリオを検討することが可能だが、全てのシナリオが規制当局の関心となる臨床的疑問を反映している状況とは限らないと言える。と述べている。



Darken ら<sup>29)</sup> は、ICH E9(R1)の定義にあるように、治療を受け治療実施計画書を遵守したと仮定する仮想ストラテジーを用いた estimand について、かつては“有効性の” estimand として記載されており、製薬企業、医師、そして患者は治療の効果として興味を持っていたのではないかと記載されている。また、Darken らはエンドポイントが連続かつ再帰的なイベントな場合に、仮想ストラテジーと複合変数ストラテジーを組み合わせた estimand の要素としてどのように定義すべきか検討し、また数値シミュレーションでの結果を示している。

なお、死亡等の終末事象に関する中間事象が発生しなかった場合を想定することの妥当性の説明は困難であり、通常、仮想ストラテジーで対応することは妥当ではない。ただし、死亡がわずかしき起きない、又は死亡が治療対象の疾患と無関係に（確率的に）発生する場合には、目標とする治療効果の推定値を得る目的で、仮想ストラテジーと同様の欠測値の取り扱いを適用することが許容される可能性がある<sup>24)</sup>。

#### 5.3.4 複合変数（Composite Variable）ストラテジー

Ratitch ら<sup>24)</sup> 及び Mallinckrodt ら<sup>13)</sup> の内容を要約した。

複合変数ストラテジーが適用される中間事象の例としては、ICH E10の2.1.5.2.2節で紹介されている早期離脱（試験治療の中止）やレスキュー治療の利用が考えられる。すなわち、治療を変更する必要が生じたことが試験のエンドポイントとなる場合である。このストラテジーが用いられる一般的な例として、例えば、ベースラインからの改善に一定の閾値が必要な場合に、臨床的指標又は患者報告アウトカムに基づいて、レスポnder／ノンレスポnderを変数として、中間事象の発現をノンレスポnderとして取り扱う例がある。

Ratitch らによれば、複合変数ストラテジーは中間事象が死亡や重篤な事象のように、明確に悪い結果であるときには適切であるとしている。一方で、試験治療の早期中止等、明確に悪い結果と断定できないような中間事象について用いることは問題となる可能性を指摘している。例えば、試験治療の早期中止が、患者が試験に参加し続けることが難しいことで発生したのならば、この事象をもって治療失敗というのは理にかなっているとは言えないと述べている。また、Ratitch らは、複合変数ストラテジーの結果を解釈し、実臨床に外挿するためには、試験における治療へのアドヒアランスが実臨床と同じであることを仮定していることを述べている。それゆえ、ランダム化プラセボ対照盲検化試験では、治療方針ストラテジーと同様に解釈に注意が必要とされている。

Keene ら<sup>30)</sup> は、順序カテゴリと複合変数ストラテジーを利用した中間事象との合成変数について検討を行っている。二値変数よりも順序カテゴリ変数となることでより多くの情報（複数の改善段階等）を変数に含めることができる。解析手法としては、比例オッズモデルが共変量調整を行うことができる適切なアプローチであると述べている。一方で比例オッズモデルの仮定であるオッズ比の比例性をおくことが適切かどうかはしばしば課題となると述べられている。Keene らはプロットでの表示や、二値での分割をいくつか行い

それぞれのオッズ比を計算することが、仮定の頑健性を示す上で有効と述べている。

また中間事象の発現と連続変数との合成変数として扱うことも可能である。連続変数の扱い方の一つとして、Keene らはカテゴリ化が可能な場合はカテゴリ化し、離散的な場合には中央値やトリム平均を用いて二値変数に変換する方法を述べているが、連続変数そのものからは情報が失われることへの危惧が述べられている。

Mallinckrodt らは中間事象が起こった患者の変数に最悪値を代入して解析する方法を示しているが、通常観測される値より極端に小さい値となる場合（例: 0-100の尺度で、60-80が平均的な観測値であるのに最悪値として0を代入する）には、分散が過大になる問題点を指摘している。これについては仮想ストラテジーで触れた Darken らの報告<sup>29)</sup> も参照されたい。

Ratitch らは複合変数ストラテジーを用いたことによるサンプルサイズ設計の影響について、「連続変数を二値アウトカムにすることで生じる感度の低下によるサンプルサイズの増加」、「順序変数や連続変数において中間事象の発現を治療失敗として何らかの値を代入する場合の、中間事象の発現割合によるサンプルサイズの変化」を指摘している。

### 5.3.5 治療下（While on treatment）ストラテジー

Ratitch ら<sup>24)</sup> には以下の記載がある。

『治療下ストラテジーを適用する場合には、要約すべき観察値が時間に依存しない尺度であることが要求される。このアプローチは、中間事象が発現するまでの被験者の反応を、その後の反応を考慮しないで反映することができ、追跡期間の長さが異なる被験者を同等に取り扱うことができる。適用例としては以下が挙げられる。

- ほぼ線形の勾配を仮定できる連続値の変化率（例: 肺機能の FVC 又は FEV1 測定値の変化率）
- 発現率が時間経過に関わらず一定であることが仮定できる再発イベントの評価（例: COPD の悪化）
- 安全性解析（例: 単位曝露あたりの有害事象の割合）

特定の時点での治療効果に関心がある場合には、治療下ストラテジーは不適切である。関心のある評価時点で値が観察されなかった場合に、治療下での最後の観察値を代入する解析手法と治療下ストラテジーは同等でない。治療下ストラテジーのアプローチでも同じ推定値を生成するが、観察が異なる時点で行われたことを反映して解釈する必要があるため、特定の時点での治療効果に関心がある場合にはこのアプローチは臨床的に意味を持たないであろう。』

Fletcher ら<sup>9)</sup> は上記の仮定において治療下ストラテジーは妥当であると述べた上で、推定の問題として患者を重みづけして扱う方法を挙げている。中間事象の発現までの期間を

用いてアウトカムをモデル化する際に、逆分散により重みを推定するか（推定値をより長期の追跡調査を受けた被験者に偏らせる可能性がある）、又は被験者ごとに等しくするか（被験者ごとの情報量の違いにより分散が大きくなる可能性がある）の選択と、背後にある仮定を指摘している。

### 5.3.6 主要層（Principal stratification）ストラテジー

ICHE9(R1)は、関心のある対象集団が治療不遵守等の中間事象が発生しないであろう主要層となる場合があることに言及している。Bornkamp ら<sup>31)</sup> は、興味のある患者層が、仮に試験治療群へ割付けられた時に試験治療に対する忍容性がある患者である例を挙げている。以下に内容を引用する。『このケースでは、試験治療に割付けられた患者については試験結果から知ることができる。他方、対照群に割付けられた患者については、試験治療に対する忍容性は観測不能であり、従って正確に知ることができない。しかしながら、興味のある患者層は、どちらの群に割付けられても忍容性がある患者とすることができるだろう。』

Bornkamp らは続けて、主要層ストラテジーを用いた estimand について今までの試験であまり利用されておらず、議論も少ない理由として以下の2点を挙げている。

- 試験集団全体のうち確実には特定できない患者層に関連するため（すなわち、主要層の一部又はすべての患者について変数が観察されず、欠測データとなる）
- 関心のある患者層における、治療群間のベースラインでの比較可能性を担保するためにはランダム化では解決できない問題が生じるため

主要層ストラテジーを用いて治療効果を推定するためには、通常、強い仮定が必要である。ICH E9(R1)では、特定の臨床上の事象（それに基づいて主要層での効果を推定する）から定義される対象集団を確実に特定するために、run-in period 等効果的なデザイン特性を用いる可能性に言及している。ただし、これらのデザインが利用できるのは特殊な状況に限られる。

Bornkamp らは主要層ストラテジーが利用できる試験の例をいくつか取り上げている。例えば、オンコロジー領域における薬剤での抗薬物抗体（Anti-Drug Antibody; ADA）について、関心のある臨床的疑問が「ADA を発生しても有用な薬物であるか？」としたとき、主要層を試験治療に割付けられたときに ADA となる可能性がある患者とし、中間事象を ADA の発現、主要評価項目が Time-to-Event である場合が述べられている。

また、Ratitch ら<sup>24)</sup> 及び Mallinckrodt ら<sup>13)</sup> は以下の効果に関心のある場合には主要層ストラテジーが有用となる可能性があるとして述べている。

- 関心のある治療（及び対照治療）に耐えることができる患者層の治療効果
- レスキュー薬を必要としない患者層の治療効果
- ランダム化された介入に関係なく感染する患者層のワクチンの発症予防効果

- 無視できない割合の患者が関心のある時点の前に死亡する場合の QOL 等の治療効果
- 特定の間件事象に関連する治療間差と他の治療間差を区別したい場合の補足的解析

2.4.5節でも述べたとおり、主要層ストラテジーを用いた解析と実際の試験での完了例に基づいた解析は異なることに注意が必要である。ICH E9(R1)でも指摘されているとおり、主要層に属する患者はランダム化前の共変量を用いたモデルから予測する必要があるが、どのモデルを用いても予測は不完全である。

この点について、Ratitch らは、解析対象となる集団がランダム化された被験者の一部分であり、ランダム化による共変量の治療群間のバランスが主要層内では完全には維持されず、そのためモデルによって間件事象の発現しうる確率について治療群間のバランスをとって因果関係を維持しようとしていると述べている。さらに、以下の議論を展開している。

主要層ストラテジーは、estimand の設定の中で対象集団を定義する。(中略) 主要層ストラテジーと将来の clinical practice の関係も考慮する必要がある。医師は治療を決定する前に、患者が臨床試験の集団のプロファイルと一致するかどうかを判断する必要がある。ただし臨床試験で対象集団を特定できなかった場合、医師は患者がどの層に属するか定義できない可能性がある。その場合、対象集団におけるベネフィットの根拠を示すだけでは不十分であり、患者が対象集団に含まれる確率と、患者が対象集団に含まれない場合のベネフィット (又はリスク) の両方によって説明する必要がある。主要層ストラテジーによって推定される因果効果を、治療を受けること又は治療を実施することのリスク・ベネフィット又は費用 - ベネフィットの評価に変換することは、個人レベル及び集団レベルのいずれにおいても非常に困難な場合がある。

この議論を踏まえて Ratitch らは、主要層ストラテジーが特に有用な場合として、機械的なゴールが設けられている場合や、たとえ患者が含まれる集団を誤特定しても臨床的な負担が小さい場合をあげている。

なお、主要層ストラテジーに関する概要は、Lipkovich らのチュートリアル<sup>32)</sup> も併せて参考にされたい。

### 5.3.7 ガイドラインに記載されていないストラテジー

ICH E9(R1)では間件事象に対応するストラテジーとして5つのストラテジーを挙げている。一方で、『Whether or not the naming convention is used, it is required that the choices of strategy are unambiguously clear once the estimand is constructed.』とあるように、ストラテジーの名称には縛られる必要はないことを示唆する記載があるほか、ストラテジーが5つに限られるという記載もない。

0章で引用した Siegel ら<sup>16)</sup> は “while on treatment” (治療下) ストラテジーに類似したストラテジーとして “while on alive” ストラテジーを提案している。これはオンコロジー領域の終末的なエンドポイントである全生存時間に対応しており、生存期間中の有効性や安

全性の評価として有益であると考えられるためである。

Mitroiu ら<sup>17)</sup> による2017年11月に実施した調査では、5つの戦略に分類できない (“Other”) 戦略が用いられていると考えられる試験があると報告された。なお、“Other” 戦略の半分以上は estimand に対応する解析手法として PPS 解析を行うことに相当し、その他にも complete case や available case 解析が含まれていた。

ガイドラインに記載されている戦略を用いる際にも、そうでない場合にも中間事象にどのように対応するか、及びその対応が治療効果にどのような影響を与えるかを明示することが重要と考えられる。

## 5.4 Estimand の設定

はじめに Training Material Module 2.3の記載から estimand の設定に関する議論をいくつか抜粋する。

- 臨床的疑問に対する治療について、基礎治療、併用治療、後治療、治療スイッチ等を含む治療条件を明確に特定し、その他の中間事象と区別する。
- 意思決定において信頼できる治療効果の推定値を導出できるようにする。Estimand が固定される前に信頼できる推定値が得られることを合意すべきである。
- Estimand の設定のプロセスを省略あるいは簡略しすぎた場合、試験目的、試験デザイン、データ収集、統計解析の間で齟齬が生じるリスクがあるので注意すべきである。

Ratitch ら<sup>24)</sup> は、estimand における治療は治療レジメンを扱っており、これまでの割付けられた治療以上の内容を含みうるとしている。併用可能な薬剤や療法等について、治療レジメンとして許容するのか、単に倫理的配慮からの許容であり本来治療レジメンとしては許容しないのか、といった区別が必要になると指摘している。

患者集団は estimand を構成する要素の1つである対象集団 (population) が、関心のある臨床的疑問に対応する対象となる患者集団 (target population) として設定される。Fletcher ら<sup>9)</sup> では ICH E9 と ICH E9(R1)での設定の差異について以下のように述べている。

ICH E9では、解析対象集団を定義し、これにより対象となる患者が特定され、複数の評価項目の推定根拠となっていた。一方、ICH E9(R1)では、それぞれの estimand に対し、解析対象集団に含める患者と中間事象に対応するための戦略を反映した観察値の両方を特定したデータから成る。

これに対応して、TransCelerate のプロトコルモデルドキュメント<sup>5)</sup> では、解析対象集団の項に上記の被験者集団 (Participant Analysis Set) に加えて、データポイントセット (Data Point Set) という概念が記載されている。データポイントセットに関連する記載について、上記モデルドキュメントのインストラクションから引用する。

データポイントセットは解析に含まれる被験者ごとの中間事象の発現状況や時期を考慮したデータポイントの集合として定義される。解析データセットは、被験者集団に属す



る患者ごとに解析に含めるべきデータポイントのセットとして定義され、すなわち被験者集団とデータポイントセットを組み合わせたものとなる。目的の章で estimand が明確に定義されているかどうかに関わらず、統計解析に含まれる被験者の集団、すなわち「被験者解析対象集団」を定義すべきである。

## 5.5 試験デザイン及びデータ収集

試験デザインにおいては、試験の目的を反映する estimand と整合させる必要がある。Training Material Module 2.4 (p.147)では、ICH E6 (筆者注：日本では「医薬品の臨床試験の実施の基準に関する省令（平成9年厚生省令第28号）」第7条、「「医薬品の臨床試験の実施の基準に関する省令」のガイダンス（令和元年薬生薬審発0705第3号）」によれば「中央薬事審議会答申（平成9年3月13日中薬審第40号）」の10も適宜参照される）に治験実施計画書に記載すべき試験デザインの重要な内容が一覧となっていることに触れたうえで試験デザインを検討する場合には、試験目的と estimand を整合させる必要があるとしている。

Training Material Module 2.4では試験を中止する患者が多い場合、ストラテジーや estimand の関連性は揺らがないが、推定精度は悪化することが指摘されている。また、中間事象の性質や発生頻度、発生時期に関するデータも含め、関心のある estimand と整合する解析手法をサポートするデータは収集するように努めるべきとも記載されている。

Mallinckrodt ら<sup>13)</sup>では次の議論があった。例えば estimand としてすべての中間事象に対して厳密に治療方針ストラテジーを適用する場合、解析で定められた時点までのデータは、試験中止の場合を除き追跡する必要が発生する。したがって、有害事象等で治療中止となった場合でも追跡する必要があり、オペレーション上困難が伴う可能性がある。また、中間事象の取り扱いによっては、治療中止の理由を収集する必要がある場合も存在する。

なお、ICH E9(R1)では正確な治療中止理由の収集において、十分な情報取得のための前向きな計画は、中間事象と欠測データを区別するのに役立つとしている。例えば、「追跡不能」ではなく、「有効性欠如による治療中止」と記録された方がより正確だろうと述べている。また、Qu ら<sup>33)</sup>も正確な治験薬中止理由の収集は、中間事象に対するストラテジーの検討や欠測値の補完方法において重要となる場合があるとし、現在、phuse working group にてより正確で詳細な治験薬中止理由の収集を目的に CDISC を含めた標準 CRF の見直しを行っていると述べている。更に、Training Material Module 2.4でも、CRF の修正について触れている。

解析手法の選択に関する一般的な議論は本報告書に記載されており、さらに専門的な議論に踏み込むと技術的な内容が多くなるため割愛する。

## 5.6 仮定の特定及び感度分析の設定

主解析における仮定を特定し、仮定を検証するための適切な感度分析を設定する際の議論を以下に述べる。Fletcher ら<sup>9)</sup>は、感度分析及び補足的解析の意味をどのように estimand



のフレームワークに導入するかは課題であったと述べられている。それによれば、補足的解析を同一の estimand に合わせるか、あるいは治療効果の理解を深める別の estimand に合わせるか不明であったとしている。また、彼らは ICH E9(R1)での定義を引用した上で、感度分析の例として tipping point analysis<sup>34)</sup> を挙げ、補足的解析が必要な例として『primary estimand が2つの治療群間の連続変数である主要評価項目のベースラインからの平均変化量により要約される場合に、有効性の評価として奏効例(レスポnder解析)を検討する』という主要評価項目に関連のある別の評価項目を用いた異なる estimand を必要とする場合を取り上げている。

また、Mallinckrodt ら<sup>13)</sup> は以下の内容を述べている。

主とする推定量を裏付ける統計的な仮定は、estimand で明記される（筆者注：ここでは estimand に紐づく estimator（推定量）に統計的な仮定があることを述べていると考えられる）。そして、主とする推定量によって算出された推定値が、その仮定からのずれに対して頑健であるかを確認する目的で、同じ estimand を対象とした一つ以上の感度分析を事前に規定すべきである。例えば、tipping point analysis 等のように、統計的又は臨床的な有意性の解釈が変わってしまう仮定からのずれの程度で頑健性を特徴付ける方法がある。

感度分析を計画し、実施する際には、主とする推定量の複数の仮定を同時に変更しないことが推奨される。複数の仮定を同時に変更した場合には、結果の違いについてどの仮定が原因となっているかを特定することが困難になるからである。したがって、主とする推定量における複数の仮定を同時に変更した下での解析結果と、主とする推定値を単に比較するのではなく、感度分析の結果が主とする推定値とずれた場合に、単一の仮定を逐次変えて結果を確認することが望ましい。また、複数の仮定を同時に変更する解析が必要である場合も、ケースごとに考慮すべきである。

## 5.7 欠測データ

ICH E9(R1)の用語集では、欠測データを次のように説明している。『規定した estimand の解析に対して意味があると考えられるが、収集されなかったデータ。存在しないデータや中間事象の発現により意味があると思えないデータとは区別されるべきである。』ICH E9(R1)及び Training Material では中間事象と欠測は異なる概念であること、及び両者の区別について明確に文書に記載し、混同しないよう注意喚起がなされている。

Ratitch ら<sup>24)</sup> は、少なくとも主要評価項目については中間事象発生後のデータを収集することが標準的な方法ではないかと述べているが、中間事象発生後のデータの収集については補足的解析を含めてどのような estimand での解析を計画しているのかを鑑みて決定する必要があるだろう。しかしながら最大限データ収集に努力を払ったとしても、欠測データを回避することはできない。ICH E9(R1)の欠測データの定義に基づくと、estimand を構成する5つの要素（特に採用した中間事象とストラテジー）によっては、収集したデータは同じでも欠測データとなる場合とそうでない場合があることにも注意が必要である。

例えば、治療方針ストラテジーの場合、loss-to follow-up 等によりデータ収集ができなかった場合が欠測データに該当する。一方で治験薬中止の中間事象に対し、複合ストラテジーを適用し、変数をレスポnder/ノンレスポnderで定義する場合には、欠測データとは見なされない。同意撤回や loss-to follow-up 等の理由により必ずしもデータが収集できるとは限らないので、欠測データが発生することを考慮に入れた解析手法（mixed model for repeated measures (MMRM) や多重代入法等）を検討する必要がある。なお、欠測値の補完方法の一つとして近年、試験治療中止等の中間事象の発現後も治験に参加している被験者（retrieved dropout）のデータも利用した手法が提案されている<sup>35)</sup> が、Fletcher ら<sup>9)</sup> は現時点では使用経験は限られたものであると指摘している。

また、Training Material Module 2.4 には注意事項として試験中止した被験者の人数は、estimand や特定のストラテジーにおける因果関係そのものには影響しないが、推定値には影響がありうるとし、注意深く検討するよう述べている。

欠測データについてはデータサイエンス部会でこれまでも多く議論されてきており、2019年に報告された「臨床試験における estimand と感度分析に関連する文献調査」<sup>36)</sup> の 3.2節 “データの取扱い及び欠測データに含めて検討されている問題” でも深く議論されているので参考にされたい。

## 5.8 試験の目的に応じた estimand

ICHE9(R1)に従うと、少なくとも検証的な試験での estimand の設定は必須となるが、検証的な試験以外での estimand の設定も有用であろう。また、ICH E9(R1)A.4にはメタアナリシスや安全性・有効性の統合解析でも estimand を構成することで適切な評価ができるとされている。本節では、優越性を検証する試験以外で estimand を設定する際に考慮すべき事項を取り上げる。ただし、Fletcher ら<sup>9)</sup> が指摘したように、非劣性/同等性試験及び安全性評価等における estimand、また解析方法として Per-protocol 解析等においては今後更なる検討が必要であろう。

### 5.8.1 PK/PD 解析における estimand

Akacha ら<sup>37)</sup> は、用量曝露 (PK) /曝露反応 (PD) 解析について estimand のフレームワークに基づいて議論している。以下に内容を要約する。

PK/PD 解析で標準的に用いられる手法である two-step アプローチでは、PK モデル及び PD モデルを構築した後に PK/PD プロファイルのシミュレーションを実施する。このとき、実際の投与量及び投与時間を用いた解析を行うことで、規定の治療レジメンを遵守しない（ノンアドヒアランス）患者について、アドヒアランスを考慮した薬物濃度が導き出される。PK/PD モデルを用いて、完全にレジメンを遵守した患者の PK プロファイルを多数シミュレーションし、個々の患者に対する反応プロファイルを算出することができる。

このような状況下では、試験治療群と対照群の群間プロファイルの差は、estimand のフ

レームワークで考えると「すべての患者が割付けられた治療に従う」といった仮想的な治療効果として考えることができる。この手法は割付けられた治療の有用性に関心があり、また、すべての患者が割付けられた治療に従った場合に、もっとも効率よく推定が可能である。

さらに、特定の有害事象（例えばアドヒアランスを低下させた有害事象）を中間事象として、その中間事象に対する直前の薬物濃度の影響をモデル化し、前述の two-step アプローチを拡張した PK/PD 解析を行うことで、特定の中間事象に対して治療方針ストラテジーを使った治療効果を算出することも可能である。治療方針ストラテジーを用いた estimand の下で推定される治療効果は、新しいレジメンについて承認審査を行う際に規制当局から求められるかもしれない。

### 5.8.2 早期相における estimand

Mallinckrodt らの報告<sup>13)</sup>には以下の記載がある。

臨床データパッケージの中でピボタル試験となりうる抗がん剤等の早期相試験は規制当局を主要なステークホルダーとして試験を組むことがある。これらの試験では検証的試験において典型的な primary estimand を採用する可能性がある。一方で、多くの早期相試験は Proof of Concept や後期相試験の用量設定のために計画され、主に企業の意思決定のために用いられる。この目的で実施される試験の estimand は、検証的試験で焦点となる可能性がある治療方針には重点を置いておらず、最初に割付けられた治療に焦点を当てていることが多いが、検証的試験で用いる estimand を早期相試験の secondary estimand として評価し、後期相試験を計画するためのデータを収集することもできる。また、早期相試験で対照群を設定しない単群デザインを採用する場合、estimand は単群での推定や事前定義された（もしくはヒストリカルな）基準との比較に焦点を当てる必要があり、同時対照試験と比較した治療効果の推定が難しくなるが、中間事象の取り扱いについて考慮すべき事項はほとんど同じである。

Ionan ら<sup>25)</sup>は第 II 相試験と第 III 相試験において primary estimand が異なる場合、estimand のフレームワークがそれぞれの状況の理解に役立つとしている。第 II 相試験の結果を考慮して第 III 相試験のデザインを計画する際に、第 II 相試験の補足的解析として第 III 相試験で用いる estimand を設定すれば、その推定値は第 III 相試験の計画に役立つであろうと述べている。ただし、この補足的解析の結果の解釈には、試験間の関心のある臨床的疑問や試験デザイン等の違いから注意が必要であるとしている。

また Kang ら<sup>38)</sup>は、仮想的な試験として小児集団における抗レトロウイルス併用療法レジメンの一環として新規抗レトロウイルス薬の安全性を検討することを目的とした、HIV を対象とした早期相の単群小児試験を想定し、その試験での estimand について検討をしている。Kang らは、変数を治療開始から24週までの間における生命を脅かすような有害事象の発現有無とし、中間事象としては死亡、24週未満での早期治療中止、妊娠を設

定した。死亡については複合変数ストラテジーとして扱い、他の2つの中間事象については治療方針ストラテジーを採用し、かつ早期に試験を中止した被験者については、中止後には生命を脅かすような有害事象は発現しなかったとして取り扱うことを検討した。なお、Kang らは、Ratitch ら<sup>24)</sup> を参照したうえで、治療中止時点までの検討に関心がある場合、又は要約指標が時間に依存しない尺度（曝露量あたりの有害事象の発現率等）である場合、治療下ストラテジーがより適切となることがあると述べている。

### 5.8.3 非劣性／同等性試験における estimand

非劣性試験及び同等性試験は優越性試験と異なり、治療群間差を過小評価することは、劣らないもしくは同等であるという結果を得やすくし、第一種の過誤確率を上振れさせることにつながる。このことから、これまで非劣性試験と優越性試験とは異なる統計的考察がなされてきた。ICH E9にも、非劣性試験及び同等性試験での留意事項が述べられており、例えば同等性試験で最大の解析対象集団（FAS）による解析を行うと、同等性を導く方向へ偏る傾向があるといった記載がある。ICH E9(R1)では、estimand のフレームワークを用いる立場からさらに掘り下げた議論がされている。Training Material Module 2.3 p.138 には次の記載がある。

- 規制当局の意思決定における問題点が異なるため、非劣性もしくは同等性を示す estimand の設定は、優越性を比較する場合と異なる可能性がある
- 非劣性及び同等性試験では、治療方針ストラテジーの使用（ICH E9 で ITT 原則の下 FAS を使って解析を行うことに相当）は、治療中止やレスキュー薬の中間事象発現後、両方の治療群で同じような反応が現れる可能性があるため、一般的に保守的ではない
- 治療効果差の過小評価につながる中間事象に直接対処するための estimand は構築しうる

Mallinckrodt ら<sup>13)</sup> は、第一種の過誤確率の上振れを避ける目的で、結果的に非劣性試験の estimand が優越性試験とは異なることはあるだろうが、優越性試験と同様に非劣性試験においても最初に臨床的疑問から estimand を設定し、それに基づいて解析手法を決定することが重要である、と述べている。

### 5.8.4 Per Protocol Set と estimand

ICH E9では Per Protocol Set (PPS)を用いることは、治験実施計画書の基礎となっている科学的なモデルを最もよく反映することになる、とする一方で、治験実施計画書の遵守状況が試験治療及び臨床結果と関連している場合には、生じる偏りは重大なものとなるおそれがあることを指摘していた。ICH E9(R1)では estimand のフレームワークを用いて PPS 解析と整合する estimand を構築することができないかもしれないと述べている。また、PPS 解析では、主要層に対する治療効果の推定を得ることができない。例えば主要層を試



験治療に忍容性があり継続して投与することができる患者層とした場合、PPS 解析では異なる治療を受けている同様の被験者集団を比較することにはならないためである。PPS 解析は治験実施計画書違反の患者を除外したものである。また、患者は例えば死亡のような中間事象も引き起こす可能性がある。これら治験実施計画書違反と中間事象は、事象自体は異なるものの、いずれの事象も estimands の要素「変数」に影響を与える可能性があると言える。もし、解析方法との整合性が取れ、PPS 解析の目的を反映したような estimand が設定できる場合には、PPS 解析により更なる見解や解釈が得られることはないかもしれない。

2017 年11月時点における estimand の記載状況を調査した、Mitroiu ら<sup>17)</sup> の論文では、PPS での治療効果を評価した試験を、ICH E9(R1)の5つのストラテジーに当てはまらない“Other”として分類した。

Little ら<sup>39)</sup> は、ランダム化比較試験における Per-Protocol 推定値は、ITT にもとづく推定値より強い仮定が必要になるとも述べている。この理由はランダム化によって解消された交絡が、部分集団を取り出すことで復活する可能性があり、推定される治療効果に影響する可能性があるためである。

#### 5.8.5 Per-protocol 解析による治療効果に対応する estimand

従来の PPS 解析とは別に、Keene ら<sup>26)</sup> は、estimand のフレームワークとして Per-Protocol の治療効果を評価する場合、主要層ストラテジーや仮想ストラテジーを用いた estimand が PPS の代替案になるかもしれないと述べている。

また、Little らは2型糖尿病患者を対象とする PIONEER III 試験<sup>40)</sup> で定義された2つの estimand (“treatment policy”, “trial product”)を取り上げ、estimand の相違点に着目した。

“Trial product estimand”ではすべてのランダム化された患者が全期間にわたって割付けられた治療を遵守し、レスキュー薬を使わないと仮定した下での治療効果を推定の対象とした。Trial product estimand の下では、試験治療の中止やレスキュー薬の使用を中間事象とし、仮想ストラテジーを用いて対応することが決められていた。Little らは trial product estimand の population を Per-Protocol Population として紹介した。Per-Protocol Population という名称では、ランダム化された患者の部分集団である PPS が population としても考えられるが、estimand ですべてのランダム化された患者を対象とすると明記したことで、PPS である可能性を定義から除いている。この患者集団は治験実施計画書において定義された治療と同一のアドヒアランスをもつ患者集団として治療効果が評価可能かもしれないと述べている。

また、Hernán ら<sup>41)</sup> は pragmatic trials における ITT 解析に対する別の解析として per-protocol effect を提案しているため紹介する。ただし、これは患者全員が割り付けられた治療を遵守した時の因果効果と定義したものであり、ICH E9で述べている PPS と対象としている集団が異なることには留意したい。

Hernán らは、pragmatic trial における per-protocol effect を患者全員が治験実施計画書を遵守したであろう場合の治療効果と定義している。Hernán らは per-protocol effect の推定が妥当であるための条件として、loss-to follow-up の違いによる選択バイアスの調整が適切になされていることに加え、試験治療へのアドヒアランスや治験実施計画書に記載されていない併用療法の利用による交絡について正しく調整されていることが必要と述べている。この2つの条件に加えて、per-protocol effect の妥当な推定値を導くための一般的に重要な3つの条件について述べている。3条件のうち2つはデータの収集に関する要請で、臨床的な理由で治療を中止した時点では打ち切るべきではないが、どのような治療を受けているか不明となったら打ち切りとするべきであることを述べている。最後の条件は、アドヒアランスが不十分であることによる交絡は補正すべきという、因果推論的な要請である。詳細な議論は論文を参照されたい。また、per-protocol effect を適切に評価するためには、治験実施計画書に per-protocol effect の正確な定義とその解析の内容を事前に統計解析計画書に規定しておくことが、適切な解析手法の選択と並んで重要であると述べている。治療効果の正確な定義によって、データの収集範囲を定められること、遺伝子、バイオマーカー、臨床上的条件等、患者中心医療に関する部分集団で per-protocol 解析ができること、事後解析との誤解を避けることが利点であると述べている。

なお、本報告書の治療方針ストラテジーに関する議論（5.3.2節）で pragmatic trial において、ITT の原則に基づいた治療効果についてもっとも試験の結果を一般化できる可能性が高いと考えられる<sup>24)</sup> という見解と、患者の試験治療に対するアドヒアランスによって結果が変わりうる<sup>41)</sup> という見解を取り上げた。これについて Hernán らは ITT に基づく解析結果が公表されることで、実臨床でのアドヒアランスが変わり、それによって結果がもはや過去のものになる可能性があり、また試験間の比較についてもアドヒアランスが違う場合には ITT の結果をそのまま比較することは誤った結論を導きかねないと述べている。

Pragmatic trial についてはデータサイエンス部会から2018年6月に「Pragmatic Trials のススメ」<sup>42)</sup> が報告されているので参照されたい。

#### 5.8.6 安全性評価における estimand

本報告書3.4節にて安全性に関する estimand について紹介した。Mallinckrodt ら<sup>13)</sup> 及び Ratitch ら<sup>24)</sup> は安全性評価を estimand のフレームワークで扱う際の留意点を指摘している。その内容は下記のとおりである。

- 治療によるリスクを評価することができる
- 安全性評価では治療下ストラテジーの estimand を適用して評価することが多い
- 有効性と安全性を総合してベネフィット・リスクを評価する際に、有効性と安全性が異なる治療レジメンの estimand に基づいていると問題が生じる可能性がある

上記3つ目の内容について事例を挙げる。有効性がレスキュー薬を含む治療レジメンに



基づいて評価されている一方で、安全性が割付けられた治療のみに基づいている場合、ベネフィット・リスクを直接比較することが困難になるので注意が必要である。

## 5.9 パンデミックの影響評価

2019年末に新たに発生した COVID-19は世界規模で爆発的流行を見せ、臨床試験においてパンデミックに関連する中間事象をどう取り扱うべきか考える契機となった。EMA の Biostatistics Working Party (BSWP)は臨床試験のデータ解析と解釈において、COVID-19パンデミックで生じた治療中止等の中間事象と欠測データの影響を考慮する必要がある、estimand のフレームワークはこの影響分析を明確にするための包括的なアプローチを提供している<sup>43)</sup>。本節では Meyer ら<sup>44)</sup> を参照してパンデミック下で実施された臨床試験の統計的課題と提言について一部紹介する。

パンデミックの影響を受けた試験の場合、主要な試験目的に紐づく primary estimand がパンデミックの影響と交絡していない治療効果を推定できるかが最初の重要な問題となる。可能な限り当初の試験目的を維持し、パンデミックに関連した事象による交絡がない治療効果を推定すべきである。ただし、予期しないパンデミックにより計画された estimand への影響が避けられない場合もあるので、その影響を適切かつ厳密に説明するためにパンデミックに関連する中間事象を定義する必要がある。また探索的な目的で追加の estimand を設定し、パンデミック下にある特定の状況又は部分集団における治療効果を推定することも可能である。

パンデミックの影響により試験の実施やデータ収集方法が変わる可能性がある。例えば、治験実施施設にて医療従事者によって投与されていた薬剤に対して、自宅での自己投与を許容する場合があるかもしれない。これらのオペレーション上の緩和策は、データ解析においてこれらの緩和策が与える影響を十分に検討した上で適用すべきである。代替するデータ収集方法の妥当性や標準的な評価方法との交換可能性については慎重に検討する必要がある。

パンデミックを考慮して estimand を更新した場合、その estimand と推定量及び欠測データの処理が一貫することを確認しながら、計画された有効性と安全性の解析方法を変更する必要があるかどうか注意深く検討すべきである。またパンデミックの影響を完全に説明するためには追加の感度解析や補足的解析が必要になる。

試験の実施、データ収集及び解析に対するこれらの変更は、必要に応じて統計解析計画書や総括報告書に文書化する必要がある。一部の変更においては治験実施計画書の修正に加え規制当局との協議が必要になる場合がある。

また2022年2月にはウクライナでの戦争が開始され、東欧に治験実施施設を抱える多くの製薬企業において、臨床試験の計画と実施に多大な影響が出ている。2022年4月にはEMA が戦争に影響を受けた事象を取り扱うのに estimand のフレームワークを利用すべきと企業に提言しており、その際には COVID-19パンデミック時のガイダンスを参照するよ

う提唱している<sup>45)</sup>。

データサイエンス部会からは2020年5月に「COVID-19が治験に及ぼす影響について—統計解析・データマネジメントの視点—」<sup>46)</sup>という声明が出ており、estimandに関係する点を列記した上で統計解析に関する問題への指針を提示している。また2022年7月には「ICH E9(R1)の理解に役立つ因果推論」<sup>7)</sup>が報告されており、COVID-19に関連する中間事象の取り扱いが4.2節に詳しく述べられているので参照されたい。

## 5.10 Estimand に言及している主な疾患ガイドライン

以下に参考資料として2022年7月時点において、estimandに言及しているEMA及びFDAが発出している主な疾患ガイドラインを列挙するので参考にされたい。

- EMA
  - Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus<sup>18)</sup>
  - Guideline on the clinical investigation of medicines for the treatment of Alzheimer’s disease<sup>47)</sup>
  - Concept paper on the need to develop a reflection paper on development of medicinal products to prevent and treat acute kidney injury<sup>48)</sup>
  - Reflection paper on regulatory requirements for the development of medicinal products for chronic non-infectious liver diseases (PBC, PSC, NASH)<sup>49)</sup>
  - Guideline on clinical investigation of medicinal products in the treatment of epileptic disorders<sup>50)</sup>
  - Guideline on quality, non-clinical and clinical aspects of medicinal products containing genetically modified cells<sup>51)</sup>
  - Guideline on registry-based studies<sup>52)</sup>
  
- FDA
  - Eosinophilic Esophagitis: Developing Drugs for Treatment<sup>53)</sup>
  - Acute Myeloid Leukemia: Developing Drugs and Biological Products for Treatment<sup>54)</sup>
  - Chronic Rhinosinusitis with Nasal Polyps: Developing Drugs for Treatment<sup>55)</sup>
  - Crohn’s Disease: Developing Drugs for Treatment<sup>56)</sup>
  - Ulcerative Colitis: Developing Drugs for Treatment<sup>57)</sup>

## 6. オンコロジー領域における estimand の意義

### 6.1 はじめに

本章では、オンコロジー領域における estimand の検討について叙述する。オンコロジー領域に絞る理由としては、以下の2点による。1点目として、estimand に関する議論や検討は歴史的経緯からプライマリーケア領域を主たる対象として行われており、オンコロジー領域に関してはガイドライン等でも例示が少なく、EPSIS や ASA のワーキンググループにて議論が重ねられている最中であることが挙げられる。日本語による報告書として、オンコロジー領域についての estimand に関する検討状況を紹介することは、当該領域の試験関係者に関して有益と考えられるからである。2点目の理由として、オンコロジー領域の試験デザインは、プライマリーケア領域とは状況が少々異なると考えられるためである。特に、比較の後期相のオンコロジー領域では、イベント発現時間 (Time-to-Event; TTE) に関する生存時間解析が代表的であり、評価項目や評価に係る解析手法には慣習的なコンセンサスが存在していると報告されている<sup>59)</sup>。

TTE による評価項目は、終末事象である死亡までの時間を扱う全生存期間 Overall Survival (OS) がハードなエンドポイントとして存在し、ゴールドスタンダードなエンドポイントとして当局にも認知されている<sup>16)</sup>。また、OS のサロゲートエンドポイントとして死亡と増悪等の事象を組み合わせてイベントと扱う複合エンドポイント (PFS, Disease-Free Survival; DFS, Event-Free Survival; EFS) を評価することが一般的である。

生存時間解析では、生存曲線の推定として Kaplan-Meier (KM) 推定量を用いることや、群間差 (治療効果) の検出にはログランク検定や比例ハザード性を仮定した Cox 比例ハザードモデルを利用してハザード比を求めることが一般的である。また、治験中止、同意撤回や画像評価の複数回にわたる欠測等測定されないデータは打ち切りとして取り扱われることが多い。

標準的なエンドポイントや解析手法があることは、estimand の一部の要素に相当するものが検討済みであるという見方ができる。しかしながら本報告書の2章等で述べたように、estimand の設定は、「試験目的」、「関心のある臨床的疑問」を明確にしたうえで、estimand の要素を検討し推定したい治療効果を定義する。解析手法は estimand を設定したあとに検討されるべき内容である。これまでの方法では経験的に解析方法を定めることで推定される治療効果を定義していたことに相当するので、試験目的と関心のある臨床的疑問に適するかといったことは十分検討されていなかったという点で estimand のフレームワークを使った検討とは開きがある。さらに、打ち切りや中間事象に相当するイベントの取り扱いに代表される解析上の取り扱いについて、想定する仮定が必ずしも明確にされておらず、論文報告でも不十分な記載となっていると主張した報告<sup>60)</sup>もある。また、Rufibach<sup>59)</sup>は ICH E9 (R1)発出前に実施された GALLIUM 試験における主解析に対して計画された「感度分析」について検討したところ、primary estimand とは異なる estimand を用いた補足的解析にあたる内容が含まれており、ICH E9(R1)の意味での感度分析は1つのみであつ

たと報告した。ICH E9 (R1)では感度解析は同じ estimand の下で仮定を変えたときの頑健性を検討するものである。具体的な感度分析の例は、Rufibach, Sun や Sigel<sup>16)</sup> らの報告を参照すると、比例ハザード性や無情報打ち切り (non-informative censoring) の仮定に対する検討や、主治医判定の PFS と中央判定の PFS を入れ替えた解析が挙げられている。

さらに、近年の免疫チェックポイント阻害薬及びがんワクチン等の治験では免疫機能を介して効果を発現するため、効果発現まで一定の時間を要すること、効果が長期的に持続するという特性を持つことから、一般的に利用されてきた比例ハザード性を仮定した解析は妥当であるといえない治験も増加していることも報告されている<sup>58)</sup>。こうした状況も踏まえて estimand のフレームワークに従って、試験目的、関心のある臨床的疑問に適合する estimand に沿った試験となるよう評価項目、中間事象の取り扱い、及び解析手法を明確にし、付随する仮定を整理することには意義があると思われる。

本章では、Rufibach<sup>59)</sup>、Casey ら<sup>61)</sup>、Sun ら<sup>22)</sup>、Manitz ら<sup>62)</sup>等を参照し、オンコロジー領域における estimand のフレームワークに関して概説する。加えて固形癌に関する既存の試験について、事後的に estimand のフレームワークを用いて解釈した例や、中間事象と関連の深い事象として治療変更、打ち切りに関する近年の考え方についてまとめて示す。なお、オンコロジー領域における estimand については、EPSPI 及び ASA のコラボレーション working group である、Oncology estimand working group (<http://www.oncoestimand.org/>) が積極的に検討を進めている。文献や各種 webinar 資料も多数掲載されているため、関心がある方はぜひ参照し、理解を深めていただきたい。

## 6.2 Estimand の要素

本節では、estimand の5つの要素ごとに、留意すべき点を要約して示す。ただし、既に本報告書で記載されている事項と同様のものは省略する。また、その他の中間事象は中間事象の扱いに依存し他の要素と比較して内容が多いため、以下のまとめでは触れず、6.3節以下に示す。

### 6.2.1 関心のある治療

Sun らは血液がんの臨床試験について、評価項目や収集するデータに加えて臨床的に意味のある治療効果や治療効果の解釈によって特徴づけられるとしたうえで、標準治療には多くの場合後治療が存在し、手術も治療戦略の重要なものと位置づけられているため、関心のある治療の要素を定義する際に課題となると述べている。また、一般に倫理的な側面から治療変更や併用薬の追加を不可とすることは難しいため、関心のある治療が治療の変更を含むかは検討が必要であると思われる。治療変更と関心のある治療の関係については Manitz ら<sup>62)</sup>の検討を6.4節に示す。

### 6.2.2 対象集団

2.3節に記載のとおりである。

### 6.2.3 変数（評価項目）

TTE を評価することが多い。全生存時間 (OS) であれば、死亡までの期間 (日数, 月数) という正確なエンドポイントとなるが、画像評価等来院単位で行われるものでは、イベントの評価は連続的ではなく離散的な (来院単位の) 評価となることに注意が必要である。また、評価項目間に競合リスクがある場合、一つの評価項目に対応するイベントが発生したことで、他の評価項目ではイベントが発生せず打ち切りとなる場合も存在する (たとえば増悪までの時間 (Time to progression; TTP) を評価するにあたり、増悪とは無関係に死亡した場合)。評価項目として PFS のように複合変数として定義することも、複合変数とはせずに競合する中間事象の1つとして死亡を定義することも可能である。Rufibach<sup>59)</sup> は変数を TTP, 死亡をその他の中間事象として扱う場合、推定したい対象によって、適切な戦略、解析上の仮定や取り扱いが異なることを述べている節を以下に記載する。

### 6.2.4 集団レベルの変数要約

一般的に KM 曲線や KM 推定量を用いて、生存時間の中央値等を算出するほか、通常ログランク検定と Cox 比例ハザードモデルによるハザード比を用いて要約する。ただし、Rufibach<sup>59)</sup> が指摘しているように比例ハザード性が成り立たない時は上記の検定統計量と推定量が直接対応しない (ハザード比は打ち切りの分布に依存する)。この場合の対応として Rufibach は2つの方法を提示している。1つは比例ハザード性の仮定が崩れても頑健な検定および推定量に置き換えることである。もう1つは2段階でのアプローチである。二群のランダム化比較試験であれば、第一段階では、例えば両群で生存時間が等しいという帰無仮説を用いて検定する。帰無仮説が棄却され、かつ点推定値が正しい方向に向いている (試験治療群の生存時間が延長する) か確認する。これが満たされた時試験は成功したと考え、治療効果を検討することになる。

## 6.3 打ち切りと中間事象

打ち切りとは、観察開始後のある時点で、観測が打ち切られることである。たとえば、イベント発現前の試験中止、同意撤回、loss-to follow up は代表的な打ち切りである。また、治験実施計画書等で規定された追跡期間を過ぎてもイベントが未発現である場合 (切断) や、競合するイベントの発生でも発生する。

打ち切りの分類の1つとして、無情報の打ち切り (non-informative censoring) と情報のある打ち切り (informative censoring) がある。無情報の打ち切りとは、試験と関連しない理由で発生した打ち切りである。例えば、被験者が状態と関係なく転居し、loss-to follow up となり、打ち切った場合が考えられる。一般に無情報の打ち切りはランダムな打ち切りと考えられ、打ち切りの有無によらず集団内の同一性が担保される。したがって、無情報打ち切りはハザード比等の一般的な推定量に偏りを生じさせない。このことから無情報打ち切りは関心のある疑問への情報を与えるものではないと考えることができる。また、



Rufibach は事前に定義された観察期間終了に伴う単純な打ち切り (administrative censoring) もランダムな打ち切りであり、推定量に critical な影響を与えないとしている。また、6.2.3 節で取り上げた死亡が中間事象となるイベントの評価に当たっては、仮想ストラテジーと治療下ストラテジーの両方を用いることができるが、両者が推定する対象は異なる。競合イベント又は関心のあるイベントの双方についていずれか早い時点までの治療効果に関心があるなら、治療下ストラテジーが目的に合致すると Rufibach<sup>59)</sup> は述べている。情報のある打ち切りは、効果不十分や安全性の懸念からの同意撤回等、試験に関連した理由による打ち切りが考えられる。情報のある打ち切りが想定される場合、すべての打ち切りが無情報の打ち切りを仮定した推定量や検定統計量に影響を及ぼす可能性がある。Casey ら<sup>61)</sup> は primary estimand において情報のある打ち切りの影響を評価するための感度分析として、打ち切り後のデータを補完する方法があると述べている。

Siegel ら<sup>16)</sup> は, estimand のフレームワークの下では, 打ち切りが推定量の一部で扱うか, estimand の一部として扱うのかを区別することになると述べている。これに対応した打ち切りの分類として, noumenal censoring と phenomenal censoring を提案している。Noumenal censoring は試験デザインや来院スケジュール, 中止基準, 解析上の仕様で発生する, 被験者間で共通のイベントによる打ち切りである。Noumenal censoring の例としては, 追跡期間満了による打ち切りが該当する。Phenomenal censoring は個々の被験者に対して試験中の事象として発生する打ち切りであり, 比較的稀なイベントである。例えば偶然誤った治療を受けたことによる被験者の同意撤回が挙げられる。

Siegel らは続けて, noumenal censoring の場合, 基本的に打ち切り後のデータは収集されないと考えられるとした。従って, 本報告書におけるストラテジーの節で述べたように, 打ち切りを中間事象として治療方針ストラテジーを適用することには問題があると考えられるため, 治療方針ストラテジー以外のストラテジーを検討することが必要だと主張し, noumenal censoring は estimand の一部にあたることを主張している。そして仮想ストラテジーを用いた方が適切な解釈ができるかもしれないと述べている。なぜなら仮想ストラテジーを用いることで, 「打ち切りがない場合にどのような状況となるか」を推定することができるためである。ただし, 仮想ストラテジーを用いる際には, 中間事象による打ち切りは non-informative な打ち切りであることを想定していることには注意が必要である。

例えば, 比例ハザード性が成り立たない状況下で, ハザード比による治療効果の推定を考える。比例ハザード性が成立しないことから推定される治療効果は試験の終了時点に依存する。このときランダムな打ち切りであるにも関わらず, 治療方針ストラテジーの下で考えると, 偏りのある推定量となる (全被験者がイベントを発現するまでフォローした場合を除く)。他方, 治療下ストラテジーの下で考えると, 死亡という事象までの生存に関心のある治療効果考えるので偏りのない推定量となる。

Sigel らは試験実施上起こりうる noumenal censoring について明確に特定することを勧めている。治験実施計画書で規定されている来院日や中止基準も考慮したうえで estimand,



中間事象およびストラテジーを記載することを勧めている。

一方 noumenal censoring とは違って、phenomenal censoring は患者全員に同様に起こる事象ではないことから、estimand の一部というよりも推定量の一部をなすものと主張した。

Sigel らはこの2つの打ち切りの分類により、データと関心のある疑問に対する解釈に差異が発生するとした一方で、中間事象となる打ち切りはしばしば情報のある打ち切りなので、無情報の打ち切りを仮定した解析手法による結果の妥当性は確認すべきであるとも述べている。

#### 6.4 治療変更

観察期間中に治療の効果が不十分な場合、試験中に被験薬又は対照薬への治療変更を許容することがある。また、治療変更を許容しなくとも、被験薬（又は対照薬）中止後の新たな抗悪性腫瘍薬の使用には（倫理上の観点からも）制限がかけられないため、既承認の類薬が後治療として使用可能である場合、OS の評価が困難になることもある（図 6-1）。

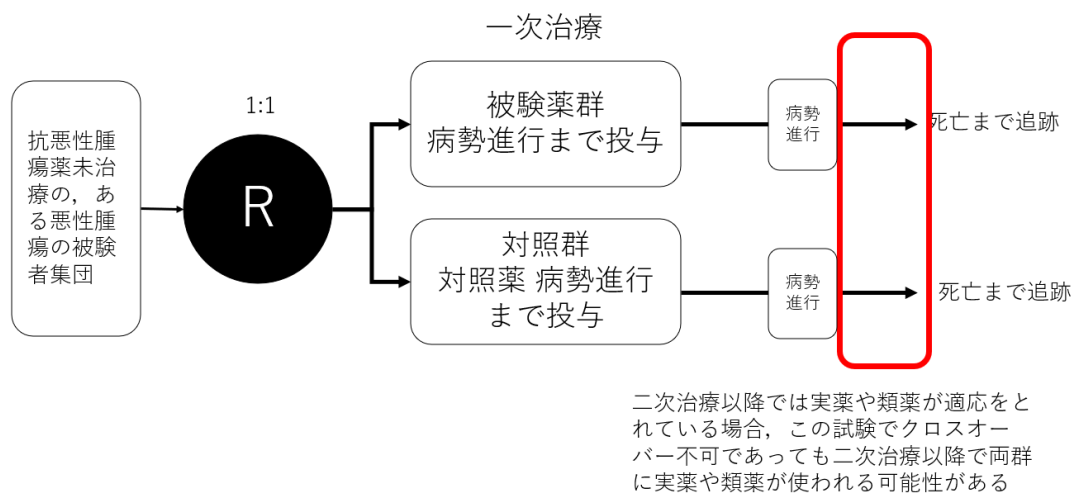


図 6-1 OS の評価が困難になる仮想事例

Manitz ら<sup>62)</sup> は、治療変更 (treatment switch) が発生しうる場合の OS に関する estimand の設定について検討した。内容を以下に要約する。

Manitz らは治療変更を、「関心のあるエンドポイントより前に、割付けられた試験治療を中止し、新しいがん治療方法（後治療）に変更したという中間事象」と定義した。またクロスオーバーは対照群の被験者が規定された臨床的イベント後に治験薬と同じ作用機序をもつ治療への治療変更として定義した。クロスオーバーを含む治療変更が存在する状況で、OS の治療効果を推定する4つの異なる目的と estimand が検討された。それぞれの

目的と estimand の要素を以下に示す。なお、治療変更に伴う estimand の検討であるため、治療変更と関係のない中間事象（例：有害事象による試験中止）は考慮されていない。

1. 治療変更が実臨床でも使われるものとした OS

クロスオーバーも許容するため、治験薬が承認された後に期待される標準治療を含み、後治療を含めた治療効果に興味がある場合と考えられる。承認を得るための試験では主要な estimand として扱われることが多く、Manitz らもこの estimand は常に示した方がよいとしている。

関心のある治療レジメン	一連の治療すべて (後治療, クロスオーバー含む)
中間事象① クロスオーバー以外の後治療開始	治療方針ストラテジー
中間事象② 増悪前のクロスオーバー開始	治療方針ストラテジー
中間事象③ 増悪後のクロスオーバー開始	治療方針ストラテジー
推定方法	Cox model や KM 推定量

2. 治療変更を補正した（治療変更が発生しなかった場合の）OS

患者が後治療に移行しなかった場合の OS に関心がある場合に用いる。対照群の患者が現在承認されておらず実臨床を反映しない治療に移ると試験計画段階で議論され、後治療に臨床的意義がないとみなしてよい場合に利用される。

関心のある治療レジメン	(治療変更がないと仮定した) 割付けられた治療
中間事象① クロスオーバー以外の後治療開始	仮想ストラテジー
中間事象② 増悪前のクロスオーバー開始	仮想ストラテジー
中間事象③ 増悪後のクロスオーバー開始	仮想ストラテジー
推定方法	IPCW (Inverse Probability of Censoring Weighting Method) で重みづけされた Cox model による調整済ハザード比 (HR), 信頼区間 (CI) や KM 推定量; IPCW を一般化した周辺構造モデル (Marginal Structural Models; MSM) の利用も可能
感度分析の例	IPCW では未測定交絡因子がないことが重要な仮定であるため, その仮定が満たされているか確認

3. クロスオーバーを補正した（クロスオーバーが発生しなかった場合の）OS

対照群の治療中止後にクロスオーバーをすることは実臨床を反映していないと考えられるときに有用である。

関心のある治療レジメン	(クロスオーバーがないと仮定した)
-------------	-------------------

	一連の治療と後治療
中間事象① クロスオーバー以外の後治療開始	治療方針ストラテジー
中間事象② 増悪前のクロスオーバー開始	仮想ストラテジー
中間事象③ 増悪後のクロスオーバー開始	仮想ストラテジー
推定方法	RPSFT (rank preserving structural failure time) モデルを用いて補正された生存時間に基づく HR ブートストラップによる CI 補正された生存時間に基づく KM IPCW の利用も可能
注意事項	RPSFT モデルは IPCW と異なり共変量の情報は不要で少数例でも実行可能だが、共通治療効果の仮定が必要
感度分析	共通治療効果の頑健性を確認する分析

#### 4. 増悪後のクロスオーバーを補正した OS

増悪後のクロスオーバー以外の治療変更は実臨床を反映していると考えられる場合に相当する。

関心のある治療レジメン	(増悪後のクロスオーバーがないと仮定した) 一連の治療と後治療
中間事象① クロスオーバー以外の後治療開始	治療方針ストラテジー
中間事象② 増悪前のクロスオーバー開始	治療方針ストラテジー
中間事象③ 増悪後のクロスオーバー開始	仮想ストラテジー
推定方法	Two-stage 法で再構成された生存時間* を用いた HR や KM 推定量 IPCW や RPSFT モデルの利用も可能
感度分析	異なる Secondary baseline の利用 (増悪時と治療の終了時を入れ替える)

\*: 加速モデルによって、増悪が認められクロスオーバーをした時点を secondary baseline timepoint として、被験者の生存時間を調整する。

Maintz ら<sup>62)</sup> は、治療変更を考慮した OS の解析について議論し、試験の設計段階で関心のある疑問への回答となる方法を予め特定することを勧めている。これにより事後的な (治療変更を調整した) 解析における偏りやデータの欠落が防げるとした。また標準的な estimand であっても、実臨床の観点等を考慮すると、それとは異なる estimand が関心のある疑問をより適切に反映することもありうるため、試験立案時にサイエンスの立場から estimand をよく検討し、適切な推定量を定めることを推奨している。

## 6.5 Estimand の記述について

Casey ら<sup>61)</sup> は、固形癌における estimand の検討を目的として、腎臓がん (RCC) の術後療法 (adjuvant treatment) に関する3つのランダム化プラセボ対照第 III 相試験の結果について、比較とレビューを行っている。

3試験とも類似した試験であり、主要評価項目は DFS である。また、イベント発現前後治療を開始した場合や、2つ以上の時点で測定が欠測となった場合は、打ち切りとして取り扱うことも共通である。一方で、DFS として対象とするイベントの定義や対象とする患者集団、中間事象の取り扱いには差異があり、エンドポイントだけの記述では臨床的疑問が十分に明確な記載とならない。加えて、3試験それぞれの治験実施計画書、統計解析計画書等の文書では違う用語を使用しているため、estimand の詳細はそれぞれ異なる文章から収集しなければならないと Casey らは述べている。その上で試験にて評価される治療効果について透明性がある記述が必要であることを指摘した。

Casey らは、estimand のフレームワークを用いることによる変化点を次のようにまとめている。「治験デザインの過程で以下の要素が対応していることが担保されているか議論を促進させる。それは関心のある臨床的疑問と解析内容や、複合変数を構成する各要素のバランス、解釈方法の対応等である。また、関心となる疑問に対して十分な検出力を担保した試験を設計することも可能になる。」このことは本報告書を通じたメッセージと共通した内容があると思われる。

## 6.6 小括

オンコロジー領域に関する estimand について概観した。既に本報告書本文や Casey ら<sup>61)</sup> が述べているように、estimand のフレームワークの導入により、治験そのものや既に合意形成されている標準的な解析手法が大幅に変更されるわけでもない。しかしながら、Fletcher ら<sup>9)</sup> にあるように「エンドポイントから計画して、エンドポイントを満たす試験とする」というマインドセットから「計画した試験は関心のある臨床的な疑問に答えることができる」というマインドセットに切り替えることが重要になる。本報告書で取り上げた estimand の thinking process はそのマインドセットに切り替えて思考する一つの方法である。「関心のある臨床的疑問は何か」、「試験目的、推定の対象は関心のある臨床的疑問に答えているか」、「estimand と estimand から設定される解析手法において、仮定されていることは何か」を試験に関係するすべての関係者が試験計画時から、意識し、議論し、認識を共有し、より適切な試験を計画運営していくことが重要となる。

## 7. まとめ

本報告書では ICH E9(R1)に記載されている estimand の概要を示した上で、Training Material にある thinking process に則って estimand の設定及び考慮事項を整理した。また新たな試みとして、治験実施計画書における estimand の記載方針を提言し、仮想事例を用いて治験実施計画書への具体的な記載内容を提示した。

試験計画段階で estimand のフレームワークを使うことで推定すべき治療効果を正確に叙述し、それを治験依頼者（企業）と規制当局の合意の下、事前に文書に規定しておくことは重要である。何故なら合意された estimand に基づいて試験を計画・実施し、得られた結果を解釈することによって、製薬企業と規制当局の両者で関心のある治療効果を推定することができるからである。これは医薬品開発、特に承認申請のためには重要なプロセスとなる。

Estimand の設定は臨床統計家だけに委ねられるものではなく、共に治験実施計画書を作成する臨床担当者やメディカルライターと協働し、計画している臨床試験でどのような治療効果を推定したいのか、という観点で深く議論されるべきものである。また estimand は共通言語として試験の目的とそれに対応する適切な試験デザイン及び解析手法を繋ぐものであり、ICH E9(R1)がステップ5になれば承認申請後のスムーズな審査のために PMDA と治験相談等で合意していく必要があるだろう。適用するストラテジーによっては、試験の実施・運営（例えば、基礎治療薬の用量規定の追加・変更やレスキュー薬使用の詳細な基準、できる限り詳細な治験薬中止理由の収集等）にも影響を与えることから、estimand は単に治験実施計画書の「統計解析（Statistical Analysis）」の章だけではなく、「目的（Objectives）」の章で試験目的と紐づけて適切に設定する必要があると考える。ICH E9(R1)がステップ5になり、今後新しく立ち上がる特に検証的試験に対して PMDA と事前に estimand を合意することが求められることを前提に考えると、目的の章への記載は必須と考える。実際に estimand の設定を行う際には Training Material の thinking process が非常に役立つため、こちらについても利用を勧めたい。

なお、本報告書において記載した指針は一例であり、記載事例をすべて満たすことが必須要件でないことを改めて強調させていただくが、本報告書が治験実施計画書作成の際の一助になれば幸いである。

## 8. 参考文献

- 1) International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human use (ICH). ICH harmonised tripartite guideline Statistical principles for clinical trials E9; 1998.  
[https://database.ich.org/sites/default/files/E9\\_Guideline.pdf](https://database.ich.org/sites/default/files/E9_Guideline.pdf)
- 2) ICH. ICH harmonised guideline addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1); Final Version.; 2019.  
[https://database.ich.org/sites/default/files/E9-R1\\_Step4\\_Guideline\\_2019\\_1203.pdf](https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf)
- 3) 日本製薬工業協会 ICH プロジェクト編集委員会. ICH 改革と ICH ガイドライン解説. 株式会社じほう; 2018.
- 4) ICH. E9(R1) Training Materials; 2021.  
[https://database.ich.org/sites/default/files/E9%28R1%29%20Training%20Material%20-%20PDF\\_0.pdf](https://database.ich.org/sites/default/files/E9%28R1%29%20Training%20Material%20-%20PDF_0.pdf)
- 5) TransCelerate, Clinical Content & Reuse Assets - Clinical Studies; Common Protocol Template.
- 6) ICH. ICH harmonised guideline clinical electronic structured harmonised Protocol (CeSHarP) M11 Template; Draft version.; 2022.  
[https://database.ich.org/sites/default/files/ICH\\_M11\\_Template\\_Step2\\_2022\\_0904.pdf](https://database.ich.org/sites/default/files/ICH_M11_Template_Step2_2022_0904.pdf)
- 7) 日本製薬工業協会医薬品評価委員会データサイエンス部会. ICH E9(R1)の理解に役立つ因果推論; 2022.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/gbkspa00000012pj-att/ds\\_202207\\_causal.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/gbkspa00000012pj-att/ds_202207_causal.pdf)
- 8) 日本製薬工業協会医薬品評価委員会データサイエンス部会. ICH E9(R1)の理解に役立つ因果推論～時間依存性治療編～; 2022.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/gbkspa0000001g13-att/DS\\_202209\\_causal-tv.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/gbkspa0000001g13-att/DS_202209_causal-tv.pdf)
- 9) Fletcher C, Hefting N, Wright M, et al. Marking 2-years of new thinking in clinical trials: The Estimand journey. Ther Innov Regul Sci. 2022;56(4):637-650.
- 10) Bell J, Hamilton A, Sailer O, Voss F. The detailed clinical objectives approach to designing clinical trials and choosing estimands. Pharm Stat. 2021;20(6):1112-1124.
- 11) Lawrance R, Degtyarev E, Griffiths P, et al. What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials?. J Patient Rep Outcomes. 2020;4(1):68.
- 12) 日本製薬工業協会医薬品評価委員会データサイエンス部会. 総括報告書の解析帳票レイアウトと使用する ADaM データの事例; 2019.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000a0pg-att/adam\\_data.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000a0pg-att/adam_data.pdf)



- 13) Mallinckrodt CH, Molenberghs G, Lipkovich I, Ratitch B. Estimands, Estimators and Sensitivity Analysis in Clinical Trials. Chapman & Hall/CRC; 2020.
- 14) Cro S, Kahan BC, Rehal S, Ster AC, Carpenter JR, White IR, Cornelius VR. Evaluating how clear the questions being investigated in randomised trials are: systematic review of estimands. *bmj*. 2022 Aug 23;378.
- 15) Lynggaard H, Bell J, Lösch C, et al. Principles and recommendations for incorporating estimands into clinical study protocol templates. *Trials*. 2022;23(1):685.
- 16) Siegel J, Grinsted L, Liu F, Weber HJ, Englert S, Casey M. Censoring and censoring mechanisms in oncology in light of the estimands framework. arXiv preprint arXiv:2203.01781. 2022 Mar 3. <https://arxiv.org/abs/2203.01781>
- 17) Mitroiu M, Oude Rengerink K, Teerenstra S, Pétavy F, Roes KC. A narrative review of estimands in drug development and regulatory evaluation: old wine in new barrels?. *Trials*. 2020 Dec;21(1):1-4.
- 18) EMA/Committee for Medicinal Products for Human Use (CHMP). Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus; Rev. 2, draft. 2018. [https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-treatment-prevention-diabetes-mellitus\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-treatment-prevention-diabetes-mellitus_en.pdf)
- 19) US Food and Drug Administration (FDA). Table of surrogate endpoints that were the basis of drug approval or licensure. <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure>
- 20) FDA. Guidance for industry: Clinical trial endpoints for the approval of cancer drugs and biologics; 2018. <https://www.fda.gov/media/71195/download>
- 21) 厚生労働省. 抗悪性腫瘍薬の臨床評価方法に関するガイドラインに関する Q & A. 厚生労働省医薬食品局審査管理課事務連絡別添 平成18年3月1日.
- 22) Sun S, Weber HJ, Butler E, Rufibach K, Roychoudhury S. Estimands in hematologic oncology trials. *Pharm Stat*. 2021;20(4):793-805.
- 23) Akacha M, Bretz F, Ruberg S. Estimands in clinical trials - broadening the perspective. *Stat Med*. 2017;36(1):5-19.
- 24) Ratitch B, Bell J, Mallinckrodt C, et al. Choosing Estimands in Clinical Trials: Putting the ICH E9(R1) Into Practice. *Therapeutic Innovation & Regulatory Science*. 2020;54(2):324-341.
- 25) Ionan AC, Paterniti M, Mehrotra DV, et al. Clinical and Statistical Perspectives on the ICH E9(R1) Estimand Framework Implementation. *Statistics in Biopharmaceutical Research*. Published online July 6, 2022:1-6.

- 26) Keene ON, Wright D, Phillips A, Wright M. Why ITT analysis is not always the answer for estimating treatment effects in clinical trials. *Contemporary Clinical Trials*. 2021 Sep 1;108:106494.
- 27) Mallinckrodt C, Molenberghs G, Rathmann S. Choosing estimands in clinical trials with missing data. *Pharmaceutical Statistics*. 2017;16(1):29-36.
- 28) Lachin JM. Statistical considerations in the intent-to-treat principle [published correction appears in *Control Clin Trials* 2000 Oct;21(5):526]. *Control Clin Trials*. 2000;21(3):167-189.
- 29) Darken P, Nyberg J, Ballal S, Wright D. The attributable estimand: A new approach to account for intercurrent events. *Pharmaceutical Statistics*. 2020;19(5):626-635.
- 30) Keene ON. Strategies for composite estimands in confirmatory clinical trials: Examples from trials in nasal polyps and steroid reduction. *Pharmaceutical Statistics*. 2019;18(1):78-84.
- 31) Bornkamp B, Rufibach K, Lin J, et al. Principal stratum strategy: Potential role in drug development. *Pharmaceutical Statistics*. 2021;20(4):737-751.
- 32) Lipkovich I, Ratitch B, Qu Y, Zhang X, Shan M, Mallinckrodt C. Using principal stratification in analysis of clinical trials. *Statistics in Medicine*. 2022;41(19):3837-3877.
- 33) Qu Y, White RD, Ruberg SJ. Accurate collection of reasons for treatment discontinuation to better define estimands in clinical trials. arXiv:220601659 [stat]. Published online June 3, 2022. <https://arxiv.org/abs/2206.01659>
- 34) Permutt T. Sensitivity analysis for missing data in regulatory submissions. *Statistics in Medicine*. 2015;35(17):2876-2879.
- 35) Wang S, Hu H. Impute the missing data using retrieved dropouts. *BMC Med Res Methodol*. 2022;22(1):82.
- 36) 日本製薬工業協会医薬品評価委員会データサイエンス部会. 臨床試験における Estimand と感度分析に関連する文献調査; 2019. [https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000b4ql-att/estimand\\_analysis.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000b4ql-att/estimand_analysis.pdf)
- 37) Akacha M, Bartels C, Bornkamp B, et al. Estimands-What they are and why they are important for pharmacometricians. *CPT Pharmacometrics Syst Pharmacol*. 2021;10(4):279-282.
- 38) Kang M, Kendall MA, Ribaud H, et al. Incorporating estimands into clinical trial statistical analysis plans. *Clinical Trials*. 2022;19(3):285-291.
- 39) Little RJ, Lewis RJ. Estimands, Estimators, and Estimates. *JAMA*. 2021;326(10):967-968.
- 40) Rosenstock J, Allison D, Birkenfeld AL, et al. Effect of Additional Oral Semaglutide vs Sitagliptin on Glycated Hemoglobin in Adults With Type 2 Diabetes Uncontrolled With Metformin Alone or With Sulfonylurea: The PIONEER 3 Randomized Clinical Trial. *JAMA*. 2019;321(15):1466-1480.
- 41) Hernán MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. *N Engl J Med*.

- 2017;377(14):1391-1398.
- 42) 日本製薬工業協会医薬品評価委員会データサイエンス部会. Pragmatic Trials のススメ; 2018.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000bxjh-att/pragmatic\\_trial\\_final\\_2.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000bxjh-att/pragmatic_trial_final_2.pdf)
- 43) EMA/CHMP. Points to consider on implications of Coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials; Rev. 1. 2020.  
[https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-implications-coronavirus-disease-covid-19-methodological-aspects-ongoing-clinical\\_en-0.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-implications-coronavirus-disease-covid-19-methodological-aspects-ongoing-clinical_en-0.pdf)
- 44) Meyer RD, Ratitch B, Wolbers M, et al. Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic. Statistics in Biopharmaceutical Research. 2020;12(4):399-411.
- 45) EMA/CHMP. Points to consider on the impact of the war in Ukraine on methodological aspects of ongoing clinical trials; 2022.  
[https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-impact-war-ukraine-methodological-aspects-ongoing-clinical-trials\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-impact-war-ukraine-methodological-aspects-ongoing-clinical-trials_en.pdf)
- 46) 日本製薬工業協会医薬品評価委員会データサイエンス部会. COVID-19が治験に及ぼす影響についてー統計解析・データマネジメントの視点ー; 2020.  
[https://www.jpma.or.jp/information/evaluation/results/message/lofurc000000cjhp-att/covid-19\\_statement.pdf](https://www.jpma.or.jp/information/evaluation/results/message/lofurc000000cjhp-att/covid-19_statement.pdf)
- 47) EMA/CHMP. Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease; 2018.  
[https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicines-treatment-alzheimers-disease-revision-2\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicines-treatment-alzheimers-disease-revision-2_en.pdf)
- 48) EMA/CHMP. Concept paper on the need to develop a reflection paper on development of medicinal products to prevent and treat acute kidney injury; 2018.  
[https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-need-develop-reflection-paper-development-medicinal-products-prevent-treat-acute\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-need-develop-reflection-paper-development-medicinal-products-prevent-treat-acute_en.pdf)
- 49) EMA/CHMP. Reflection paper on regulatory requirements for the development of medicinal products for chronic non-infectious liver diseases; 2018.  
[https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-regulatory-requirements-development-medicinal-products-chronic-non-infectious-liver\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-regulatory-requirements-development-medicinal-products-chronic-non-infectious-liver_en.pdf)
- 50) EMA/CHMP. Guideline on clinical investigation of medicinal products in the treatment of epileptic disorders; 2018.  
[https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-treatment-epileptic-disorders-revision-3\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-treatment-epileptic-disorders-revision-3_en.pdf)

- 51) EMA/CHMP. Guideline on quality, non-clinical and clinical aspects of medicinal products containing genetically modified cells; 2020.  
[https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-quality-non-clinical-clinical-aspects-medicinal-products-containing-genetically-modified\\_en-0.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-quality-non-clinical-clinical-aspects-medicinal-products-containing-genetically-modified_en-0.pdf)
- 52) EMA/CHMP. Guideline on registry-based studies; 2021.  
[https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies\\_en-0.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en-0.pdf)
- 53) FDA. Guidance for industry: Eosinophilic Esophagitis: Developing Drugs for Treatment; 2020.  
<https://www.fda.gov/media/120089/download>
- 54) FDA. Guidance for industry: Acute Myeloid Leukemia: Developing Drugs and Biological Products for Treatment; 2022.  
<https://www.fda.gov/media/162362/download>
- 55) FDA. Guidance for industry: Chronic Rhinosinusitis with Nasal Polyps: Developing Drugs for Treatment; 2021.  
<https://www.fda.gov/media/154724/download>
- 56) FDA. Guidance for industry: Crohn's Disease: Developing Drugs for Treatment; 2022.  
<https://www.fda.gov/media/158001/download>
- 57) FDA. Guidance for industry: Ulcerative Colitis: Developing Drugs for Treatment; 2022.  
<https://www.fda.gov/media/158016/download>
- 58) 厚生労働省. がん免疫療法開発のガイダンスについて. 薬生薬審発0308第1号/薬生機審発0308第1号, 平成31年3月8日.
- 59) Rufibach K. Treatment effect quantification for time-to-event endpoints-Estimands, analysis strategies, and beyond. Pharm Stat. 2019;18(2):145-165.
- 60) Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. J Clin Oncol. 2008;26(22):3721-3726.
- 61) Casey M, Degtyarev E, Lechuga MJ, et al. Estimand framework: Are we asking the right questions? A case study in the solid tumor setting. Pharm Stat. 2021;20(2):324-334.
- 62) Manitz J, Kan-Dobrosky N, Buchner H, et al. Estimands for overall survival in clinical trials with treatment switching in oncology. Pharm Stat. 2022;21(1):150-162

日本製薬工業協会 医薬品評価委員会 データサイエンス部会

2021年度タスクフォース4 / 2022年度継続タスクフォース4

“Estimand の実装” プロトコルサブチーム

タスクフォースメンバー

氏名	所属
今井 翔悟	旭化成ファーマ株式会社
尾崎 凌斗	中外製薬株式会社
作井 将	武田薬品工業株式会社
佐野 文哉	住友ファーマ株式会社
高戸 俊一	杏林製薬株式会社
竹ノ内 一雅	日本イーライリリー株式会社
手島 梨恵	ノバルティスファーマ株式会社
寺田 亘	アステラス製薬株式会社
富田 裕章	中外製薬株式会社
西田 朋由	ノボ ノルディスク ファーマ株式会社
平岡 毅彦	丸石製薬株式会社
福岡 鮎美	中外製薬株式会社
白石 亜矢子*	ヤンセンファーマ株式会社
土屋 悟**	住友ファーマ株式会社
菅波 秀規**	興和株式会社

\* タスクフォース推進委員, \*\* データサイエンス部会担当副部長