

## HTA における Target Trial Emulation の利用状況と評価上の課題

立命館大学 生命科学部 生命医科学科

医療政策・管理学研究室

森脇 健介

### 1. 要旨

ランダム化比較試験(Randomized controlled trial: RCT)のエビデンスは、長期的な効果を把握できないことや関連する比較対照技術やアウトカムを必ずしも含まないことから、医療技術評価(HTA: Health technology assessment)のニーズを満たさないことが多い。これを補完するソースとして、リアルワールドデータが注目されているが、患者の選択や交絡によるバイアスが懸念される。リアルワールドデータを用いた非ランダム化研究において、RCT のデザインの原則を適用する方法として、Target Trial Emulation(TTE)のアプローチが開発されている。本研究では、HTA の文脈における TTE をはじめとする非ランダム化研究の利用状況と課題を整理した。

### 2. 研究目的

医療技術評価のプロセスにおいて、医療技術の追加的有用性と費用対効果の評価は重要な 2 つの柱である。これらの検証において、ランダム化比較試験は内的妥当性の観点から最も重要なエビデンスであることは周知の事実である。一方で、RCT は必ずしも HTA の文脈で意思決定者のニーズを満たさないことが多い。例えば、目標となる対象集団への一般化可能性や長期的な効果における不確実性、関連する比較対照技術やアウトカムが含まれない可能性などが課題として挙げられる。こうした課題をクリアするために、リアルワールドデータ(Real world data: RWD)の活用に期待が寄せられるが、非ランダム化研究に特有の、患者の選択や交絡といったバイアスの問題が不可避である。近年、臨床・薬剤疫学の分野では、RWD を用いた非ランダム化研究において、RCT デ

ザインの原則を適用する方法として、Target trial emulation(TTE)のアプローチが開発されている。HTA の分野でも TTE のコンセプトを応用した評価手法の提案や事例の蓄積が進んでおり、今後、本邦でもこうした国際的潮流を踏まえた評価体制や方法論の整理を行う必要がある。本研究では、HTA における TTE の利用状況や具体的事例を把握することと方法論上の留意点を整理することを目的とした文献調査を実施した。

### 3. 研究方法

PubMed および Google scholar を利用して、“Health technology assessment” と “Target trial emulation”等をキーワードとした文献検索を実施し、諸外国の HTA 機関における TTE や RWD の活用に関する報告を特定し、記載事項の要約・整理を行った。

### 4. 研究結果

#### 4-1. RCT の限界

NICE では、HTA において RCT が利用できないケースを、以下のように整理している[1]。

##### <RCT が実施できない状況>

1. ランダム化が、例えばアンメット・ニーズが高いなどの理由で、倫理的に問題があると考えられる。
2. 患者が試験中の介入のいずれかに割り付けられることを望まない場合。
3. 医療専門家は、患者を効果が低いと思われる介入にランダムに割り付けたがらない。
4. 少数の対象患者
5. 研究の財政的または技術的制約
6. すべての治療の組み合わせ(治療順序を含む)を直接評価できるわけではない。

### <RCTがあっても利用できない状況>

1. NHSにおける標準治療を代表しない比較対照(プラセボ対照を含む)の使用
2. 検証されていない代替アウトカムの使用
3. 限定的なフォローアップ
4. 対象集団の除外(例えば、合併症のある人、妊婦、小児など)
5. NHSの対象集団への結果の移植性に影響を与える集団、ケア経路、設定の違い
6. 患者の技術使用の違い
7. 日常診療とは異なる臨床支援
8. 学習効果(すなわち、ユーザーが経験を積むにつれて、介入の効果が時間とともに変化すること)
9. 治療法の切り替え、追跡不能、データの欠落などのランダム化後の事象に対処するために使用された方法。

#### 4-1. TTE とは

Target trial emulation は、観察研究などの非ランダム化研究のデータを用いて、目標とする試験(関心のあるリサーチクエスチョンに解を与えるであろう仮想の RCT を模倣することにより、因果推論を行う方法である[2]。一般に、TTE では目標とする試験のデザインと解析に関してプロトコルを事前に規定する。その上で、目標試験で収集されるであろうデータを模倣するように、このプロトコルを RWD に適用する。表 4-1-1 に TTE のプロトコルのイメージを示す。

表 4-1-1. TTE のプロトコルのイメージ[2]

要素	target trial のプロトコル	具体的な取り扱い
	リサーチクエスチョン：メトホルミンが奏効しなかった 2 型糖尿病患者に対する治療強化の戦略間で HbA1c、死亡や心血管疾患(CVD)のリスク、コストは異なるか？	

適格基準	<p>選択基準：</p> <ul style="list-style-type: none"> <li>• 2型糖尿病の診断がある成人(18歳以上)</li> <li>• 少なくとも3か月のメトホルミン使用歴がある</li> <li>• 血糖コントロールが不十分である(HbA1c <math>\geq</math> 7.5%)</li> <li>• 過去に治療強化を行っていない</li> </ul> <p>除外基準：</p> <ul style="list-style-type: none"> <li>• メトホルミン不耐性のもの</li> </ul>	<p>選択基準：同左 除外基準：同左。ただし、ベースライン時点の前6か月における治療に関するデータが利用できないものも除外する。</p>
治療戦略	<p>3つの治療戦略を設定する：</p> <ol style="list-style-type: none"> <li>1. SU剤によって治療強化を開始</li> <li>2. DPP-4阻害薬によって治療強化を開始</li> <li>3. SGLT-2阻害薬によって治療強化を開始</li> </ol>	<p>ベースライン時点での治療に関するデータが左記の治療強化戦略と一致する場合、該当する治療戦略に割付けられる。異なる患者(例えばインスリンを開始しているなど)は不適格となる。</p>
割付け方法	<p>参加者はベースライン時点でランダムに各治療戦略へ割付けられ、その内容を知っている(非盲検)。</p>	<p>全ての交絡因子のばらつきが群間で調整されることによってランダム化する。</p>
追跡	<p>開始：治療戦略が割付けられた時点(ランダム化時点) 終了：ランダム化後2年経過、死亡、CVDイベントの発生、脱落のいずれか最も早い時点</p>	<p>開始：各治療戦略の開始時点 終了：同左</p>
アウトカム	<p>HbA1c、全死亡、CVDイベント、コスト</p>	<p>同左</p>

Estimands	<p>intention-to-treat(ITT)effect : いずれかの治療戦略に割付けられたときの治療効果</p> <p>per-protocol(PP)effect : いずれかの治療戦略を 2 年間継続的に受けたときの治療効果</p>	<p>ITT effect : 異なる治療戦略を導入した群間での比較</p> <p>PP effect: 割付けられた治療戦略を遵守した患者間での比較</p>
分析方法	<p>ITT analysis: ランダム化された群間でのリスクを比較</p> <p>PP analysis : ベースライン時点と追跡期間中のアドヒアランスに関する因子を調整したもとのリスクを比較</p> <p>双方の分析ともに、追跡が不十分な場合は脱落に関する調整が必要となる。</p>	<p>ITT analysis : 同左。ただしベースライン時点での交絡因子の調整が必要となる。</p> <p>PP analysis : 同左。ただしベースライン時点と追跡期間中の脱落に関する因子の調整も必要となる。</p> <p>双方の分析ともに、未測定の交絡因子がないことを仮定する。</p>

TTE を利用することの強みとしては、以下が挙げられている。

- 非ランダム化試験でよく見られるデザインの欠陥や明らかなパラドックスを回避できる。例えば、適格性評価と治療割り付けのずれに起因する選択バイアスである。
- 非ランダム化試験の潜在的な限界を特定し、トレードオフを明確にするための構造化されたプロセスを提供する。特に、RWD には無作為化をエミュレートするための交絡因子に関する十分な情報が含まれていない可能性がある。割り付け手順を明確に指定することで、**TTE** は測定されていない交絡を探索したり、意味のある方法で対象試験を再定義するための適切な分析戦略を考えるのに役立つ。
- TTE は非ランダム化試験のデザインと解析を対象試験と明示的に結びつけることで、その基礎となる仮定と結果の解釈を容易にする。
- 十分な質の大規模 RWD を利用することで、TTE は仮定の RCT と比較して、より大規模で多様な集団を含み、より長い追跡調査が可能である。加えて、RCT に含まれる患者と比較して、臨床の日常診療で観察される患者のタイプやケアをより忠実に反映させることができる。

先行研究のレビュー報告によると、TTEによりRCTの結果を近似可能であることが実証されている[2]。異なる研究チームが同じ目標試験をエミュレーションした取り組みや、RCTの結果が判明する前に試験のエミュレーションが完了した取り組みがある[2]。また、RCTとTTEで乖離した事例は、その原因を明らかにするとともに、RWDの限界を理解するのに役立つものと考察されている。例えば、RCTのすべての適格基準をエミュレートすることは困難であるが、できる範囲や方法を明らかにすることは可能である。また、いくつかの研究では、エミュレーション研究とRCTの結果の違いは、未測定交絡の存在ではなく、ゼロ時点(ベースライン)の定義など、デザインの特徴の違いによるものであることが示された。ただし実際には、非ランダム化試験は測定不能な交絡の可能性を排除できないため、TTEは十分に実施されたRCTを補完するものとして捉えるのが最善である。

#### 4-2. TTE や RWD の活用状況

Griffiths EA によると、2010～2015年のNICEの技術評価(Technology appraisal: TA)の38%(118件中45件)は、追加的有用性の評価において、非ランダム化研究を考慮していることが示されている[3]。これは、単群試験のデータと外部対照としてのレジストリーデータを利用するような分析を含んでいる。また、カナダではCanadian Real-world Evidence for Value of Cancer Drugs(CanREValue)共同研究の下で、カナダの各州が一貫した統合的な方法で、抗がん剤の給付のためのリアルワールドエビデンス(RWE)の作成と利用の枠組みを開発し、検証している[4]。大規模なRWDの連結など、利用環境の改善に沿ってこのような動きは加速化するものと予想される。

特に外部対照(External control arm: ECA)の評価において、非ランダム化研究の活用が進んでおり、Jaskaらは7つの技術評価の事例(ブリナツモマブ2適応、アベルマブ、エルダフィチニブ、エントレクチニブ、トラスツズマブ・デルクステカン、イデカブタゲン・ビクリューセル)を特定し、規制当局・HTA機関における論点を整理した[5]。その結果、おおむね以下の論点に集約された。

- 一般化可能性の問題：標準治療が時代や地域によって異なることや関心となる対象集団が異なることによる一般化可能性に対する批判。
- 交絡の問題：Naïveな比較に対する批判。未測定あるいは未調整の交絡因子に対する批判。
- その他の問題：選択バイアス、不適切な調整解析、不適切なアウトカム定義、情報量不足に対する批判。

なお、選択バイアスと交絡が最も共通する論点であった。一方、規制当局と HTA 機関の間における批評の一致は低かった。また、これら外部対照の論点が意思決定に与える影響も様々であった。

また、Campbell らは FDA ガイダンスで求められる科学的頑健性に加えて、RWD が許容されうる条件を考察・整理し、企業向けにチェックリストを作成している[6]。それによると、以下の条件が重要とされている。

- 致命的疾患であるか？アンメット・ニーズがあるか？
- RCT が倫理的に実施できないのか？
- RWD でエンドポイントを適切かつ客観的に評価できるか？
- 疾患の自然史が十分に理解されているか？
- 意思決定を支援するその他のエビデンスはあるか？
- 治療効果が十分に大きいことを示すエビデンスがあるか？

#### 4-3. NICE の RWD 利用のフレームワーク

近年、NICE は、エビデンスギャップの解消とイノベーションへのアクセス改善を目指し、RWD を活用するための枠組みを整理している[1]。本枠組みの狙いは、不確実性を減らし、ガイダンスを改善するために、リアルワールドのデータをどのような場合に使用できるかを特定することにある。また、エビデンスの質と透明性を向上させるために、リアルワールド研究を計画、実施、報告するためのベストプラクティスを示すことを意図して作成されている。その基本原則は表 4-3-1 に要約される。

表 4-3-1. RWD 活用の基本原則

エビデンス作成の段階	主な検討事項
計画	<ul style="list-style-type: none"> <li>● 研究課題を明確に定義し、関連する場合は、集団の適格基準、介入、アウトカム、推計対照量の定義を含む。</li> <li>● 事前に研究を計画し、プロトコル(データ分析計画を含む)を公開する。</li> <li>● 十分な実績があり、研究課題を解決するのに十分な品質と関連性のあるデータを選択する</li> </ul>

	<ul style="list-style-type: none"> <li>● 一次データ収集を計画する際には、患者や医療従事者の負担を最小限に抑えつつ、患者中心の方法で収集を実施する方法を検討する。</li> <li>● 現地の法律、データガバナンスプロセス、行動規範、およびデータ管理者の要件に従ってデータを使用する。</li> </ul>
実施	<ul style="list-style-type: none"> <li>● バイアスの主要なリスクを考慮し、研究課題に適した研究デザインと統計的手法を使用する。</li> <li>● 感度分析及び／又はバイアス分析を用いて、バイアスの主要なリスクや不確実なデータキュレーション又は分析上の決定に対する研究の頑健性を評価する。</li> <li>● 研究の完全性と質を保証するための品質保証基準とプロトコルを作成し、実施する。</li> </ul>
報告	<ul style="list-style-type: none"> <li>● 独立した研究者が、何がなせ行われたかを十分に理解し、研究を批判的に評価し、再現できるように、研究デザインと分析方法を十分に詳細に報告すること。</li> <li>● 報告書はまた、それをカバーするものでなければならない： <ul style="list-style-type: none"> <li>○ データの出所、品質、関連性</li> <li>○ データキュレーション</li> <li>○ 初期データから最終解析までの患者の減少</li> <li>○ 患者の特徴(欠測データを含む)、全体および主要集団におけるフォローアップの詳細</li> <li>○ 計画され実施されたすべての解析の結果(事前に計画されていなかった解析は明確に示すこと)</li> <li>○ バイアスのリスクと NHS の対象集団への一般化可能性の評価</li> <li>○ 研究の限界と結果の意味の解釈</li> </ul> </li> </ul>

なお、NICEでは、一般的に使用されうるRWDの情報源も属性ごとに整理したうえで、提示している(表4-3-2)。

表 4-3-2. イギリスにおける一般的な RWD の情報源

データソース	説明	例
電子カルテ	<p>コンピュータ化された個々の患者の記録。これらは通常、患者の臨床管理に利用される。</p> <p>検査室、ゲノム、画像システムなど、他の情報システムからのデータを統合することもある。</p>	<p><a href="#">Clinical Practice Research Datalink (CPRD)</a> GOLD には、英国内の一般診療所に登録された患者の人口統計学的情報と臨床情報が含まれている。</p>
行政データ	<p>医療・福祉サービスによって管理目的で収集されたデータ。</p>	<p><a href="#">Hospital Episode Statistics (HES)</a> Admitted Patient Care データセットには、NHS の病院に入院したすべての患者、あるいは NHS の資金で私立病院で治療を受けたすべての患者の診断と処置に関する情報が含まれている。HES の主な目的は、PBP(payment by results)やその他の運営活動を通じて、病院の診療報酬に反映させることである。</p>
クレームデータ	<p>医療サービスの利用に関する行政データの一つで、保険システムから収集されることが多い。</p>	<p><a href="#">メディケア&amp;メディケイドサービスセンターのデータ</a>には、払い戻し情報または請求書の支払いから得られたメディケアサービスを受けた個人に関するデータが含まれる。</p> <p><a href="#">NHS Business Services Authority</a> は、イングランドのプライマリーケアで調剤された医薬品に関するデータ</p>

		を提供している。
患者登録	<p>登録は、特定の疾患、状態、または曝露によって定義された集団の特定の結果を特定するために、統一されたデータ(臨床およびその他)を収集する組織化されたシステムである。</p> <p>登録には、研究、臨床ケア、政策などいくつかの目的がある。登録には介入研究も含まれる。</p>	<p><a href="#">Systemic Anti-Cancer Therapy(SACT)データセット</a>には、NHS イングランドのプロバイダーから抗がん剤治療を受けた全患者の情報が含まれている。このデータはNICE 内で広く使用され、Cancer Drugs Fund 内で使用が承認された薬剤に関する情報を提供している。</p> <p><a href="#">UK Cystic Fibrosis Registry</a> は、英国内の嚢胞性線維症の専門センターにおいて、同意を得た嚢胞性線維症患者のデータを収集しています。登録データは、研究を促進し、ケアセンターにおける質の向上を導き、新薬の安全性を監視することで、嚢胞性線維症の人々の健康を改善するために使用されます。</p>
患者が作成した健康データ	<p>ウェアラブル医療 instrument や個人用 instrument、モバイルアプリ、ソーシャルメディア、その他のインターネットベースのツールから、患者やその介護者が直接生成したデータ。データは、能動的に(例えば、人々がフォームにデータを入力することによって)収集されることもあれば、受動的に(例えば、人々の活動レベルを測定するスマートウォッチによって)収集されることもある。</p>	<p>入院の必要性を警告するため、在宅で治療を受けている COVID-19 患者のモニターにパルスオキシメーターが使用されている(<a href="#">Greenhalgh ら、2021 年</a>)。</p> <p><a href="#">ZOE アプリ</a>から得た COVID-19 と long-COVID 症状に関する自己申告データ。</p>

カルテレビュー	<p>患者の医療記録(紙または電子記録を含む)のレビューからレトロスペクティブに抽出されたデータ。</p> <p>カルテレビューは自然史研究において広く用いられている。ルーチンのデータソースでは報告されていないデータを抽出することができる。</p>	<p>レトロスペクティブ・カルテレビューは、疾患の自然史や治療経路をモデル化するために、希少疾患の研究では特によく行われる(<a href="#">Garbade et al.</a>)</p>
監査とサービス評価	<p>臨床監査は、現在の医療水準がベストプラクティスや設定された基準に照らしてどのようなものであるかを理解し、質の改善につなげるために行われる。データは前向きに収集されることもあれば、後向きに収集されることもある。</p> <p>サービス評価は、現在のケアを定義し判断するために行われる。</p>	<p><a href="#">医療の質向上パートナーシップ</a>は、<a href="#">センチネル脳卒中全国監査プログラム(SSNAP)</a>などの全国臨床監査プログラムを管理している。SSNAPは、イングランド、ウェールズ、北アイルランドにおける集学的脳卒中入院医療サービスの組織と提供の質を評価するために使用される。</p>
一次データ収集による観察コホート	<p>1 つ以上の研究課題に答えるためにデザインされた伝統的な前向き研究。</p>	<p><a href="#">英国バイオバンク</a>は、患者の病歴と遺伝学に関するデータを収集している。健康アウトカムについては患者の記録とリンクしている。特定の研究課題のために設計されたのではなく、疫学研究を可能にするためである。</p> <p><a href="#">EMBRACE-I</a>は、局所進行子宮頸部腫瘍に対してMRIを用いた画像誘導適応ブラキセラピーを受けた患者の局所腫瘍制御と罹患率を評価する多施設共同前向きコホート研究である。</p>

<p>健康調査、インタビュー、フォーカス・グループ</p>	<p>健康調査とは、調査を通じて、ある集団における健康と病気に関するデータを系統的に収集することである。集団の健康状態の傾向を把握したり、患者の医療体験を理解したりするなど、さまざまな目的がある。</p> <p>インタビューやフォーカスグループは、患者の認識や経験などの質的データを収集するために行われる。</p>	<p><a href="#">イングランド健康調査</a>は、イングランドにおける健康の傾向を測定するために毎年行われる代表的な世帯調査である。</p> <p>患者チャリティー団体 Lipoedema UK による'<a href="#">Living with Lipoedema'2021年調査</a>は、脂肪浮腫患者の体験データを収集するものである。この調査は、脂肪浮腫のために美容整形以外の脂肪吸引やその他の治療を受けた患者の経験を評価するものである。</p>
-------------------------------	---	--

さらに、RWD を用いた研究の計画、分析、報告に際して、特に考慮すべき事項を以下の通り示している。

- **研究デザイン**

- 望ましいランダム化比較試験を模倣した試験をデザインする(ターゲットトリアルアプローチ)。
- 患者の適格基準が満たされ、治療が割り当てられ、フォローアップが開始されるまでの間の差に起因する時間に関連したバイアスを避ける。
- 外部統制を用いる研究では、主要な研究変数の利用可能性と運用上の定義、データ収集プロセス、患者の特徴、治療設定、ケア経路、期間など、データソース間の差異を最小限にするようデータを選択し、キュレーションし、研究の質と妥当性への影響を考慮する。

- **分析**

- 系統的なアプローチにより潜在的交絡因子(時変交絡因子を含む)を同定し、因果関係の仮定を明確にする。
- 観察された交絡因子と観察されていない交絡因子を考慮した交絡に対処する統計的方法を用いる。
- 情報打ち切り、欠損データ、測定誤差によるバイアスの影響を考慮し、必要であれば適切に対処する。
- 感度分析とバイアス分析を用いて、バイアスの主なリスクや不確実なデータキュレーションと分析の決定に対する結果の頑健性を評価する。

- **報告**

- 非ランダム化エビデンスの必要性を正当化する。
- 最終的な解析を行う前に、試験プロトコルと統計解析計画を提供する。
- 独立した研究者が研究を再現し、何がなぜ行われたかを理解できるよう、研究を十分に詳細に報告する。
- バイアスのリスクと研究課題との関連性を評価する。

※ただし、許容されるエビデンスの質は、用途や様々な文脈的要因に依存する可能性がある。

## 5. 考察

### 5-1. HTA における活用の方向性

今後、HTA の文脈で想定される TTE/RWD の活用の方法としては、主に以下の 4 点が想定される。

- **単群試験と外部対照との比較**：レジストリーから比較対照群のデータを得て、マッチングする際に、適格性や時間ゼロ、追跡開始の正しい調整など、見落とされがちなデザイン上の要点を TTE のアプローチにより認識することができる。
- **間接比較法の改善**：NMA など従来の間接比較法の仮定(試験間の効果修飾因子の分布がそろっているか等)の妥当性を評価できる。
- **HTA の再評価**：RWD における効果の迅速な評価・再評価が可能となる。
- **複雑な治療戦略の評価**：長期にわたり動的に変化するような治療戦略(糖尿病など)の評価を可能にする。

### 5-2. HTA における活用のハードル

一方で、HTA の文脈で想定される TTE/RWD の活用のハードルとして、以下の 5 点が挙げられる。

- **研究デザイン**：疫学分野ではかなり前から使用されているが、HTA では比較的新しいコンセプトである。
- **因果推論手法**：因果推論の分析手法の一部は、HTA ではなじみがない。
- **情報の範囲**：HTA での関心データと RW で収集されるデータにギャップがある。
- **データへのアクセス**：最も豊富な情報を持つ RWD へのアクセスを得るには、長い時間を要する傾向があり、非常にコストがかかる。
- **データ管理**：データ生成メカニズムを十分に理解することが必要であるが、HTA の人間はこれに精通していないことが多い。

## 6. 結論

医療技術の追加的有用性や費用対効果を評価する際に、RCT だけでは意思決定を十分に支援することが困難になっている。関心のある対象集団・比較対照技術での有効性・安全性のエビデンスを創出するために、TTE, ECA などを含む因果推論のための新しい

考え方・方法が使用されるようになっている。内的・外的妥当性における一定の限界を認識しつつも、利用可能な RWD をフルに活用して、ステークホルダー同士が知恵を出し合い評価を進める必要がある。今後の公的分析では、とくに疫学・臨床疫学・薬剤疫学分野の専門家の関与が重要になると考えられた。

## 7. 研究発表

1. 森脇 健介. 医療経済評価におけるリアルワールドデータの活用と課題. 第 28 回日本薬剤疫学会学術総会 シンポジウム 5. (2023/11/18)
2. 森脇 健介. 費用対効果評価の論点 ―比較対照技術・外部対照選択の考え方―. 臨床統計シンポジウム. (2023/10/13)

## 8. 参考文献

- [1] The National Institute for Health and Care Excellence. NICE real-world evidence framework. Corporate document Published: 23 June 2022 [www.nice.org.uk/corporate/ecd9](http://www.nice.org.uk/corporate/ecd9) 2022.
- [2] Gomes, M., et al., Target Trial Emulation for Transparent and Robust Estimation of Treatment Effects for Health Technology Assessment Using Real-World Data: Opportunities and Challenges. *Pharmacoeconomics*, 2022. 40(6): p. 577-586.
- [3] Griffiths, E.A., et al., The Role of Noncomparative Evidence in Health Technology Assessment Decisions. *Value Health*, 2017. 20(10): p. 1245-1251.
- [4] Chan, K., et al., Developing a framework to incorporate real-world evidence in cancer drug funding decisions: the Canadian Real-world Evidence for Value of Cancer Drugs (CanREValue) collaboration. *BMJ Open*, 2020. 10(1): p. e032884.
- [5] Jaksa, A., et al., A Comparison of Seven Oncology External Control Arm Case Studies: Critiques From Regulatory and Health Technology Assessment Agencies. *Value Health*, 2022. 25(12): p. 1967-1976.
- [6] Campbell, U.B., N. Honig, and N.M. Gatto, SURF: A Screening Tool (for Sponsors) to Evaluate Whether Using Real-World Data to Support an

Effectiveness Claim in an FDA Application Has Regulatory Feasibility.  
Clin Pharmacol Ther, 2023. 114(5): p. 981-993.

## 参考資料

翻訳 NICE リアルワールド エビデンスの枠組み

**NICE real-world evidence framework**

**Corporate document [ECD9]Published: 23 June 2022]**

<https://www.nice.org.uk/corporate/ecd9/chapter/overview>

※DeepL を用いて下訳を作成し、一部、用語の修正を行った。無断転載・配布は禁止で、個人の自己学習の範囲内で利用する事。

## 目次

1.	概要.....	3
1-1.	キーメッセージ.....	3
1-2.	NICE ガイダンスにおけるリアルワールドデータとその役割.....	6
2.	NICE の意思決定におけるリアルワールドのエビデンスの紹介.....	8
2-1.	背景.....	8
	リアルワールドデータとは何か?.....	8
2-2.	リアルワールドエビデンスとは何か?.....	13
2-3.	NICE ガイダンスにおける現リアルワールドのエビデンスの活用.....	13
2-4.	リアルワールドデータを用いた介入効果の推定.....	16
2-5.	リアルワールドのエビデンスを生み出す上での課題.....	21
3.	定量的なリアルワールドエビデンス研究の実施.....	25
3-1.	キーメッセージ.....	25
3-2.	はじめに.....	25
3-3.	研究計画.....	27
3-4.	研究の実施.....	31
3-5.	研究報告.....	33
4.	データ適合性の評価.....	37
4-1.	キーメッセージ.....	37
4-2.	はじめに.....	37
4-3.	データの実証性.....	38
4-4.	データの目的適合性.....	41
4-5.	データの質.....	41
4-6.	データの関連性.....	44
5.	比較効果に関するリアルワールド研究の方法.....	47
5-1.	キーメッセージ.....	47
5-2.	はじめに.....	48
5-3.	非ランダム化研究デザインの種類.....	49
5-4.	研究デザイン.....	52
5-5.	分析.....	57
5-6.	研究の頑健性の評価.....	61
5-7.	報告.....	65
5-8.	品質評価.....	65
付録 1 -	データ適合性評価ツール (DataSAT).....	67
	DataSAT 評価テンプレート.....	67
	DataSAT - ケーススタディ.....	70

付録 2-バイアスのリスクを最小化するために用いた方法の報告 .....	77
メソッド報告テンプレート .....	77
メソッド報告 - ケーススタディ 1.....	78
メソッド報告 - ケーススタディ 2.....	81
付録 3-特定の分析法に関する報告情報 .....	84
枠組みの開発経緯.....	89
背景 .....	89
NICE 開発チーム.....	89

## 1. 概要

### 1-1. キーメッセージ

- [リアルワールドのデータ](#)は、医療と社会的ケアの提供、患者の健康と経験、そして日常的な設定における患者とシステムの結果に対する介入の効果についての理解を深めることができる。
- [2021年から2026年までのNICE戦略で述べられている](#)ように、私たちは、知識のギャップを解決し、患者のためのイノベーションへのアクセスを促進するために、リアルワールドのデータを使用したいと考えている。
- 私たちは、この野心を実現するために、リアルワールドのエビデンス枠組みを開発した。それは以下のようなものである：
  - 不確実性を減らし、ガイダンスを改善するために、リアルワールドのデータをどのような場合に使用できるかを特定する。
  - エビデンスの質と透明性を向上させるために、[リアルワールドのエビデンス](#)研究を計画、実施、報告するためのベストプラクティスを明確に説明する。
- この枠組みは、ガイダンスに情報を提供するリアルワールドのエビデンスの質を向上させることを目的としている。この枠組みは、エビデンスの質に関する最低許容基準を設定するものではない。勧告がどのようになされるかについての詳細は、NICEの関連マニュアルを参照されたい（[NICEガイダンスの項](#)を参照）。
- この枠組みは、NICEガイダンスに情報を提供するためのエビデンスを開発する人々を主な対象としている。また、患者、データ収集者、エビデンスの査読者にも関連している。
- [表1](#)は、リアルワールドエビデンス研究を実施する際の主な考慮事項をまとめたものである。高品質で信頼できるリアルワールドエビデンスを創出するためには、以下の基本原則に従うべきである：
  - データの[出所](#)が明確で、関連性があり、研究課題に答えるのに十分な品質であることを確認する。
  - 研究の計画から実施、報告に至るまで、透明性のある方法で、誠実にエビデンスを作成する。
  - バイアスのリスクを最小化し、不確実性を特徴付ける分析方法を用いる。
- この枠組みは、リアルワールドのエビデンスのさまざまな利用方法にわたって、これらの基本原則の実施を支援するための詳細なガイダンスとツールを提供する。構成は以下の通りである：

- [序章では](#)、リアルワールドデータとリアルワールドのエビデンスに関する背景を説明し、その長所と短所を論じ、NICE ガイダンスにおける現在の使用法と潜在的な使用法を要約する。
  - [研究実施に関するセクションでは](#)、エビデンスの受容可能性はエビデンスの種類やその他の文脈的要因に依存することを認識した上で、リアルワールドエビデンス研究の計画、実施、報告に対する NICE の期待を記述している。
  - [データの適切性の評価に関するセクションでは](#)、データの出所、特定の研究課題に取り組むための品質と関連性を評価するために必要な情報について説明する。
  - [比較効果に関するリアルワールド試験の方法の項では](#)、[非ランダム化比較試験](#)を実施するための、より具体的な推奨事項を示している。これには、従来の観察研究だけでなく、[外部対照](#)を形成するためにリアルワールドデータを用いる臨床試験も含まれる。
- この枠組みは生きた枠組みであり、ユーザーからのフィードバック、模範となるケーススタディを含む実施からの学び、リアルワールドのエビデンス手法の発展、優先的なトピックに関する追加ガイダンスを含む範囲の拡大などを反映するために、定期的に更新される予定である。
  - 我々は、リアルワールドデータを申請書類に使用する予定の企業に対し、エビデンス作成計画の一環として、リアルワールドデータをどのように最良に活用するかについて、[NICE 科学的助言](#)に早期に関与するよう奨励している。

表 1

リアルワールドのエビデンス研究を計画、実施、報告する際に考慮すべき重要事項のまとめ

エビデンス作成の段階	主な検討事項
<a href="#">プランニング</a>	<ul style="list-style-type: none"> <li>• 研究課題を明確に定義し、関連する場合は、集団の適格基準、介入、アウトカム、推計対象量の定義を含む。</li> <li>• 事前に研究を計画し、プロトコル（データ分析計画を含む）を公開する。</li> <li>• 十分な実績があり、研究課題を解決するのに十分な品質と関連性のあるデータを選択する（<a href="#">データの適切性の評価に関するセクション</a>を参照）。</li> <li>• <a href="#">一次データ</a>収集を計画する際には、患者や医療従事者の負担を最小限に抑えつつ、患者中心の方法で収集を実施する方法を検討する。</li> <li>• 現地の法律、<a href="#">データガバナンスプロセス</a>、行動規範、および<a href="#">データ管理者</a>の要件に従ってデータを使用する。</li> </ul>
<a href="#">行動</a>	<ul style="list-style-type: none"> <li>• バイアスの主要なリスクを考慮し、研究課題に適した研究デザインと統計的手法を使用する。</li> <li>• 感度分析及び／又はバイアス分析を用いて、バイアスの主要なリスクや不確実な<a href="#">データキュレーション</a>又は分析上の決定に対する研究の頑健性を評価する。</li> <li>• 研究の完全性と質を保証するための品質保証基準とプロトコルを作成し、実施する。</li> </ul>
<a href="#">報告</a>	<ul style="list-style-type: none"> <li>• 独立した研究者が、何がなぜ行われたかを十分に理解し、研究を批判的に評価し、再現できるように、研究デザインと分析方法を十分に詳細に報告すること。</li> <li>• 報告書はまた、それをカバーするものでなければならない：             <ul style="list-style-type: none"> <li>○ データの出所、品質、関連性（<a href="#">データの適切性の評価に関するセクション</a>を参照）</li> <li>○ データキュレーション</li> <li>○ 初期データから最終解析までの患者の減少</li> <li>○ 患者の特徴（欠測データを含む）、全体および主要集団におけるフォローアップの詳細</li> <li>○ 計画され実施されたすべての解析の結果（事前に計画されていなかった解析は明確に示すこと）</li> <li>○ バイアスのリスクとNHSの対象集団への一般化可能性の評価</li> <li>○ 研究の限界と結果の意味の解釈</li> </ul> </li> </ul>

## 1-2. NICE ガイダンスにおけるリアルワールドデータとその役割

- **リアルワールドデータとは**、高度に管理された臨床試験以外で収集された、患者の健康や経験、ケア提供に関するデータを指す。患者健康記録、管理記録、患者登録、調査、観察コホート研究、デジタルヘルス技術など、様々な情報源から得られる。
- リアルワールドのデータは、例えば NICE ガイダンスに情報を提供するために、すでに広く利用されている：
  - 健康状態、介入、ケア経路、患者のアウトカムと経験を特徴づける。
  - **経済モデル**（資源使用、QOL、イベント発生率、有病率、発生率、長期アウトカムの推定を含む）の設計、入力、検証
  - デジタルヘルス技術を開発または検証する（例えば、デジタル技術は、リアルワールドのデータを使用して開発された臨床アルゴリズムを使用することができる）。
  - **健康の不平等を**特定し、特徴づけ、対処する。
  - 医薬品、instrument、インターベンショナル手技を含む医療技術の安全性を理解する。
  - 介入（検査を含む）がサービス提供やケアに関する決定に与える影響を評価する。
  - NHS の患者に対する臨床試験の適用性を評価する。
- 対象集団を代表するリアルワールドのデータは、NICE がこれらの用途のほとんどに推奨するエビデンス源である。このようなデータは NICE ガイダンスの中で定期的に使用されているが、特に日常的に収集されたデータについては、その使用がより一般的になる可能性がある。
- ランダム化比較試験は、介入の効果に関するエビデンスとして好ましいものである。ランダム化により、群間のベースライン特性の差が偶然によるものであることが保証される。盲検化が適用されていれば）治療割り付けの知識が行動に影響するのを防ぐことができる。しかし、ランダム化比較試験が利用できない場合や、NHS における患者ケアの決定に直接関係しない場合もある。
- ランダム化比較試験は、以下のような理由で実施できない場合がある：
  - ランダム化が倫理的でない、あるいは実行不可能であると考えられる場合（例えば、アンメットニーズのある希少疾患や重篤な疾患など）
  - 技術的な課題によりランダム化がリアルワールド的でない。これは医療 instrument や介入手技で最もよく見られる。
  - 臨床試験のための資金が得られない場合（例えば、介入がすでに日常診療で使用されている場合）。

- ランダム化されたエビデンスがあったとしても、NHSにおける意思決定には十分ではないかもしれない：
  - 比較対照が NHS の標準治療を反映していない。
  - 該当する集団は除外される
  - 患者の行動、ケア経路、または設定において、日常診療の実施とは異なる大きな違いがある。
  - フォローアップに限界がある
  - 検証されていない代替アウトカムが使用されている
  - 学習効果がある
  - 臨床試験の質が低い
- [非ランダム化比較試験](#)は、医療 instrument や処置、公衆衛生上の介入の効果を推定するために、すでに広く用いられている。非ランダム化比較試験は、非対照[単群試験に基づいて承認される医薬品](#)が増えるにつれ、医薬品の初期評価においてより広く用いられるようになってきている。最後に、技術のライフサイクル評価と患者の生活体験に焦点が当てられるようになったため、最初の承認後の非ランダム化比較試験に依存している。比較効果を評価するためのリアルワールドデータを用いた非ランダム化比較試験で最も一般的なものは、観察コホート試験と[実外部対照](#)を用いた単群試験である。
- エビデンスのギャップを埋め、患者へのアクセスを早めるために、リアルワールドデータがより日常的に利用される可能性がある。この約束を実現するためには、リアルワールドエビデンス研究は、透明かつ誠実に実施され、目的に合ったデータを使用し、バイアスの主要なリスクに対処しなければならない。
- 私たちは、リアルワールドエビデンス研究が透明性をもって行われ、質の高いものであることを保証するために、リアルワールドエビデンス研究の実施に関するベストプラクティスについて、私たちの見解を伝えている。これは、リアルワールド試験とその意思決定への利用に対する信頼を向上させるために不可欠である。

## 2. NICE の意思決定におけるリアルワールドのエビデンスの紹介

- [背景](#)
- [リアルワールドデータとは何か？](#)
- [リアルワールドエビデンスとは何か？](#)
- [NICE ガイダンスにおける現リアルワールドのエビデンスの活用](#)
- [リアルワールドデータを用いた介入効果の推定](#)
- [リアルワールドのエビデンスを生み出す上での課題](#)

### 2-1. 背景

[リアルワールドのデータ](#)は、医療と社会的ケアの提供、患者の健康と経験、そして日常的な設定における患者とシステムの結果に対する介入の効果についての理解を深めることができる。

[2021 年から 2026 年までの NICE 戦略で述べられているように](#)、私たちは、知識のギャップを解決し、患者のためのイノベーションへのアクセスを促進するために、リアルワールドのデータを使用したいと考えている。

私たちは、この野心を実現するために、リアルワールドのエビデンス枠組みを開発した。それは以下のようなものである：

- 不確実性を減らし、ガイダンスを改善するために、リアルワールドのデータをどのような場合に使用できるかを特定する。
- エビデンスの質と透明性を向上させるために、[リアルワールドのエビデンス](#) 研究を計画、実施、報告するためのベストプラクティスを明確に説明する。

この枠組みは、ガイダンスに情報を提供するリアルワールドのエビデンスの質を向上させることを目的としている。この枠組みは、エビデンスの質に関する最低許容基準を設定するものではない。勧告がどのようになされるかについての詳細は、NICE の関連マニュアルを参照されたい ([NICE ガイダンスにおけるリアルワールドデータ活用の項参照](#))。

この枠組みは、NICE ガイダンスに情報を提供するためのエビデンスを開発する人々を主な対象としている。また、患者、データ収集者、エビデンスの査読者にも関連している。

リアルワールドデータとは何か？

リアルワールドデータとは、高度に管理された[臨床試験](#)以外で収集された、患者の健康や経験、ケア提供に関するデータと定義する。リアルワールドデータは、医療や社会的ケアの提供中に日常的に収集される。また、1つ以上の特定の研究課題に取り組むために前向きに収集されることもある。ほとんどのリアルワールドデータ源は観察研究（または非介入研究）であり、介入（または曝露）は研究プロトコルによって決定されない。その代わりに、医療介入は患者と医療専門家によって決定される。また、公衆衛生や社会的ケアでは、介入は個人の行動や環境暴露、政策立案者によって決定されることもある。

プラグマティック臨床試験のような介入試験の中にも、[リアルワールドのエビデンス](#)が得られるものがある。このような試験では、試験の計画、参加者の募集、アウトカムデータの収集にリアルワールドのデータ源を利用することもある。詳細については、[英国医薬品・ヘルスケア製品規制庁（MHRA）の「規制上の決定を支援するためのリアルワールドデータを用いたランダム化対照試験に関するガイドライン」](#)を参照のこと。

[表 2](#)に非介入リアルワールドデータの一般的な情報源を示す。これらのデータには、オリジナルのデータ集（患者の健康記録など）と、オリジナルの情報源から収集したデータ（レトロスペクティブなカルテレビューから得られたデータなど）が含まれる。各タイプのデータ源には一般的な長所と短所があるが、ある研究課題に対する価値は特定のデータの特徴によって決まる（詳細は[データの適合性の評価のセクション](#)を参照）。リアルワールドの様々なデータ源は、データの質と範囲を向上させるために、リンクやプールによって組み合わせることができる。

リアルワールドのデータには定量的なものと定性的なものがある。一般的なデータには、患者の人口統計、健康行動、病歴、臨床アウトカム（患者報告によるアウトカムを含む）、患者や利用者の経験、資源利用、コスト、[オミックス](#)、検査室測定、画像、フリーテキスト、検査結果、患者が作成したデータなどがある。推奨を行う際には、国内のデータ収集と国際的なデータの両方を考慮する。

## 表 2

## リアルワールドデータの一般的な情報源

データソース	説明	例
電子カルテ	<p>コンピュータ化された個々の患者の記録。これらは通常、患者の臨床管理に利用される。</p> <p>検査室、ゲノム、画像システムなど、他の情報システムからのデータを統合することもある。</p>	<p><a href="#">Clinical Practice Research Datalink (CPRD)</a> GOLD には、英国内の一般診療所に登録された患者の人口統計学的情報と臨床情報が含まれている。</p>
行政データ	<p>医療・福祉サービスによって管理目的で収集されたデータ。</p>	<p><a href="#">Hospital Episode Statistics (HES)</a> Admitted Patient Care データセットには、NHS の病院に入院したすべての患者、あるいは NHS の資金で私立病院で治療を受けたすべての患者の診断と処置に関する情報が含まれている。HES の主な目的は、PBP (payment by results) やその他の運営活動を通じて、病院の診療報酬に反映させることである。</p>
クレームデータ	<p>医療サービスの利用に関する行政データの一つで、保険システムから収集されることが多い。</p>	<p><a href="#">メディケア&amp;メディケイドサービスセンターのデータ</a>には、払い戻し情報または請求書の支払いから得られたメディケアサービスを受けた個人に関するデータが含まれる。</p> <p><a href="#">NHS Business Services Authority</a> は、イングランドのプライマリーケアで調剤された医薬品に関するデータを提供している。</p>
患者登録	<p>登録は、特定の疾患、状態、または曝露によって定義された集団の特定の結果を特定するために、統一されたデータ（臨床およびその他）を収集する組織化されたシステムである。</p>	<p><a href="#">Systemic Anti-Cancer Therapy (SACT) データセット</a>には、NHS イングランドのプロバイダーから抗がん剤治療を受けた全患者の情報が含まれている。このデータは NICE 内で広く使用され、Cancer Drugs Fund 内で</p>

## リアルワールドデータの一般的な情報源

データソース	説明	例
	<p>登録には、研究、臨床ケア、政策などいくつかの目的がある。登録には介入研究も含まれる。</p>	<p>使用が承認された薬剤に関する情報を提供している。</p> <p><a href="#">UK Cystic Fibrosis Registry</a> は、英国内の嚢胞性線維症の専門センターにおいて、同意を得た嚢胞性線維症患者のデータを収集しています。登録データは、研究を促進し、ケアセンターにおける質の向上を導き、新薬の安全性を監視することで、嚢胞性線維症の人々の健康を改善するために使用されます。</p>
<p>患者が作成した健康データ</p>	<p>ウェアラブル医療 instrument や個人用 instrument、モバイルアプリ、ソーシャルメディア、その他のインターネットベースのツールから、患者やその介護者が直接生成したデータ。データは、能動的に（例えば、人々がフォームにデータを入力することによって）収集されることもあれば、受動的に（例えば、人々の活動レベルを測定するスマートウォッチによって）収集されることもある。</p>	<p>入院の必要性を警告するため、在宅で治療を受けている COVID-19 患者のモニターにパルスオキシメーターが使用されている (<a href="#">Greenhalgh ら、2021 年</a>)。</p> <p><a href="#">ZOE アプリ</a> から得た COVID-19 と long-COVID 症状に関する自己申告データ。</p>
<p>カルテレビュー</p>	<p>患者の医療記録（紙または電子記録を含む）のレビューからレトロスペクティブに抽出されたデータ。</p> <p>カルテレビューは自然史研究において広く用いられている。ルーチンのデータソースでは報告されていないデータを抽出することができる。</p>	<p>レトロスペクティブ・カルテレビューは、疾患の自然史や治療経路をモデル化するために、希少疾患の研究では特によく行われる (<a href="#">Garbade et al.</a>)</p>

## リアルワールドデータの一般的な情報源

データソース	説明	例
<p>監査とサービス評価</p>	<p>臨床監査は、現在の医療水準がベストプラクティスや設定された基準に照らしてどのようなものであるかを理解し、質の改善につなげるために行われる。データは前向きに収集されることもあれば、後向きに収集されることもある。</p> <p>サービス評価は、現在のケアを定義し判断するために行われる。</p>	<p><a href="#">医療の質向上パートナーシップ</a>は、<a href="#">センチネル脳卒中全国監査プログラム (SSNAP)</a> などの全国臨床監査プログラムを管理している。SSNAP は、イングランド、ウェールズ、北アイルランドにおける集学的脳卒中入院医療サービスの組織と提供の質を評価するために使用される。</p>
<p>一次データ収集による観察コホート</p>	<p>1 つ以上の研究課題に答えるためにデザインされた伝統的な前向き研究。</p>	<p><a href="#">英国バイオバンク</a>は、患者の病歴と遺伝学に関するデータを収集している。健康アウトカムについては患者の記録とリンクしている。特定の研究課題のために設計されたのではなく、疫学研究を可能にするためである。</p> <p><a href="#">EMBRACE-I</a> は、局所進行子宮頸部腫瘍に対して MRI を用いた画像誘導適応ブラキセラピーを受けた患者の局所腫瘍制御と罹患率を評価する多施設共同前向きコホート研究である。</p>
<p>健康調査、インタビュー、フォーカス・グループ</p>	<p>健康調査とは、調査を通じて、ある集団における健康と病気に関するデータを系統的に収集することである。集団の健康状態の傾向を把握したり、患者の医療体験を理解したりするなど、さまざまな目的がある。</p> <p>インタビューやフォーカスグループは、患者の認識や経験などの質的データを収集するために行われる。</p>	<p><a href="#">イングランド健康調査</a>は、イングランドにおける健康の傾向を測定するために毎年行われる代表的な世帯調査である。</p> <p>患者チャリティー団体 Lipoedema UK による <a href="#">'Living with Lipoedema' 2021 年調査</a> は、脂肪浮腫患者の体験データを収集するものである。この調査は、脂肪浮腫のために美容整形以外の脂肪吸引やその他の治療を受けた患者の経験を評価するものである。</p>

## 2-2. リアルワールドエビデンスとは何か？

リアルワールドのエビデンスとは、リアルワールドのデータの分析から得られたエビデンスと定義する。このエビデンスは、疾病疫学、医療サービス調査、因果関係の推定など、様々なエビデンスを網羅することができる（[NICE ガイダンスのリアルワールドデータ利用の項を参照](#)）。研究課題や事例に応じて、様々な研究デザインや分析方法（量的・質的方法を含む）から作成することができる。リアルワールドエビデンス研究では、ルーチンに収集されたデータ、特注のデータ収集、またはその2つの組み合わせを用いることができる。我々は、外部対照を作成するためにリアルワールドデータソースを使用する[単群試験](#)をリアルワールドエビデンス試験とみなしている。

## 2-3. NICE ガイダンスにおける現リアルワールドのエビデンスの活用

### 2-3-1. NICE ガイダンス

NICE には、入手可能な最善のエビデンスを用いて、保健、公衆衛生、社会的ケアにおける意思決定の指針となる勧告を作成するガイダンスがいくつかある：

- 臨床、社会的ケア、公衆衛生のトピックに関するガイドラインで、医療および社会的ケアの専門家に助言ガイダンスを提供する。
- 医薬品、診断薬、医療 instrument、デジタルヘルス技術、インターベンション治療などの医療技術の評価。

ガイドラインは NICE 内部で作成される。技術評価は通常、企業が提出したエビデンスに基づいて行われるが、メーカーやその他の関係者から提出されたエビデンスや、独立した学術センターから委託された研究を利用することもある。

技術評価のプロセスと方法は、NICE のプログラムによって異なる。技術評価プログラムでは、主に医薬品（高度に専門化された技術を含む）を評価するが、医療 instrument や診断薬も含まれる。技術評価プログラムと診断指導プログラムは、ともに医療技術の費用対効果を検討する。医療技術評価プログラムでは、医療 instrument、デジタルヘルス技術、診断薬など、コスト削減またはコスト中立が期待される医療技術を評価し、患者とシステムのアウトカムを考慮した費用効果分析を行う。処置・介入プログラムは、コスト分析を行わずに処置・介入の有効性と安全性を評価する。

NICE が技術評価プログラム（Technology Appraisal Programme）を通じて治療法を " 選択肢として " 推奨した場合、NHS はその発表日から 3 ヶ月以内（特に指定がない限り）にその治療法が利用できるようにしなければならない。費用対効果が見込めるが、その価

値について相当かつ解決可能な不確実性がある技術については、マネージドアクセス契約での使用を推奨することができる。一定期間のリアルワールドデータ収集後、その技術は技術評価プログラム（Technology Appraisal Programme）によって再評価される。NICE ガイダンスで推奨され、導入後 3 年以内に手頃な価格でコスト削減が期待できる instrument、診断技術、デジタル技術を選択した場合は、[NHS イングランドの MedTech funding mandate](#) を通じて資金を提供することができる。

様々な NICE プログラムについて、方法とプロセスのマニュアルが作成されている。本枠組みの利用者は、適宜これらのマニュアルを参照されたい：

- [NICE ガイドラインの開発：このマニュアルは](#)、NICE ガイドラインの開発と更新に使用されるプロセスと方法について説明しています。
- [NICE の医療技術評価マニュアルには](#)、診断薬評価プログラム、医療技術評価プログラム、高度専門技術プログラム、技術評価プログラムなど、医療技術評価を開発するための方法とプロセスが記載されている。
- [NICE の介入手技プログラムマニュアルには](#)、介入手技プログラムにおけるガイダンス作成のプロセスと方法が記載されている。

[NICE のデジタルヘルス技術に関するエビデンス基準の枠組みは](#)、デジタルヘルス技術に関する優れたエビデンスレベルがどのようなものかを定めている。これは、デジタルヘルス技術の革新者と委託者を対象としている。

### 2-3-2. リアルワールドデータの使用例

NICE のガイダンスプログラム間の違いは、リアルワールドのエビデンスの用途や受容性にばらつきをもたらす。

リアルワールドデータは、NICE のプログラム全体にわたって、特に介入の効果についてではない質問について、さまざまなタイプのエビデンスを生み出すためにすでに使用されている。過去の NICE ガイダンスの例を以下に挙げる：

- 健康状態、介入、ケア経路、患者のアウトカムと経験（自然史を含む）の特徴づけ：[脊髄性筋萎縮症治療のためのオナセムノゲン・アベパルボベクに関する NICE 高度専門技術ガイダンスは](#)、脊髄性筋萎縮症を特徴づけるために複数のリアルワールドデータを用いた。
- 経済的負担の推定 [重症好酸球性喘息治療薬ベンラリズムマブに関する NICE 技術評価ガイダンスは](#)、HES にリンクされた CPRD GOLD のデータを報告した。

- [経済モデル](#)を設計し、入力し、検証する。一般的なエビデンスの種類は以下の通り：
  - 患者開始時の特徴 [侵襲的冠動脈造影時の冠動脈狭窄を評価するための QAngio XA 3D QFR および CAAS vFFR 画像ソフトに関する NICE 診断ガイドランス](#)の IRIS-FFR レジストリのデータが報告された。
  - ベースラインのイベント発生率 [NICE ガイドライン「16 歳以上の慢性閉塞性肺疾患（COPD）：診断と管理」](#)では、CPRD GOLD からベースラインの疾患重症度別 COPD 増悪率のデータが報告されている。
  - 日常診療における治療の特徴 [Dravet 症候群の発作治療薬フェンフルラミンに関する NICE 技術評価ガイドランス](#)では、日常診療における比較対照治療薬の平均投与量を評価するために、複数のリアルワールド試験が用いられた。
  - 健康状態または疾患進行の間の遷移確率：[高カリウム血症治療薬パティロマーに関する NICE 技術評価ガイドランス](#)では、CPRD データを用いて慢性腎臓病患者の病態遷移をモデル化した。
  - リソースの使用とコスト [NICE の医療技術ガイドランスである HeartFlow FFRCT（冠動脈 CT 血管造影から血流予備能を推定するための FFRCT）](#)は、NHS の参照費用から冠動脈血行再建術の費用データを使用した。
  - QOL を含む患者報告アウトカム [ムコ多糖症 4a の治療に対するエロスルファーズアルファに関する NICE の高度専門技術ガイドランス](#)では、調査から得られた QOL データが使用された。
  - 外挿：[未治療の広範囲小細胞肺癌に対するアテゾリズマブとカルボプラチンおよびエトポシドの併用療法に関する NICE 技術評価ガイドランス](#)は、米国の電子カルテから得られた Flatiron Health データベースのデータを使用した。
- 患者の経験を測定する [COPD 管理のための myCOPD に関する NICE 医療技術ガイドランス](#)では、技術の使いやすさに関する患者調査データを使用した。
- 予後予測モデルを含むデジタルヘルス技術の開発と検証：詳細については、[デジタルヘルス技術に関する NICE のエビデンス基準の枠組み](#)を参照のこと。
- 健康格差の特定、特徴づけ、対処 [鎌状赤血球症における鎌状赤血球クリーゼ予防のためのクリザンリズマブに関する NICE 技術評価ガイドランス](#)は、特定の少数民族グループにおける鎌状赤血球症の健康状態および不均衡な負担に関する全国血色素症登録からのエビデンスを報告した。
- バイオマーカーのような検査精度の推定や検査結果の再現性：[NICE の医療技術ガイドランス](#)で報告された Zio XT による不整脈の検出に関するレトロスペクティブ観察コホート研究のデータ

- 人工関節置換術（一次）：股関節と膝関節 [人工関節置換術（一次）に関する NICE ガイドライン：股関節、膝関節、肩関節](#)人工膝関節置換術の再置換率に関する National Joint Registry のデータを使用した。
- 介入（検査を含む）がサービス提供やケアに関する決定に与える影響を測定する：[早期乳癌における術後補助化学療法の決定を導くための腫瘍プロファイリング検査に関する NICE 診断学ガイダンスは](#)、いくつかの前向き観察研究の結果を報告した。

リアルワールドデータは、NHS の患者に対する試験結果の適用性を評価したり、介入効果を推定したりするためにも用いることができる（詳細については、[リアルワールドデータを用いた介入効果の推定の項](#)を参照）。

リアルワールドエビデンスは、こうしたエビデンスの種類多くにすでに広く利用されているが ([Leahy et al. 2020](#), [Makady et al. 2018](#))、その利用はより一般的になる可能性がある。データが対象集団を代表するものであり、十分な質のものであれば、それが望ましいデータ源となる可能性がある。バックグラウンドのイベント発生率や臨床試験からの自然歴データは、選択的採用のため、対象集団のイベント発生率を過大評価したり過小評価したりすることがある ([Bloudek et al. 2021](#))。場合によっては、研究課題への適用性が低い公表されたエビデンスに依存するよりも、日常的に収集されたデータを用いて研究を実施することに価値があるかもしれない。

## 2-4. リアルワールドデータを用いた介入効果の推定

### 2-4-1. ランダム化比較試験の利用と課題

[ランダム化比較試験は](#)、介入の因果効果を推定するために望ましい研究デザインである。これはランダム化により、群間の既知および未知のベースライン特性における差が偶然によるものであることが保証されるからである。盲検化が適用されていれば）治療割り付けの知識が行動に影響することを防ぎ、標準化されたプロトコルが一貫したデータ収集を保証する。

しかし、ランダム化比較試験は常に利用できるとは限らず、また関心のある研究課題を解決するのに十分でない場合もある。

ランダム化比較試験は、以下のような理由で実施できない場合がある：

- ランダム化は、例えばアンメットニーズが高いなどの理由で、倫理的に問題があると考えられる。

- 患者が試験中の介入のいずれかに割り付けられることを望まない場合。
- 医療専門家は、患者を効果が低いと思われる介入にランダムに割り付けたがらない。
- 少数の対象患者
- 研究の財政的または技術的制約
- すべての治療の組み合わせ（治療順序を含む）を直接評価できるわけではない。

ランダム化比較試験は、希少疾患、革新的で複雑な技術、あるいは特定の集団においては、特に困難な場合がある。

同様に、医療 instrument や介入手技の場合、盲検化の難しさ、学習効果の重要性、標準治療の変化による比較対照の選択の難しさ、性能に影響を与える可能性のある経時的な技術の特性の変化、研究能力や資金調達制限などの理由から、質の高いランダム化比較試験は困難な場合がある [\(Bernard et al.\)](#)

たとえ臨床試験があったとしても、以下の理由から、研究課題や NHS の日常診療に直接適用できない場合がある：

- NHS における標準治療を代表しない比較対照（[プラセボ](#)対照を含む）の使用
- 検証されていない代替アウトカムの使用
- 限定フォローアップ
- 対象集団の除外（例えば、合併症のある人、妊婦、小児など）
- NHS の対象集団への結果の移植性に影響を与える集団、ケア経路、設定の違い
- 患者の技術使用の違い
- 日常診療とは異なる臨床支援
- 学習効果（すなわち、ユーザーが経験を積むにつれて、介入の効果が時間とともに変化すること）
- 治療法の切り替え、追跡不能、データの欠落などのランダム化後の事象に対処するために使用された方法。

NHS における標準治療を代表するものではない比較対照の使用など、これらの課題のいくつかは、試験の比較可能性に関する一定の仮定の下で、[ネットワークメタ解析](#)のような他のアプローチで対処できる可能性がある。詳細については、[エビデンスの情報源と統合に関する NICE 意思決定支援ユニットの報告書](#)を参照のこと。

リアルワールド試験のように、デザインにリアルワールドデータを用いたり、アウトカムの測定にリアルワールドデータを用いたりするランダム化比較試験からも、リアルワール

ドエビデンスを得ることができる。このような臨床試験は、ランダム化による内的妥当性と、日常診療から得られたデータの一般化可能性を組み合わせることで、大きな価値を提供することができる。[英国 MHRA は、ランダム化対照試験からリアルワールドエビデンスを作成するためのガイダンスを発表している。](#)

## 2-4-2. リアルワールドのエビデンス

リアルワールドのデータは、ランダム化比較試験の状況を説明するため、試験がない場合の介入の効果を推定するため、あるいは日常的な設定における介入の影響についてのより広範な疑問に答えるために試験を補完するために用いることができる。

### 2-4-2-1. 文脈化

文脈化（Contextualisation）とは、臨床試験の結果が NHS の対象集団にうまく反映されるかどうかを評価することである。これは NICE プログラム全体におけるリアルワールドデータの重要な利用法であるが、NICE は、費用対効果の可能性のある医薬品について、マネージドアクセスの取り決めを通じて、不確実性がさらなるデータ収集によって対処できる場合には、さらなるデータ収集を要求することがある。このデータは、NHS における臨床試験の妥当性を理解するために用いられることが多い。

NICE ガイダンスでは、臨床試験の背景を説明するためにリアルワールドデータが用いられている：

- NHS における対象集団、治療経路、ケア環境、アウトカムの違い [発作性夜間ヘモグロビン尿症治療薬ペグセタコプランに関する NICE 技術評価ガイダンス](#)では、英国の登録データを用いて、英国の診療所における尿中ヘモグロビン濃度がランダム化比較試験の適格基準値と一致していることを示した。
- サロゲート・アウトカムと最終 [アウトカム](#)（患者報告アウトカムを含む）の関係をモデル化した：[原発性高シュウ酸尿症 1 型治療薬ルマシランに関する NICE の高度専門技術ガイダンス](#)では、代用アウトカムである血漿シュウ酸塩と腎機能の関係をモデル化するために、レジストリに基づく研究が用いられた。
- 介入策の使用とアドヒアランスを測定する：[不眠症と不眠症状を治療する Sleepio に関する NICE 医療技術ガイダンス](#)は、アプリやウェブサイトから収集した使用状況に関するデータを使用した。
- 試験期間を超えた長期的アウトカムや治療効果に関する仮定の妥当性を評価する：[進行性線維性間質性肺疾患治療薬ニンテダニブに関する NICE 技術評価ガイダンス](#)では、長期アウトカムの外挿を検証するためにレジストリデータを用いた。

[EGFR T790M 変異陽性進行非小細胞肺癌に対するオシメルチニブの NICE 技術評価ガイドランスは、全身抗癌剤治療（SACT）データセットのデータを用いて、AURA3 試験の結果の NHS 患者への関連性を評価した。特に、SACT データは比較に使用された：](#)

- 全生存期間
- 年齢、民族、パフォーマンスステータス、治療歴などの患者特性の違い。

#### 2-4-2-2. 推定

効果は、以下のようなさまざまな結果について推定することができる：

- 患者アウトカム - 臨床アウトカム、バイオマーカー、患者報告アウトカム、行動変容、利用者満足度、エンゲージメント
- システムのアウトカム-資源の使用、コスト、ケアのプロセス。

リアルワールドデータは、ライフサイクルにわたる介入の効果をよりよく理解するために用いることができる。介入の効果を推定するためのリアルワールドデータの利用可能性は、そのライフサイクルの段階によって異なる。

新規の介入（例えば、英国で最近販売許可を得たもの）については、NHS におけるその使用とアウトカムに関するリアルワールドデータは限られている。リアルワールドデータの利用には以下のようなものがある：

- [単一群間比較試験に対する](#)効果を推定するため、あるいはランダム化比較試験の対照群に追加するために、比較対照群（すなわち[外部対照](#)）を作成する：[NICE の高度専門技術ガイドランスでは、脂肪異栄養症治療のためのメトレプチンについて、単群試験の外部対照を形成するために自然史研究を用いている。](#)
- [医薬品への早期アクセススキーム](#)からのデータを使用：[遺伝性血管性浮腫の再発予防のためのベロトラルスタットに関する NICE 技術評価ガイドランスでは、長期的なアウトカムに関する不確実性を低減するために、医薬品への早期アクセススキームのデータが用いられた。](#)
- 英国よりも早く技術が利用可能になった他国での比較効果の推定 ([Jonsson et al.](#))
- 例えば、すべての適格患者の特徴を反映させるために、臨床試験の結果を再重み付けすることによって、日常的な設定におけるアウトカムと治療効果を予測することができる：[未治療の転移性扁平上皮非小細胞肺癌に対するペムブロリズマブとカルボプラチンおよびパクリタキセルの併用療法に関する NICE の技術評価ガイドランス](#)

[では](#)、Cancer Drugs Fund の処方データを用いて、サブグループの有病率で重み付けしたアウトカムを推定した。

医療技術が日常的に、あるいはパイロットプロジェクトで使用されるようになれば、リアルワールドのデータを得る機会はより大きくなり、以下のようなものが含まれる：

- 日常的な設定における介入の効果の推定（[冠動脈造影および末梢血管造影における急性腎障害のリスク低減のための DyeVert システムに関する NICE 医療技術ガイダンス](#)を参照）。
- HES データは、好ましい比較対照との直接比較を提供するものである：[前治療歴のある菌状息肉症およびセザリー症候群に対するモガムリズマブの NICE 技術評価ガイダンス](#)では、ランダム化比較試験の介入群に対する英国特有の標準治療比較群を提供するために HES データを使用した。
- カシリビマブ+イムデビマブ、ニルマトレルビル+リトナビル、ソトロビマブ+トシリズマブは、COVID-19 の治療薬として使用された：[カシリビマブ+イムデビマブ、ニルマトレルビル+リトナビル、ソトロビマブ、トシリズマブの COVID-19 治療に関する NICE 技術評価ガイダンス](#)では、OpenSAFELY の電子カルテデータを用いて臨床試験データで観察されたアウトカムを支持し、臨床試験データから除外された高リスク集団を含めた。
- 介入効果の不均一性を探る[幹細胞移植または少なくとも 2 回の前治療歴のある再発または難治性の古典的ホジキンリンパ腫に対するペムブロリズマブの NICE 技術評価ガイダンス](#)では、幹細胞移植歴のない患者における全生存期間をモデル化するために SACT データが使用された。
- 最終的な[アウトカム](#)（代替アウトカムではなく）に対する効果を、より長い期間にわたって推定する。
- ランダム化比較試験で検討されていない併用療法（シーケンスを含む）または決定戦略に関する効果の推定（[Fu et al.](#)）
- 例えば、プライアへの情報提供、検出力の向上、[ネットワークメタ分析](#)におけるエビデンスのギャップを埋めるなど、エビデンスの統合に組み込むことができる（[NICE Decision Support Unit report on sources and synthesis of evidence, Sarri et al.](#)）

### 2-4-3. 介入効果を推定するためのリアルワールドエビデンスの妥当性

ランダム化対照試験と比較して、リアルワールドのエビデンス（より一般的には[非ランダム化比較試験](#)）の内的[妥当性](#)を理解することを目的とした文献が増加している。これに

は、同じ問題に取り組む異なるデザインの研究の結果を比較するメタ疫学研究 ([Woolacott et al. 2017](#))、個々の事例研究 ([Dickerman et al.](#)

これらの研究により、質の高い非ランダム化比較試験が、多くの場合（もちろん全ての状況ではないが）、相対的治療効果の有効な推定値を算出できることが示された。有効な推定値の可能性を高めるために、以下のような共通のデザイン原則がある：

- アクティブ比較対照（同一または類似の適応に対する代替介入、通常は同一モダリティ）の使用
- 介入を以前から利用している人（先行利用者）ではなく、新規利用者（または介入開始者）の比較

妥当性は、疾患の特徴、アウトカムの種類（客観的な臨床アウトカムが望ましい）、治療状況、データの内容や質など他の要因にも左右される。

## 2-5. リアルワールドのエビデンスを生み出す上での課題

リアルワールドのデータは、日常的な設定における介入の価値についての理解を向上させる大きな可能性を秘めている。しかし、確実な結果を生み出し、エビデンスの信頼性を向上させるためには、対処しなければならない重要な課題がある。以下に主な課題を述べる。

### 2-5-1. リアルワールドのエビデンス研究への信頼

リアルワールドのデータは複雑であることが多く、分析する前にかなりの準備が必要である。また、比較効果の推定など、用途によっては分析方法が高度になることもある。既に収集されたデータを利用する場合、研究者は[統計分析計画](#)を確定する前にデータにアクセスすることがある。データの準備や分析上の決定は、結果として得られる推定値に重要な影響を及ぼす可能性がある。

したがって、得られたエビデンスの完全性と信頼性（例えば、データの浚渫やチェリーピッキングに起因するもの）に関する懸念に対処する必要がある。データの合法的使用に関する懸念は、COVID-19 の治療における再利用薬の有効性に関する著名な研究が著名な医学雑誌から撤回されたことで浮き彫りになった。

リアルワールドのエビデンス研究に対する信頼は、次のようにして向上させることができる：

- 試験実施前に試験実施計画書を登録する ([Real-World Evidence Transparency Initiative](#) を参照)。
- 報告用チェックリストまたはツール ([健康研究の質と透明性の向上\[EQUATOR\]ネットワーク](#)参照)
- データアクセスと研究実施の完全性を確認するための著者声明を要求する ([2020年、ランセット・グループ編集者による撤回からの学び](#)を参照)。
- データ、[コードリスト](#)、解析コードの公開
- 安全なデータ環境を通じてデータへのアクセスを提供し、監査証跡を維持する ([保健社会福祉省の報告書「より良く、より広く、より安全に：研究と分析のための保健データの利用」](#)を参照)。

リアルワールドエビデンス研究の計画、実施、報告に関するガイダンスは、リアルワールドエビデンスを創出するための[リアルワールドエビデンス研究の実施に関するセクション](#)を参照のこと。

## 2-5-2. データの質と関連性

リアルワールドのデータを使用する際には、いくつかの共通の課題がある。常にではないが、リアルワールドのデータソースにはないタイプのデータもよくある（腫瘍の大きさや機能状態の測定値など）。しかし、医師のメモのような[非構造化データからデータ要素](#)を抽出する方法は、ますます使用されるようになってきている。

他の変数は、十分に詳細なレベルで収集されないかもしれない。例えば、研究では特定の薬剤や医療 instrument に関する知識が必要であるが、データには薬剤や instrument のクラスしか含まれていない場合がある。同様に、ある研究では出血性脳卒中と虚血性脳卒中を区別する必要があるが、データソースにはそれ以上の詳細がなく、すべての脳卒中に関するデータが含まれている場合がある。関連する項目が必要な粒度で収集されていても、データが欠落していたり不正確であったりすることがあり、[情報の偏りの原因](#)となることがある。さらに、データの記録方法や質には施設間や個人間でばらつきがあり、異なるデータ源の質管理プロセスにもばらつきがある可能性がある。

関連する研究要素に関するデータの入手可能性に加え、あるデータ源の研究課題への関連性は、いくつかの要因に影響される。これには、研究サンプルの代表性、治療パターンや医療提供の NHS における日常診療との類似性、データの適時性、サンプルサイズ、追跡期間などが含まれる。重要な疑問点は、データが意思決定問題に関連する頑健な推定値を出すのに十分かどうか、そして結果が NHS の対象集団に翻訳または一般化されると期待できるかどうかである。

詳細については、[データ適合性の評価に関するセクション](#)を参照のこと。

### 2-5-3. バイアスのリスク

リアルワールドデータを用いた研究は、事例によって様々なバイアスリスクにさらされる。個々のリアルワールドエビデンス研究における内部妥当性を脅かす主なバイアスのリスクについて、以下に記述する。[非ランダム化比較試験におけるバイアスのリスク](#)については、[薬剤疫学とファーマコビジランスのための ヨーロッパセンターネットワーク \(ENCEPP\) の薬剤疫学における方法論基準に関する手引きや、コクランハンドブックの介入に関する系統的レビューの第 25 章](#)など、詳細な記述がある。

#### 2-5-3-1. 選択バイアス

非比較研究では、調査対象者が対象集団を代表していない場合、選択バイアスが生じる可能性がある。これは、調査対象集団のランダム抽出でないサンプリング、質問票への無回答、調査研究に志願した人々の行動や結果の違いから生じる可能性がある。

比較効果研究において、参加者の選択または追跡期間が介入とアウトカムの両方に関連している場合、選択バイアスが生じる。対象集団の代表性の欠如自体は、必ずしも比較研究における選択バイアスの原因ではない。比較研究における選択バイアスは[交絡とは異なる](#)。

試験開始時の選択バイアスの一般的な原因には、以下のようなものがある：

- 非ユーザーと比較した技術の有病者ユーザーを含む（すでにイベントを経験したユーザーや介入に耐えられなかったユーザーは分析から除外される）。
- アウトカムが起こりえない追跡期間を除外する（生存アウトカムの不死時間バイアスとして知られる）
- 介入とアウトカムに関連する特性（例えば、入院など）に基づく研究への選択。

試験終了時の選択バイアスの一般的な原因は、追跡不能である。選択バイアスは、データの欠落がある参加者などを解析から除外することによっても起こりうる。

#### 2-5-3-2. 情報バイアス

情報バイアスは、集団の適格基準、介入または曝露、アウトカム、[共変量](#)（関連する場合）についてのデータの欠落または不正確さから生じる可能性がある。このような限界

は、データの質、ケアパターン、データ収集プロセスが低いために生じる可能性がある。また、追跡期間の誤設定に起因することもある。

これらの問題の結果は、研究の種類、介入群間で制限が異なるかどうか、ランダムか系統的か（つまり、欠損データのメカニズム）、制限の大きさ、どの変数で発生するかなどの要因によって異なる。群間で誤分類が異なる一般的な原因の1つは検出バイアスである。これは、介入の有無によってケアの過程が異なり、ある群では別の群よりもアウトカムが同定されやすい場合に発生する。さらなる情報については、[測定誤差と誤分類のセクション](#)を参照のこと。

### 2-5-3-3. 交絡

[交絡](#)は、介入の選択と結果に共通の原因がある場合に生じる。医療従事者と患者は、ベネフィットとリスクに対する期待に基づいて治療の開始と継続に関する決定を行うため、これは医療において一般的であると予想される（適応による交絡または channelling bias として知られている）。交絡バイアスは、適応の異なる治療や介入の種類（例えば、薬物治療と比較した介入手技など）を比較する場合や、環境暴露の研究において難航する可能性がある。

例えば、因果経路上の共変量（降圧剤が脳卒中に及ぼす影響における血圧など）、共変量（暴露と結果の両方から独立して影響を受ける変数）、または instrument（暴露と関連するが、暴露を通して以外は結果と無関係な変数と定義される）を対照とする研究の場合、共変量の調整が不適切なためにバイアスが生じることもある。

### 2-5-3-4. その他のバイアス

逆因果（または原病性バイアス）は、介入が結果の結果または結果の症状である場合に起こる。これは、いくつかの癌のように潜伏期間が長い状態で最も問題となる。これが存在する場合、内的妥当性に重大な影響を及ぼす深刻なバイアスとなる。

バイアスは、データの統計的分析（例えば、モデルの誤特定）から生じることもある。

ある研究課題に関する文献群を評価する際、リアルワールドのエビデンス研究が報告されないため、[出版バイアス](#)がさらに懸念される。

### 3. 定量的なリアルワールドエビデンス研究の実施

- [はじめに](#)
- [学習計画](#)
- [研究実施](#)
- [研究報告](#)

#### 3-1. キーメッセージ

- エビデンスの信頼性を向上させ、査読者が研究を批判的に評価できるようにするためには、リアルワールドのエビデンスを透明で再現可能な形で作成することが不可欠である。
- 以下の原則は、リアルワールドでのエビデンス研究の実施を支えるものである：
  - データの[出所](#)が明確で、関連性があり、研究課題に答えるのに十分な品質であることを確認する。
  - 研究の計画から実施、報告に至るまで、透明性のある方法で、誠実にエビデンスを作成する。
  - バイアスのリスクを最小化し、不確実性を特徴付ける分析方法を使用する。
- 必要なエビデンスレベルは、用途や様々な背景要因によって異なる（[リアルワールドにおけるエビデンスの質と受容性に関する考慮事項の項](#)を参照）。勧告がどのようになされるかについての詳細は、NICE の関連マニュアルを参照されたい。

#### 3-2. はじめに

##### 3-2-1. エビデンス生成の原則

本節では、実社会におけるエビデンス研究の計画、実施、報告に関する NICE の望ましいアプローチについて述べる。

以下の原則は、すべてのリアルワールドエビデンス研究の実施を支えるものである：

- データの[出所](#)が明確で、関連性があり、研究課題に答えるのに十分な品質であることを確認する。
- 研究の計画から実施、報告に至るまで、透明性のある方法で、誠実にエビデンスを作成する。
- バイアスのリスクを最小化し、不確実性を特徴付ける分析方法を使用する。

ここでの焦点は、現在のところ、量的データを用いたリアルワールドのエビデンス研究である。しかし、計画、実施、報告のいくつかの側面は質的研究にも適用可能である。異なる側面については、質的エビデンスの分析、統合、提示に関する一般的な方法を適用すべきである。

患者は、試験の計画および実施のすべての側面を通じて相談されるべきである。

### 3-2-2. リアルワールドのエビデンスの質と受容性に関する考察

すべての研究は、最高レベルの透明性と厳密性を目指すべきである。しかし、1つのガイダンスに反映させることができるリアルワールドエビデンス研究の数と種類が多いということは、分析・報告の程度と、以下のような使用状況との間に、合理的なトレードオフが存在する可能性があるということである：

- 最終勧告に対する研究の貢献
- 健康およびシステムの成果に対する勧告の影響
- その他の文脈的要因

特定の種類のエビデンスの貢献度は、不確実性の主な要因（すなわち、エビデンスギャップ）に応じて、用途によって異なる。例えば、腫瘍学では、全生存期間などの長期アウトカムに関する仮定や、NHS への国際共同試験の適用性が鍵となることが多い([Morrell et al. 2018](#))。費用対効果モデルや費用比較モデルでは、イベント発生率、有病率、疾患の自然史、検査性能、費用、QOL など、多くの異なるパラメータが費用対効果の重要な決定因子となりうる。

一般に、臨床効果に関する[非ランダム化比較試験には](#)、単純な特性評価試験よりも高いレベルの厳密性と透明性が必要とされる。臨床効果の推定は、通常、推奨の重要な原動力であり、非ランダム化比較試験はバイアスの危険性がある。

エビデンスの受容可能性に影響する文脈的要因には、意思決定の不確実性のレベル、疾患の有病率、健康格差への影響、質の高いエビデンスを生み出す可能性などがある。利用者は、勧告がどのようになされるかについて、関連する NICE マニュアルを参照すべきである ([NICE ガイダンスの項を参照](#))。

特定の状況では、質の高いリアルワールドエビデンスを作成することがより困難な場合がある。これには、希少疾患、一部の医療 instrument（デジタルヘルス技術を含む）、介入手技、その他の複雑な介入が含まれる。このような状況では、ランダム化比較試験の実施も困難な場合がある ([ランダム化比較試験の用途と課題の項を参照](#))。

リアルワールドデータを用いた医療 instrument や介入試験の評価における一般的な課題には、以下のようなものがある：

- 医療 instrument の使用とアウトカムに関する全国的な統合データ収集は限られている。
- 日常的に収集される多くのデータソースには、特定の instrument（および固有の instrument 識別子）または手技を特定するための粒度がない。
- 適切な比較対照の特定、時間の経過に伴う技術の変化、学習効果。

これらの課題は普遍的なものではなく、登録や電子カルテシステムなど、医療 instrument や手技に関する質の高いデータ収集の利用可能性は現在も改善されている。可能な限り、最高品質のデータを使用すべきである。

希少疾患における一般的な課題には、以下のようなものがある：

- 対象者の体系的な特定が不十分であること
- サンプルサイズが小さい、または異なる [データモデル](#) やデータ収集プロセスで複数のデータソースを組み合わせる必要がある。
- 合意された共通 [データ要素](#) の欠如
- 疾患の自然歴にはかなりのばらつきがある
- 複雑な治療経路

### 3-3. 研究計画

#### 3-3-1. リサーチ・クエスチョンの定義

[エビデンス開発者](#)は、研究デザインに関わらず、研究課題を明確に特定すべきである。リサーチクエスチョンの具体的な要素は様々であるが、以下は多くの研究デザインに共通するものである：

- 関連する場合、集団の適格基準、介入または曝露、アウトカム（患者またはシステムのアウトカム）、共変量（交絡因子および効果修飾因子を含む）を含む主要な研究変数の概念的定義。
- サブグループのカテゴリが検証されているか、または関連する研究分野で一般的に使用されているかどうかを明記する。
- 推定される目標量、例えば疾患有病率や全生存期間に対する介入遵守の平均効果など。

患者のアウトカムは、患者がどのように感じているか、どのように機能しているか、あるいはどの程度長生きしているかを反映するものでなければならない。これには、生存期間などの臨床的アウトカムのほか、患者報告アウトカムも含まれる。アウトカムは、使用される状況に対して信頼でき、妥当でなければならない。アウトカムの選択は、[Core Outcome Measures in Effectiveness Trials \(COMET\)](#) データベースに掲載されているような、質の高い中核的アウトカムセットによって支援される。

推計する対象量は、関心のある全体的な研究課題に対応するものでなければならない。例えば、有病率は、ある時点でサービスの利用を必要とする集団の数を反映することができる。有病率は、発症率および症状の持続期間の関数であり、公衆衛生計画に有用である。罹患率は、異なるサブグループまたは異なる曝露を受けた集団における発生率を把握するものであるが、定義された時間間隔にわたって一定の率を仮定している。したがって、平均発生率の妥当性を検討する必要がある。

比較効果に関する非ランダム化比較試験の場合、開発者は、ランダム化エビデンスがない理由、既存の試験の限界、研究課題に対する確実なリアルワールドエビデンスを作成する能力を考慮し、研究の明確な正当性を示すべきである。

### 3-3-2. 計画調査の実施

開発者は可能な限り研究計画を事前に規定することを目指すべきである。[プロトコルには](#)、研究の目的、データの特定や収集、[データ管理](#)、研究デザイン、サブグループ解析や感度解析を含む事前に計画されたすべての解析方法を記述すべきである。多くのリアルワールド試験におけるデータ収集の複雑さは、全ての分析上の決定を事前に規定できないことを意味する。データによって決定される場合は、明確に記述され、計画されたアプローチが正当化されるべきである。[HARmonized Protocol Template to Enhance Reproducibility \(HARPER\)](#) ツールは、透明で再現性のある比較効果のリアルワールド試験を支援するためのプロトコル構造を提供する。

実施前に研究を計画することで、研究の質が向上し、開発者が複数の分析を実施し、最も好ましい結果をもたらすものを選択するリスクを減らすことができる。

解析計画を事前に特定することは、比較効果の研究で特に重要である。このような研究では、プロトコルの変更を登録し、正当性を証明した上で、一般にアクセス可能なプラットフォームで研究プロトコルを公表することを推奨する。特定のプラットフォームを推奨するものではないが、[ClinicalTrials.gov データベース](#)、[ISRCTN \(International Standard Randomised Controlled Trial Number\) レジストリ](#)、[EU-PAS \(European](#)

[Union electronic Register of Post-Authorisation Studies](#))、[OSF \(Open Science Framework\)](#)などが選択肢として挙げられる。これらのデータベースの中には、現在のところ、リアルワールドのエビデンス研究に適しているものもある。

研究プロトコルの登録に関するさらなるガイダンスは、[Real-World Evidence Transparency Initiative](#)が提供している。[NICEの科学的助言](#)チームは、技術開発者がエビデンス創出計画の一環としてリアルワールドデータをどのように最大限に活用できるかについて助言している。

調査を計画する際、開発者は、調査の設計、分析、解釈において取り組むべき平等または多様性の問題を検討すべきである。

### 3-3-3. 目的に合ったデータの選択

開発者は、最終的なデータソースの選択を正当化し、データの出所が確かで、研究課題の目的に合っていることを確認する必要がある ([データの適合性の評価に関するセクション](#)を参照)。

開発者には、以下のような体系的で透明性のある再現可能な検索を通じて、候補となるデータソースを特定することを奨励する：

- データセットの選択と優先順位付けのための検索戦略と定義された基準の事前指定。
- 検索戦略や選択基準に情報を与え、適切な既知のデータセットを強調するための専門家によるコンサルテーション。
- オンライン検索と体系的な文献検索、および必要に応じて関連出版物の筆頭著者と連絡を取り、潜在的なデータソースへのアクセスや適切性に関する情報を得る。
- データソースを直接検索する。英国では [Health Data Research UK Innovation Gateway](#) のようなデータソースの登録によってサポートされている。
- 特定された潜在的データソースの総数、除外されたデータソースの数、およびその理由 (データの適合性やアクセス可能性が低いという理由を含む) を概説するフロー図。

このアプローチは、[データ適合性の評価に関するセクション](#)で概説した検討事項や、外部ガイダンス ([Hall et al .](#))

データソースを特定するための努力は、研究全体の重要性に比例すべきである。また、現在、データ源の登録は必ずしも利用可能でなかったり、メタデータが限定的であることも認識している。

データは現地の法律、[ガバナンスの](#)取り決め、実践規範、[データ管理者の](#)要求事項に従ってアクセス・利用されるべきである。英国では、Health Research Authority (HRA) が「[UK Policy Framework for Health and Social Care Research \(医療・社会ケア研究のための英国政策枠組み\)](#)」に従い、研究およびデータ使用に関するガイダンスを提供している。

データ管理者やデータ処理者と早めに連絡を取り、必要な時にデータが利用できるようにすることが賢明である。開発者は、必要であれば、研究調査に対する適切な倫理的（またはその他の）承認を確保すべきである。開発者はまた、適切な場合には、独立した研究者や NICE 共同研究センターとデータを共有するための計画を立てるべきである。

#### 3-3-4. データ収集

事例によっては、[一次データ](#)収集が必要な場合もある。例としては以下のようなものがある：

- 新しい観察コホート研究
- 既存のデータソースを補完するための追加データ収集、例えば患者登録に QOL アンケートを追加したり、サブサンプル検証研究を実施する。
- 健康調査。

一次データ収集を計画する際には、患者と医療従事者の負担を最小限に抑えながら、患者中心の方法で収集を実施する方法を検討する。追加データ収集の実行可能性を評価してから進める。

データ収集は事前に定義されたプロトコルに従うべきであり、データ収集の完全性と一貫性を確保するために質保証プロセスを導入すべきである。これは、後方視的カルテレビューにおける構造化情報の抽出や、既に収集されたデータソースから[構造化データ](#)要素を導き出すためにデータサイエンスの手法を用いる場合にも適用される。

データ収集は、オープンデータ標準を用いた 'Findable, Accessible, Interoperable, and Reusable (FAIR)' データのベストプラクティス標準に従うべきである ([UK Health Data Research Alliance のデータ標準白書 2021](#) を参照)。

データは、適切な透明性情報を提供し、保護措置を講じた上で、現行のデータ保護法に従って収集、保管、処理、削除されるべきである。適宜、HRA または現地組織の審査・同意による承認が行われるべきである。適切な場合には、参加者の同意を得るべきである。

医療データの研究および非研究利用におけるガバナンス要件およびデータ保護規制に関する [Health Research Authority のガイダンス](#) を参照のこと。

### 3-4. 研究の実施

#### 3-4-1. 研究デザインと分析方法の選択

リアルワールドデータは、疾病の有病率や発生率、医療利用や医療費、治療経路、患者の特性、アウトカム、経験など、いくつかのタイプのエビデンスを作成するために利用することができる ([リアルワールドデータの事例の項](#) を参照)。使用する適切な研究デザインと分析方法は、研究課題に関連し、以下のようなデータの特徴を反映したものでなければならない：

- 結果変数の性質と分布
- サンプルサイズ
- データの階層またはクラスタリングを含むデータの構造（例えば、患者は病院内にクラスタ化される、またはデータは複数のタイムポイントで患者について収集される）。
- 集団間のアウトカムの不均一性
- データが横断的か縦断的かを問わない。

適切な場合は、選択した統計モデルの適切性を評価するために診断チェックを行う。適切なチェックは、研究の目的や使用方法によって異なる。

効果比較試験のデザインと分析に関する詳しい情報は、[方法のセクション](#) に記載されている。

#### 3-4-2. バイアスリスクの最小化

バイアスの原因による内部妥当性への脅威を特定し、適切なデータ収集と分析を通じて対処する。内部妥当性に対する主な脅威は、選択、情報、交絡、その他事例に応じたバイアスに由来する ([バイアスのリスクに関するセクション](#) を参照)。

特定のデータソースを使用することによるバイアスのリスクは、[データ適合性評価](#)の際に考慮された情報によって知らされる。

比較効果の研究におけるバイアスの最小化に関するより詳細なガイダンスは、[方法のセクション](#)に記載されている。

### 3-4-3. 研究結果の頑健性の評価

開発者は、研究計画と解析の段階でバイアスを最小化するよう努めるべきである。しかし、起こりうるバイアスの範囲や、リアルワールドのデータ源や分析方法の複雑さから、残存バイアスに関する懸念が残ることも多い。

[感度分析](#)は、バイアスのリスクが最も懸念される分野、または顕著な不確実性があるにもかかわらずデータキュレーションや解析の決定が行われた場合を反映すべきである。一般的な検討事項は以下の通りである：

- 主要な調査変数の運用上の定義の違い
- 研究変数とフォローアップを定義するための異なるタイムウィンドウ
- 別の患者適格性基準を用いる
- 欠損データと測定誤差への対処
- 代替モデルの仕様
- 治療の切り替えや追跡調査不能への対応
- 非アドヒアランスの調整

残存バイアスに関する懸念が依然として高く、推奨を行う能力に影響を及ぼす場合、開発者は定量的バイアス解析の使用を検討することができる。これらの方法は、研究結果に対するバイアスの影響を定量的に推定するものである [\(Lash et al.\)](#) バイアスに関する外部データを取り入れる場合は、透明性のある体系的な方法で特定する必要がある。相対効果を含む[経済](#)モデルのパラメータについては、感度分析により、パラメータ値だけでなく、費用対効果に対するバイアスの影響も考慮することができる。

### 3-4-4. 適切な品質保証プロセスの使用

データ管理、解析コード、解析の質保証は、試験の完全性を確保し、コーディングエラーのリスクを軽減するために不可欠である。品質保証プロセスは試験のリスクに比例すべきである。

品質保証の詳細については、[Office for National Statistic Quality Assurance of Code for Analysis and Research](#) および[英国政府の Aqua Book](#) を参照のこと。これは、検証済みの分析プラットフォームを使用することでサポートされる場合がある。

### 3-5. 研究報告

研究の報告は、データにアクセスできる独立した研究者が研究を再現し、結果を解釈し、その長所と限界を十分に理解できるような十分なものでなければならない。いくつかの報告チェックリストには、以下のような重要な報告項目が示されている：

- 観察研究（研究デザイン別の報告チェックリストについては [EQUATOR ネットワーク](#)、[疫学における観察研究の報告の強化 \(Strengthening the Reporting of Observational Studies in Epidemiology \[STROBE\]\)](#) ガイドライン参照
- 日常的に収集されたデータを用いた観察研究 ([REporting of studies Conducted using Observational Routinely collected Data \[RECORD\]](#))、および
- 比較効果の研究 ([RECORD statement for pharmacoepidemiology \[RECORD-PE\]](#)、このツールは当初、薬剤疫学研究のために設計されたが、項目は他の比較研究にも関連する)。

また、[START-RWE ツール](#)は、事例を横断して研究データ、方法、結果のプレゼンテーションを支援するために開発された。

以下では、データソース、データキュレーション、方法、結果など、この研究を報告する上で特に重要な事柄について述べる。

#### 3-5-1. データ・ソースに関する報告

データ源、その出所、研究課題との関連における質と関連性を理解するために十分な情報を提供する。これは[データ適合性評価](#)で記述された考慮事項に基づくものでなければならない。

開発者は追加情報を提供すべきである：

- 研究に対する倫理的（またはその他の）承認、またはそのような承認が不要であった理由を説明すること。
- データが承認および情報[ガバナンス](#)の要件に従ってアクセスされ、使用されたことを示す声明。

- 他の人がどのようにデータにアクセスできるかの説明（すなわち、データ[共有](#)説明。）

### 3-5-2. データキュレーションと分析の報告

多くのリアルワールドのエビデンス研究、特に日常的に収集されたデータを使用する研究は、分析を行う前にかかなりの処理（またはキュレーション）を必要とする。データ管理における決定（連結、変換、除外を含む）は、研究結果に大きな影響を及ぼす可能性がある。データキュレーションは、査読者が何が行われ、それがどのように結果に影響するかを理解できるように、十分に説明されなければならない。これには可能な限り、エビデンス開発者がデータにアクセスする前に行われたキュレーションも含めるべきである。

個々の研究ごとに、開発者はバージョンシステムや使用した外部パッケージなど、解析に使用したソフトウェアに関する情報を提供すべきである。理想的には、解析コードはコード構造、書式、コメントにおいてベストプラクティスに従い、（例えば GitHub のようなコードリポジトリを通じて）一般に公開されているか、複製が可能なように要求に応じて入手できるようにすべきである。[構造化されていないデータから](#)変数を構築するために、人間による抽象化または人工知能ツールが使用されている場合は、使用された方法とプロセスが明確に記述され、その妥当性が文書化されるべきである。

例えば、プロプライエタリなソフトウェアや特定可能な個人情報を使用する場合など、すべての状況において完全にオープンなコードを提供することは不可能かもしれません。開発者は、使用した手法とその妥当性に関する明確な情報を提供すべきである。また、必要な知的財産権を保護した上で、分析の再現と検証に必要なアルゴリズムへのアクセスを、要求に応じて提供するよう努めるべきである。

例えば、可能であれば解析の監査証跡を示すなど、研究が適切に行われた証拠を提供することで、研究実施の誠実さに対する信頼はさらに向上する。これは例えば、開発者が関連する結果が明らかになる前に解析の準備をし、プロトコルを確定していたことを示すことができる（[MacCoun and Perlmutter 2015](#)）。

### 3-5-3. 方法に関する報告

以下に、報告すべき主な項目を記す。この情報は、サブグループ解析や感度解析を含むすべての解析について提示されるべきである。方法は試験プロトコルと一致しているべきであり、逸脱は特定され、正当化されるべきである。

#### 3-5-3-1. 研究デザイン

すべての研究変数について明確な操作上の定義を示し、関連する場合はフォローアップの詳細を示すべきである。研究変数には通常、患者の適格基準、介入または曝露、アウトカム、共変量が含まれる。

各変数について、以下の情報を提供すること：

- 可能であれば、[コードリストと](#)アルゴリズムを含む変数の運用定義
  - コードリストやアルゴリズムがどのように開発され、可能であれば検証されたか。
- 各変数の情報を求める期間を、指標日（例えば、治療開始の12ヶ月前）に関連して定義する。
- 関連する場合は、介入の継続的な使用を示すと想定されるオブザベーション間の猶予期間。

比較効果の研究では、潜在的な交絡因子が同定された過程を、研究変数間の因果関係に関する仮定とともに記述すべきである。

フォローアップに関する以下の情報は、該当する場合に記載する：

- 指標日に関連したフォローアップの開始と終了。
- 介入については、介入と結果発生間の最短時間（潜伏期間）および効果の持続期間（曝露-影響窓）に関する仮定が必要である。

[縦断](#)研究では、この情報を研究デザイン図を用いて要約することが有用である

[\(Schneeweiss et al.\) 再現可能なエビデンス](#) : Reproducible Evidence: Practices to Enhance and Achieve Transparency (REPEAT) イニシアチブのプロジェクトページでは、論文とデザインダイアグラムのテンプレートを公開している。

### 3-5-3-2. 統計的手法

使用した統計的手法は明確に記述されるべきである。情報は以下のために十分でなければならない：

- どのような方法が使われ、なぜそれが選ばれたのかを理解する
- モデリングの仮定の妥当性を示す
- 解析が、選択バイアス、情報バイアス、関連する場合は交絡を含む様々なバイアスのリスクにどのように対処しているかを理解する ([質の評価に関するセクションも参照](#))。

### 3-5-4. 結果報告

すべての研究において、以下の情報を提示すべきである：

- フロー（または患者減少）図：生データから最終分析標本までの各段階における患者数を、除外理由とともに報告する。
- 患者特性（欠測データを含む）、イベント発生率（またはアウトカムに関するその他の分布情報）を含む追跡調査の詳細。比較研究の場合は、群間または曝露レベル間、また関連する場合は調整前と調整後で提示する。
- 分析サンプルと対象集団の患者特性の違い。

結果には、中心点推定値、精度の尺度、その他必要であれば関連する分布情報を含める。主解析とすべてのサブグループ解析、感度解析の結果を提示する。これらの解析のうち、どれが事前に指定され、どれが指定されなかったかを明らかにする。交絡に対処するために調整を用いた解析については、未調整の結果も提示する。

図表中の情報が不用意に患者を特定できないようにする。[国家統計局（Office for National Statistics）](#)には、[保健統計を広める際の機密保持に関するガイダンスがある](#)。

### 3-5-5. 結果の解釈

結果の意味を解釈するのに役立つ情報を提供する。データソース、研究デザイン、分析の限界について説明する。

### 3-5-6. リアルワールドのエビデンス研究を明確に伝える

リアルワールドのエビデンス研究は技術的に複雑な場合がある。読者の理解を助けるために、研究は以下の方法で明確に文書化されるべきである：

- 理解しやすい科学的資料を書くためのアドバイスに従うこと（[Gopen and Swann 1990](#), [Greene 2013](#) 参照）。
- 専門用語を避ける。それが不可能な場合は、用語を平易な英語で説明する。
- 略語を避ける（[Narod et al.](#)）
- 表、グラフ、その他のテキスト以外のコンテンツを明確に表示し、その解釈方法を説明する。

## 4. データ適合性の評価

- [はじめに](#)
- [データの実証性](#)
- [データの目的適合性](#)
- [データの質](#)
- [データの関連性](#)

### 4-1. キーメッセージ

- データ源の信頼性を確保し、研究課題に取り組む目的への適合性を理解するためには、データ源の透明性のある報告が不可欠である。
- データは[出所](#)が明確であること
  - データソースに関する報告は、データの特徴、データ収集、カバー率、[ガバナンス](#)をカバーすべきである。
- 目的に対するデータの適合性は、データの質と関連性によって要約できる。
  - [データの質とは](#)、主要な研究変数の完全性と正確性に関するものである。
  - [データの妥当性は](#)、データの内容、データと NHS の対象集団との患者、介入、ケア設定の違い、サンプルサイズや追跡期間などのデータの特徴によって決定される。
- [付録 1 のデータ適合性評価ツール \(DataSAT\)](#) は、データ適合性に関する一貫した構造化された情報を提供するために使用することができる。
- 異なるデータソース間には、質、サイズ、臨床の詳細さ、地域性という点で合理的なトレードオフがある。
- あるデータ・ソースが受け入れられるかどうかは、用途や様々な[状況要因](#)に左右される。

### 4-2. はじめに

NICE ガイダンスに使用されるデータは、透明性をもって報告され、[出所](#)が明確で、研究課題との関連において目的に適合していなければならない。枠組みのこのセクションの主な目的は以下の通りである：

- データに関する明確で透明性のある報告とその目的適合性への期待について、[エビデンス開発者](#)に明確なガイダンスを提供する。
- エビデンスのレビュー担当者や委員会が、研究を批判的に評価したり推奨を作成したりする際に、データの信頼性や適切性を理解できるようにする。

このセクションは、[定量的実地エビデンス研究の実施に関するセクション](#)と並行して読む必要がある。

データ保護と情報[ガバナンスに関する](#)国内法・規制に従ってデータを使用すべきであるという以外に、データ適合性の最低基準は定めていない（[データソースに関する報告の項](#)を参照）。データ適性に関する考慮事項は、様々な種類のリアルワールドのデータと使用事例に広く適用可能であるが、主に定量的研究に焦点を当てている。

データソースの受け入れ可能性は、事例と文脈的要因に依存する（[リアルワールドエビデンス研究の質と受け入れ可能性に関する考慮事項のセクション](#)を参照）。我々は、質、規模、臨床の詳細性、地域性など、データ源の様々な特性間のトレードオフの必要性を認識している。十分な国内データがない場合、あるいは結果が設定された地域間でうまく翻訳されることが期待される場合には、国際的なデータが適切な場合もある。また、希少疾患や一部の医療 instrument・介入手技など、一部の用途において、最高品質のエビデンスを特定・収集することに課題があることも認識している（「[リアルワールドにおけるエビデンスの創出における課題](#)」の項を参照）。

この情報の全体的な提示については、特定の形式を要求するものではない。しかし、我々はデータ適合性評価ツール（DataSAT）を開発し、評価時点におけるデータ適合性の一貫した構造的な提示を支援する。このツールで提示される概念は、開発者がデータソースの候補を選択する際や、フィージビリティ・スタディを実施する際にも役立つが、これは第一の目的ではない。本ツールのテンプレートと適用例は[付録1](#)に示す。

### 4-3. データの実証性

[データの](#)信頼性を高め、特定のアプリケーションの目的適合性を理解するためには、データ・プルバンスを完全に理解することが不可欠である。このセクションでは、データ・ソースの基本的特性、データ収集、カバレッジ、ガバナンスという4つのテーマにわたって、データ・プロテナンスに関する考察を紹介する。

リアルワールドのエビデンス研究の多くは、[データ連結](#)や[データプールによって](#)、複数のデータ源を組み合わせる。データ連結は、例えば前向き観察コホート研究のデータと退院記録や死亡記録、あるいは患者が作成した健康データを組み合わせることによって、個々の患者について利用可能な情報を拡大するために行われることが多い。データプーリングは、サンプルサイズやデータのカバー範囲を拡大するために用いられ、希少疾患の研究で一般的である。

データソースの報告は、主に調査研究に使用された複合データに言及すべきである。ただし、貢献したデータセット間の重要な相違点は明確に記述する。

#### 4-3-1. データの基本特性

データ源の特定を可能にする情報を明確に報告すること。これには、全体的なデータ源と貢献したデータ源の名称、バージョン（もしあれば）、データ抽出の日付が含まれる。

共通のデータ・モデルは、異なるデータ・ソースの構造や、時にはコーディング・システムを標準化するために使用される。データが共通のデータモデルに変換された場合は、そのモデルとバージョンを報告し、情報の喪失を含め、マッピングの全詳細を利用できるようにすべきである。この情報は研究を再現するために不可欠である。

共通のデータモデルは、連携データネットワークの使用もサポートする。これにより、個々の患者の健康データは、データセット間で結果が集約される前に標準化された分析を実行する提携データ保有者の保護下に置かれる。連携データネットワークの報告は、データパートナー募集のプロセス、実現可能性評価、使用された共通の分析枠組みを理解するのに十分でなければならない。

完全に正確なデータ連結はデータの質と価値を向上させるが、不完全な連結は患者記録を除外したり、データの誤分類につながる可能性がある。従って、複数のデータソースをリンクする場合は、以下の情報を報告する：

- 誰がリンケージを行ったか（例えば、NHS デジタル）
- 決定論的か確率論的かを含む連結の方法と、連結に使用される変数
- データリンケージの性能特性（[データリンケージの品質評価に関する政府分析機能ガイドンス](#)を参照）。

#### 4-3-2. データ収集

データソースを理解するには、データ収集の目的と方法についての知識が必要である。

データ収集の本来の目的に関する情報は、以下を含むべきである：

- データが日常的に収集されたものか、特定の研究目的のために収集されたものか（またはその組み合わせ）。
- データソースの種類や主な用途など：
  - 患者ケアのための電子カルテ

- プロバイダーへの償還のための管理データ
- 医療 instrument 安全性評価レジストリ
- 介入後の QOL を推定する前向き観察コホート研究
- ある疾患の自然史をモデル化するための後方視的カルテレビュー。

重要なデータ型に関する追加情報を網羅すべきである：

- 臨床診断、検査、処置、処方など、どのような種類のデータが収集されたか。
- 例えば、臨床診断に ICD-10 コードを用いたり、がんの病期やバイオマーカーに関するフリーテキストデータを用いたりする。
- データがどのように収集されたのか、例えば、臨床検査において医療従事者が直接、遠隔モニタリングによって、あるいは管理スタッフによって収集されたのか。デジタルヘルス技術によってデータを収集した場合は、その技術の妥当性を報告する。
- 例えば、時間の経過に伴うデータ収集の変化：
  - 新しいデータ要素の追加（例えば、QOL アンケートなど）
  - データ要素の削除
  - データ収集方法の変更（例えば、患者のアウトカムの日常的モニタリングへの切り替えなど）
  - コーディングシステムの変更（例えば、英国のプライマリーケアにおける Read v2 から SNOMED-CT コードへの切り替え）。コーディングシステム間のマッピングに関する情報を入手できるようにする。
  - データ収集に大きな影響を与えたデジタルヘルス技術を含むデータ収集システムのソフトウェア更新。
- 実施されていたデータ収集の質保証プロセス（研修または盲検レビューを含む）
- 共通データモデルやその他のデータ標準への変換など、データに対して行われる変換。

データ提供者間で、どのように、どのようなデータが収集されたか、またその質について違いがある場合は、それを説明する必要がある。これは、データソースが異なるシステムや国間でプールされている場合に特に重要である。

#### 4-3-3. データ・カバレッジ

データの対象範囲に関する明確な情報を提供することは、人口、ケア設定、地理、時間などを含めて不可欠である。このような情報は、[データの妥当性](#)を評価する上で重要な意味を持つ。

情報を提供すべきである：

- データソースが対象集団をどの程度捉えているか：
  - データソースに対象人口が完全に含まれていない場合は、取得したデータの代表性に留意すべきである。
  - 患者登録を含む前向きデータ収集を伴う研究では、患者登録に関する情報を報告すべきである。
- データ収集のベースとなった環境：
  - この場合、ケア環境（例えば、プライマリーケアとセカンダリーケア）、医療提供者のタイプ（例えば、専門医療センターと一般病院）、および関連するその他の要素を区別する必要がある。
  - 情報が医療や社会的ケアのシステム外で収集された場合は、その旨を記述すること（例えば、日常生活動作の遠隔モニタリング）。
- 該当する場合は、国や地域を含むデータの地理的範囲。
- データ収集の期間。

#### 4-3-4. データガバナンス

データガバナンスに関する情報は、データの成熟度とその信頼性を理解する上で重要である。これには以下の情報が含まれるべきである：

- [データ管理者の氏名](#)
- データ収集と維持のための資金源
- [データディクショナリ](#)や [データモデルなどのデータドキュメント](#)
- 監査を含む品質保証およびデータ管理プロセスの詳細。

#### 4-4. データの目的適合性

[データの出所に関するセクション](#)では、計画された研究とは異なるデータソースの重要な特徴について説明した。このセクションでは、データの質と関連性を考慮し、特定の研究課題に答えるためのデータの目的適合性に焦点を当てる。あるデータセットはある用途には価値があるが、別の用途には価値がない場合がある。

[データ洗浄](#)、除外、変換を含む実質的な[データ管理](#)は、解析のための原データソースを準備するために必要である。データ管理及び品質保証は、[研究報告のセクション](#)に記述されているように、透明性をもって報告されなければならない。

#### 4-5. データの質

データの質に対する制限には、データの欠落、測定エラー、分類の誤り、日付の不正確な報告などがある。これらの問題は、患者の適格基準、アウトカム、介入または曝露、共変量など、すべての研究変数に適用される可能性がある。これらの問題は、リアルワールドのエビデンス研究が偏った推定値を出す原因となる情報バイアスを生み出す可能性がある。データの質に関する透明性のある報告は、査読者がバイアスのリスクを理解し、それがデータ解析によって適切に対処されているか、感度解析によって検討されているかを理解するために不可欠である。我々は、データの質の2つの主要な側面、すなわち完全性と正確性に焦点を当てている。

すべての主要な研究変数について、完全性と正確性に関する情報を提供すること。研究変数は、[構造化データ](#)および[非構造化データ](#)の両方を含む複数のデータ要素を組み合わせることで構築することができ、異なるリンクされたデータソースに由来することもある。これらの研究変数の複雑さは、データソースや用途によって異なる。例えば、喘息増悪は一つのデータフィールド（質問票への回答など）から特定できる場合もあれば、診断コード、処方、検査、フリーテキスト、その他のデータの組み合わせから構築する必要がある場合もある。

[研究報告](#)の項で述べたように、データの種類、[コードリスト](#)、非構造化データからの抽出、可能であれば期間など、各研究変数について明確で曖昧さのない定義を示すことが不可欠である。コードリストを含むこれらの運用上の定義は、他者が利用できるようにし、適切であれば再利用する。既存のコードリストの有効性は、使用前に確認されるべきである。非構造化データを使用する場合は、データ抽出とその方法の信頼性に関する情報を提供する。

これらの考慮事項は、患者報告アウトカムやデジタルバイオマーカーなど、患者が生成したデータを生成するデジタルヘルス技術からのデータにも適用される。技術から生成されるデータの妥当性と利用者のアクセシビリティに関する更なる情報を提供すべきである。

研究結果を解釈するためには、データの欠落や不正確さの理由、それらがランダムなのか組織的なのかについて、さらなる情報が必要である。比較研究の場合、欠測や不正確さが介入群間でどの程度異なるかを理解することが重要である。[情報の偏りへの対処の項](#)には、データの欠損、測定誤差、分類の誤りに対処する方法についての詳細情報がある。データの完全性や正確性については、許容できる水準が用途によって異なるため、最小限の閾値は設定していない（[リアルワールドのエビデンス研究の質と受容性に関する考慮事項のセクション](#)を参照）。

#### 4-5-1. 完全性

データの完全性とは、ある時点におけるデータの欠落がない記録の割合を指す。データの正確性についての情報を提供するものではない。このパーセンテージはデータソースから簡単に計算できることが多く、関連するデータを除外したり、インプューテーションを行う前に計算すべきである。心筋梗塞の経験などのアウトカムについては、データの欠落の問題は分類の誤りと明確に区別する必要がある。二値変数の場合、イベントがない（発生した）ことは、データ精度の問題（偽陰性による誤分類）として要約するのが最も適切であろう。

#### 4-5-2. 精度

精度の測定、つまりデータがどれだけリアルワールドに近いかは、変数のタイプによって異なる。以下では、さまざまなタイプの変数について、精度の一般的な測定基準を説明する：

- 連続変数またはカウント変数（平均誤差、平均絶対誤差、平均二乗誤差）
- カテゴリー変数（感度、特異度、陽性的中率、陰性的中率などの診断精度指標；[Fox ら、2022 年](#)）
- time-to-event 変数（リアルワールドのイベント発生時刻と記録されたイベント発生時刻の差）。

データの正確さを測るためのゴールドスタンダードのアプローチには、以下のようなものがある：

- 確立されたゴールドスタンダードの情報源（例えば、英国国家統計局の死亡率記録）との比較
- 医療記録の見直し。

これらのアプローチは、分析対象集団のサブセットに対して行われることもあれば、同一または類似の集団とデータソースにおける過去の研究に基づいて行われることもある。

これらのゴールドスタンダードアプローチは、常に可能であるわけでも、実現可能であるわけでもない。おおよその精度を示すことができる他のアプローチには、以下のようなものがある：

- 例えば、追加コードを使用する、複数のコードを必要とする、異なるデータタイプを組み合わせるなどして、異なる変数定義を比較する。
- 標本の分布を母集団の分布や過去の研究と比較する
- 専門家の意見を参考に、データの信憑性を探る。

- 一貫性の確認（データソース間の記録における患者ステータスの一致）
- 適合性の評価（データ要素の記録がデータソースの仕様と一致しているかどうか）
- 永続性（データが長期にわたって一貫しているか）をチェックする。

主要な研究変数に関するデータの正確性に関する透明性のある報告：

- 平均値および信頼区間を含む、精度に関する定量的情報。追加的な分布情報も価値があるかもしれない。
- 精度を定量化するために使用した方法とプロセス（仮定を含む）を記述する。また、研究変数の定義、母集団、データ源、期間、その他関連する検討事項の違いを考慮する。

#### 4-6. データの関連性

データの目的適合性の第二の要素は、データの関連性である。データの関連性についての主な疑問点は以下の通りである：

- データが堅牢で適切な結果を出すのに十分な情報を提供している。
- この結果は、NHS の患者に一般化される可能性が高い。

データの関連性の評価は、[データの出所に関するセクション](#)で提供された情報に基づいて行う。

NICE は、NHS における現在のケアを反映した、英国の人口に直接関連するデータを好む。しかし、NHS で入手できる情報が限られている場合や、結果が環境間でうまく翻訳されることが期待できる場合には、国際的なデータの潜在的な価値を認識している。アプリケーションによっては、ローカルデータの使用と、データの質、最新性、臨床的詳細性、サンプルサイズ、フォローアップなどの他の重要な特性との間にトレードオフが存在する。国際的なデータは、英国で利用可能になる前に他国で利用可能であった介入や、希少疾患の場合、特に価値があると思われる。NHS 内の地域医療機関や専門医療機関のデータを使用する場合にも同様の配慮が必要である。

以下に、データの関連性の主要な側面について、データのコンテンツ、カバレッジ、特性を区別して説明する。

##### 4-6-1. データ内容

データ内容がリサーチクエスチョンにとって十分かどうかを理解するために、3つの重要な考慮事項がある：

- データソースは、関連する集団の適格基準、アウトカム、介入、[共変量の適切な定義](#)を可能にする十分なデータ要素を含んでいるか。
- データ要素は十分な粒度で収集されているか？
- 測定は適切な時点で行われているか？

データ要素が十分かどうかを理解するためには、まず対象となる概念を定義し、それがリアルワールドのデータを使ってどの程度代用できるかを判断することが有用である。データ不足が意味するものは、研究変数や事例によって異なる。研究課題に答えるために必要な主要評価項目は入手可能であるべきであり、評価をサポートするために十分客観的で詳細であるべきである。母集団、介入、アウトカムを適切に定義するための情報が不十分だと、研究結果の妥当性が制限される。交絡因子に関する情報が不十分だと、妥当な知見を得る能力が制限される。

必要なデータの粒度は、研究課題によって異なる。例えば、人工膝関節置換術がQOLに及ぼす影響を検討する場合、理学療法のみ、人工膝関節全置換術と部分置換術の比較、あるいは異なる埋め込み器具の比較に関心があるかもしれない。同様に、研究課題によっては脳卒中をアウトカムとすることが適切な場合もあれば、出血性脳卒中と虚血性脳卒中を分けて考える必要がある場合もある。

最後に、人工膝関節置換術が術後1年のQOLに与える影響に関心があるかもしれない。日常的に収集されるデータでは、このような情報の記録は厳密なプロトコルに従っておらず、測定値が欠落していたり、不規則な時点で測定されていたりする。

#### 4-6-2. データ・カバレッジ

NHSの患者に対する研究結果の一般化可能性は、以下のようないくつかの要因に左右される：

- 分析サンプルと対象集団の患者特性の類似性
- ケア経路と治療環境の類似性
- ケア経路（診断検査を含む）とアウトカムの経時的変化。

分析標本と対象集団の類似性は、疾患の有病率を推定するような記述的研究では特に重要である。比較研究では、介入効果が異なる特徴を持つ患者間で移行すると予想される場合、これはあまり重要でない可能性があり、内部妥当性の確保に重点を置くべきである。

サブグループ間で治療効果にかなりの異質性がある場合は、患者特性の類似性がより重要になる。相対尺度の効果推定値は、通常、絶対尺度の推定値よりもサブグループ間でよりよく移行する ([Roberts and Prieto-Merino 2014](#))。予後モデリングなどの他の用途では、重要な患者サブグループの十分な代表性を確保するために、非代表的サンプリングが好まれる場合がある。

分析サンプルと NHS で見られる治療経路やケア設定の違いが、結果の妥当性にどのような影響を与えるかを考慮する必要がある。これは国際的なデータを使用する場合に特に重要である。NHS であっても、データは国を代表しない特定の地域に関するものであったり、すべての医療機関ではなく専門医療機関に焦点を当てたものであったりする。最後に、診断検査を含むケア経路の変化や、アウトカムの背景傾向（死亡率など）は、NHS のものであっても、過去のデータの価値を制限することがある。NICE ガイダンスに使用するデータの妥当性を議論する際には、これらの問題を注意深く考慮し、報告する必要がある。

#### 4-6-3. データの特徴

データの妥当性の最後のカテゴリーは、分析サンプルのサイズとフォローアップの期間（および分布）に関するものである。サンプルサイズは、頑健な推定値を出すのに十分な大きさでなければならない。しかし、文脈によってはサンプルサイズが常に制限されることを認識する。追跡期間は、対象となるアウトカムが発生または発生するのに十分な期間でなければならない（医療費などのアウトカムの場合）。交絡因子に関する情報を提供し、介入の新規利用者を特定するためには、追跡開始前に利用可能なデータ量も重要であろう。データ収集から研究に利用できるようになるまでのタイムラグの少ないデータソースを使用することで、より長い追跡期間を分析に利用できるようになるかもしれない。

## 5. 比較効果に関するリアルワールド研究の方法

- [はじめに](#)
- [非ランダム化研究デザインの種類](#)
- [研究デザイン](#)
- [分析](#)
- [研究の頑健性の評価](#)
- [報告](#)
- [品質評価](#)

### 5-1. キーメッセージ

- [非ランダム化比較試験は](#)、ランダム化比較試験がない場合に比較効果に関するエビデンスを提供するため、または日常的な設定における介入の効果に関するより広範な質問に答えるために、試験エビデンスを補完するために使用することができる。
- ここに示した勧告は、[外部対照群](#)を形成するためのリアルワールドデータを用いたものを含め、主にコホート研究に焦点を当てたものである。
- [研究デザイン](#)
  - 望ましいランダム化比較試験を模倣した試験をデザインする ([ターゲットトリアルアプローチ](#))。
  - 患者の適格基準が満たされ、治療が割り当てられ、フォローアップが開始されるまでの間の差に起因する時間に関連したバイアスを避ける。
  - 外部統制を用いる研究では、主要な研究変数の利用可能性と運用上の定義、データ収集プロセス、患者の特徴、治療設定、ケア経路、期間など、データソース間の差異を最小限にするようデータを選択し、キュレーションし、研究の質と妥当性への影響を考慮する。
- [分析](#)
  - 系統的なアプローチにより潜在的交絡因子（時変交絡因子を含む）を同定し、因果関係の仮定を明確にする。
  - 観察された交絡因子と観察されていない交絡因子を考慮した交絡に対処する統計的方法を用いる。
  - 情報打ち切り、欠損データ、測定誤差によるバイアスの影響を考慮し、必要であれば適切に対処する。
  - 感度分析とバイアス分析を用いて、バイアスの主なリスクや不確実なデータキュレーションと分析の決定に対する結果の頑健性を評価する。
- [報告](#)
  - 非ランダム化エビデンスの必要性を正当化する。

- 最終的な解析を行う前に、試験プロトコルと[統計解析](#)計画を提供する。
- 独立した研究者が研究を再現し、何がなぜ行われたかを理解できるよう、研究を十分に詳細に報告する。
- バイアスのリスクと研究課題との関連性を評価する。
- 許容されるエビデンスの質は、用途や様々な文脈的要因に依存する可能性がある（[リアルワールドのエビデンスの質と許容可能性に関する検討事項の項を参照](#)）。

## 5-2. はじめに

先に、様々な事例において、[定量的なリアルワールドエビデンス研究](#)を頑健かつ透明性をもって実施するための原則を概説した。本節では、リアルワールドデータを用いた比較効果の研究実施に関するより詳細な推奨事項を示す。これには、一次または[二次データ](#)収集に基づく従来の観察研究と、[外部対照](#)を形成するためにリアルワールドデータを用いる試験が含まれる。純粋な介入研究（ランダム化か否かを問わない）や介入データのみを用いた外部対照研究に対する特別な考察は行っていない。ここでは量的研究に焦点を当てるが、質的エビデンスが介入の価値に対する理解を深める上で重要な役割を果たしうることを認識する。

ランダム化比較試験は、比較効果を推定するために望ましい研究デザインである。[ランダム化対照試験](#)がない場合、研究課題に直接関連しない場合、または質が低い場合、非ランダム化エビデンスが付加価値を生むことがある（[ランダム化対照試験の用途と課題のセクション](#)を参照）。また、非ランダム化エビデンスは臨床試験エビデンスを補完し、より広範な疑問に答えることができる（[リアルワールドデータを用いた介入効果の推定に関するセクション](#)を参照）。

比較効果に関するリアルワールドエビデンスがエビデンスベースを改善する可能性がある場合、研究が頑健で透明性のある方法を用いて行われることが不可欠である。私たちは、理想的に行われたであろうランダム化比較試験を模倣するようにリアルワールドエビデンス研究をデザインし（[研究デザインのセクション](#)を参照）、交絡と情報バイアスに対処するために適切な統計的方法を使用し（[解析のセクション](#)を参照）、感度分析とバイアス分析を用いて結果の頑健性を評価することを推奨する（[頑健性の評価のセクション](#)を参照）。このアプローチは、[リアルワールドデータを用いたコホート研究を計画・報告する際の重要な検討事項を図1に視覚的にまとめたものである](#)。

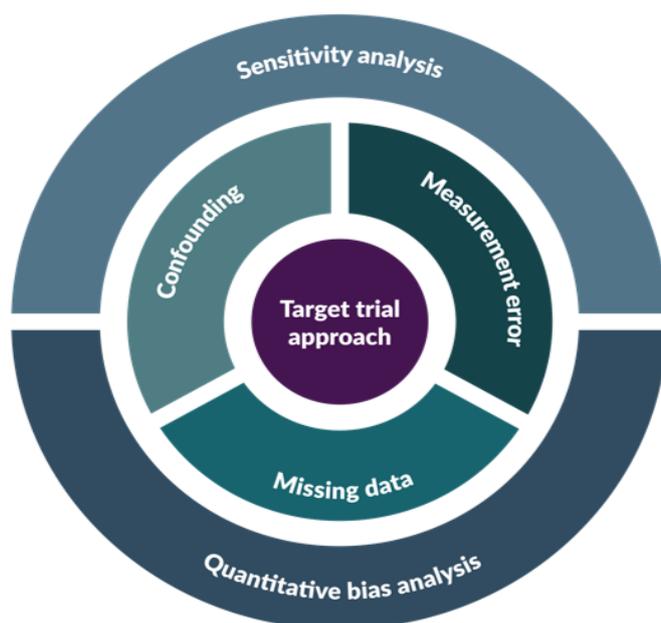
ここに提示された勧告は、方法論的な質と妥当性、および研究実施の透明性の両面から、比較効果に関するリアルワールド研究の質を向上させることを意図している。これらは、公表文献、国際的な研究コンソーシアム、国際的な規制・支払機関からのベストプラクテ

イスのガイダンスから導き出されたものであり、開発中の方法論に沿って定期的に更新される予定である。これらは [NICE 意思決定支援ユニットの技術支援文書 17](#) に基づくもので、観察データ解析のための統計的手法を提示している。

我々は、全ての研究が全ての推奨事項を完全に満たすことができるわけではないことを認識している。最も質の高い研究を実施できるかどうかは、適切なデータ（[データ適合性の評価の項を参照](#)）の入手可能性と、病態と介入の特徴に依存する。ユーザーエクスペリエンスやシステムのアウトカムのような非健康的アウトカムの評価など、他の用途ではより簡便な方法が適切な場合もある。さらに、特定の研究の受容性と意思決定への寄与は、いくつかの文脈的要因と同様に、用途に依存する（[リアルワールドのエビデンス研究の質と受容性に関する考慮事項のセクションを参照](#)）。

## 図 1

リアルワールドデータを用いたコホート研究を計画・報告する際に考慮すべき重要事項の視覚的要約



## 5-3. 非ランダム化研究デザインの種類

### 5-3-1. 概要

介入、曝露、政策の効果を推定するためには、多種多様な研究計画を用いることができる。望ましい研究デザインは文脈に依存する。曝露の変動が、個人内における経時的変動

なのか、個人間における変動なのか、医療提供者など他の集団間における変動なのかによって異なる。一般に、[riangulation](#)として知られるように、異なる研究デザインや分析方法を用いて結果が再現されれば、非ランダム化研究結果の信頼性は高まる ([Lawlor et al.](#))

一つの重要な区別は、介入研究と観察研究である。介入研究では、個人（または個人のグループ）はプロトコルに従って1つ以上の介入に割り付けられる。介入への割り付けは、ランダム、準ランダム、非ランダムがある。観察研究では、介入はプロトコルによって決定されるのではなく、医療・福祉専門家や患者の希望によって決定される。ハイブリッド研究では、介入データと観察データの両方を利用することができる。このセクションでは、観察研究とハイブリッド研究のみに焦点を当てる。

介入試験も観察試験も非対照試験とすることができる。非対照試験が適切なのは、疾患の自然経過がよく理解され、予測可能性が高く、治療効果が非常に大きいまれなケースに限られる ([ICH E10 臨床試験における対照群の選択](#)、[Deeks et al.2003](#) 参照)。ほとんどの場合、信頼性が高く有益な治療効果の推定値を得るためには、比較群が必要である。対照研究は、個人（または群）間、個人（または群）内の経時的、またはその両方における曝露と結果の変動を利用することができる。このセクションでは、対照研究に焦点を当てる。

以下では、比較研究の種類について述べる。一部の分類法では、前向き研究（一次データ収集を含む）と後向き研究（既に収集したデータに基づく）を区別している。この区別は、必ずしも研究の質に関する情報を伝えるものではないので、私たちはその使用を避けるよう助言する ([Dekkers and Groenwold 2020](#))。

### 5-3-2. コホート研究

コホート研究では、個体は曝露に基づいて同定され、追跡期間中にアウトカムが比較される。通常、コホート研究では、同じデータ源から得られた異なる曝露を受けた個人を比較する。しかし、介入データ源と観察データ源など異なるデータ源からのデータを組み合わせることもできる。この場合、観察データは試験で使用される介入に対する外部対照を形成するために使用される。試験は多くの場合、非対照[単群試験](#)であるが、対照試験の1群であることもある。外部データはランダム化比較試験内の同時対照を補強するためにも使用できる。

外部対照は、過去の臨床試験のデータから形成することもできる。このような試験の潜在的な利点は、患者の組み入れ基準、追跡調査、結果判定がより類似していることである。

多くの場合、過去の臨床試験からは個々の患者レベルのデータではなく集計データしか得られない。[NICE 意思決定支援ユニットの技術支援文書 18](#)では、集約されたデータを用いたアンカーのない間接比較の方法が記載されている。

以下の研究デザインと解析の項では、NICE ガイダンスに情報を提供する最も一般的な非ランダム化研究デザインである、リアルワールドデータソースからの外部対照を用いたものを含むコホート研究に焦点を当てる。準実験計画や自己対照研究を含むその他の研究計画は、以下に概説するように、文脈によっては関連する可能性がある。

### 5-3-3. 自己対照研究

自己対照計画、すなわち「被験者内」計画では、経時的な個人内の曝露状態の変動を利用する。これには、ケースクロスオーバー、自己対照ケースシリーズ、およびこれらのデザインの変種が含まれる。これらは、急性発症事象を伴う一過性の曝露に最も適している ([Hallas and Pottegard 2014](#))。主に医薬品（ワクチンを含む）の副作用の研究に用いられるが、前治療歴のある人の経験を用いて、がん治療薬の効果を評価するためにも用いられている ([Hatswell and Sullivan 2020](#))。これは、適切な標準治療の比較対照が入手できない場合に最も適切である。

自己管理法の主な利点は、遺伝的要因のように経時的に変化しない交絡因子（未測定または未知の交絡因子を含む）や、多くの健康行動のようにゆっくりと変化する交絡因子を管理できることである。しかし、経時的に変化する可能性のある共変量（例えば、疾患の重症度）を調整することは必要である。このような方法は一般的に、アウトカムに時間ベースのトレンドがないと仮定するか、トレンドを統計的にモデル化しようとするものである。これらのアプローチは、介入にさらされていない人々の対照群を加えることで強化できることが多い。

### 5-3-4. 横断研究

[横断的研究では](#)、現在の曝露と結果に関する情報が1つの時点で収集される。介入効果の推定には利用できるが、曝露とアウトカムを明確に時間的に分離する必要がある場合は、縦断的研究（コホート研究など）に比べて信頼性が低い。

### 5-3-5. 症例対照研究

[症例対照研究](#)では、アウトカムに基づいて個人が選択され、曝露のオッズが比較される。コホートの中に組み込まれた症例対照研究は、ネステッド症例コホート研究と呼ばれる。既存のデータベース研究内で実施される症例対照研究は、コホート研究よりも使用する情

報が少ないため、一般的には推奨されない ([Schuemie et al.](#)) 症例対照研究は、まれなアウトカムの場合、あるいは手作業によるカルテレビューや一次データ収集など、曝露に関するさらなる情報を収集する必要がある場合に最も有用である。

### 5-3-6. 準実験的研究

準実験的研究と自然実験は、因果効果を推定するために、結果とは無関係な、人全体または時間経過に伴う曝露の外的変動（「instrument」）を利用する ([Reeves et al.2017](#),[Matthay et al.2019](#))。一般的な準実験的手法には、instrument 変数分析、回帰不連続、中断時系列、差分推定などがある。これらは、ランダム化が必ずしも実行可能でない公衆衛生の場面で頻繁に使用されるが、医療技術評価でも使用されている ([不眠症と不眠症状を治療する Sleepio に関する NICE 医療技術ガイダンスを参照](#))。

instrument ベースのアプローチは、以下のような場合に有効である：

- 未知の交絡因子や測定が不十分な交絡因子による交絡が予想される。
- 対象となる曝露に関連し、その曝露を通じて以外は結果に影響を与えない適切な instrument が利用可能であること。

ヘルスケアのアプリケーションで使用されてきた手法の例としては、医師の治療嗜好や病院の処方ばらつき、遺伝子、医療提供者への距離や地理的な治療率、治療を受けるための任意のしきい値、あるいは時間（例えば、ケアパターンに即時かつ大きな影響を与える臨床ガイドラインの変更時間）などがある。

これらのアプローチの主な利点は、観察されていない共変量や十分に測定されていない共変量による交絡に対処できることである。しかし、使用する特定のデザインに依存する他の方法論的課題に加え、instrument の妥当性についても考慮する必要がある ([NICE 意思決定支援ユニットの技術支援文書 17 を参照](#))。これらの方法の適用は、通常、検証が困難な仮定に強く依存しており、実質的な知識と経験的な正当性に基づく妥当性の明確なケースが必要である。

### 5-4. 研究デザイン

このセクションでは、リアルワールドデータを用いたコホート研究と外部対照研究のための研究デザインに関する考察を示す。これらのアプローチは他の非ランダム化研究デザインにも有用であろう。

#### 5-4-1. ターゲット・トライアル・アプローチ

非ランダム化比較試験は、倫理的または実施可能性の問題に制約されずに理想的に実施されたであろうランダム化比較試験を模倣するようにデザインされるべきである ([Hernán and Robins 2016](#), [Gomes et al.](#))。ターゲットトライアルアプローチ（またはトライアルエミュレーション）として知られるこのプロセスは、[開発者](#)に試験デザインを明確に示すことを要求し、デザインの不備による選択バイアスの回避に役立つ ([Bykov et al.](#))。通常、対象となる試験は、対象集団を代表し、日常診療を反映したプラグマティックランダム化比較試験である。このアプローチは、非ランダム化比較試験のための Cochrane ROBINS-I risk of bias tool の基礎を形成している ([Sterne et al.](#))。

研究は対象とする臨床試験をできるだけ忠実にエミュレートすることを目指すべきであり、それが不可能な場合はトレードオフを明確に記述すべきである。場合によっては、データソースが十分な関連性や質を持たず、臨床試験のエミュレーションができないこともある。患者、設定、ケア、データ収集、期間などの違いにより、臨床試験とリアルワールドデータとの比較可能性が制限されるためである ([Gray et al.](#))。時には、リアルワールドデータを用いて対象試験を適切にエミュレートできず、特注のデータ収集が必要になることもある。

対象となる試験は、適格基準、治療戦略、割り付け方法、追跡期間、アウトカム、関心のある因果効果、解析計画の7つの側面から定義することができる。以下に各次元について説明し、NICE ガイダンスに情報を提供するためのエビデンスを開発する際の留意点を示す。

#### 5-4-1-1. 適格基準

ほとんどの研究において、適格基準は、NHS の日常診療で見られる臨床経路（診断検査を含む）と患者を反映させることにより、仮想的なプラグマティック試験を模倣すべきである。外部対照試験では、より広範な対象集団ではなく、介入試験の適格基準を満たすことに重点を置くべきである。試験と同様に、適格基準は治療割り付け前に記録された変数に基づくべきである。

介入効果に異質性が予想される場合は、[サブグループ解析](#)を行うことができる。研究を計画する際には、前もってサブグループを定義しておくべきである。

#### 5-4-1-2. 治療戦略

治療戦略には、対象となる介入と比較対照が含まれる。比較対照は、異なる曝露レベル（例えば、異なる用量の薬）、異なる介入、または介入がないことである。観察データで

は、選択バイアスや交絡のリスクが高いため、プラセボ対照試験を模倣することは非常に困難である。

同じ（または類似の）治療適応の比較対照（すなわち、積極的比較対照）は、介入を受けていない患者との比較よりも好ましい。積極的比較対照は、異なる介入を受ける患者の類似性をより確実にすることで、適応による交絡のリスクを低減する。ルーチンのフォローアップ手順が介入間で類似していれば、これも検出バイアスのリスクを減少させる。積極的比較対照は NHS で確立された診療を反映することが理想的である。

介入に関する研究では、一般的に新規（または偶発）利用者のデザインは、選択バイアスのリスクが低く、試験デザインの模倣がしやすいため、有病者（すでにしばらく介入を使用している人）の研究よりも好まれる。有症者は、定義上、治療を継続し、ある程度の期間追跡調査を受けて生存している。既に収集されたデータを利用する場合、新規使用者は通常、その個人が対象とする介入を使用することが観察されなかった最初の期間（薬剤疫学では「ウォッシュアウト」期間として知られる）を用いて定義される。新規使用者デザインのさらなる利点は、治療開始からの時変ハザードを推定できることである。介入の効果が累積的であったり、データ中の偶発的使用者が少なすぎたり、フォローアップが限定的であったりする場合には、有病者の組み入れが必要になることがある

[\(Vandenbroucke and Pearce 2015, Suissa et al.\)](#)

比較対照に関するデータは、介入と同じ期間、同じ医療制度や環境から得られることが理想的である。これは、ケアアクセス、パスウェイ（診断検査を含む）、アウトカムの時間的傾向の違いから生じる治療群間の差異を最小限にするためである。

### 5-4-1-3. 割り当て手続き

ランダム化比較試験では、個人（または群）が介入群にランダムに割り付けられる。可能であれば、医療提供者、患者、分析者はこの割り付けについて盲検化される。観察研究では、ランダム割り付けも盲検化も不可能である。しかし、交絡因子に関する十分な情報があれば、様々な分析的アプローチによってランダム割り付けを近似的に行うことができる（[分析のセクション](#)を参照）。

申請者によっては、複数の時点で適格基準を満たす場合もある。例えば、十分な曝露なし期間（または「ウォッシュアウト」期間）の後に、複数回治療を開始する場合がある。これに対処するには、最初の適格時点のみを使用する方法、ランダムな適格時点のみを使用する方法、すべての適格時点を使用する方法など、いくつかのアプローチがある

[\(Hernán and Robins 2016\)](#)。

#### 5-4-1-4. フォローアップ期間

フォローアップの開始と終了を定めなければならない。フォローアップの開始は、理想的にはすべての適格基準を満たし、介入を割り付けたのと同時（またはその直後）に開始すべきである。治療開始とアウトカムの間にかかなりの潜伏期間が予想される場合は、アウトカムをカウントしない誘導期間を定義することが必要であろう。これにより、アウトカムが曝露に影響する[逆因果](#)のリスクを減らすことができる。

追跡期間は、対象となるアウトカムを捕捉するのに十分な期間であるべきであるが、アウトカムが介入によって合理的に影響を受ける可能性のある期間（曝露-効果ウィンドウとして知られる）を超えてはならない。打ち切り事象は明確に定義されるべきであり、[関心のある因果効果によって異なる](#)。

#### 5-4-1-5. アウトカム

一次アウトカムと二次アウトカムを定義し、患者アウトカムと医療システムアウトカム（資源の使用やコストなど）の両方を含めることができる。患者のアウトカムは、患者がどのように感じ、どのように機能し、どの程度長く生きるかを反映するものでなければならない。これには QOL やその他の患者報告アウトカム指標が含まれる。客観的な臨床アウトカム（生存など）は、アウトカムの検出や報告が既知の治療歴に影響される可能性がある場合、主観的アウトカムよりもバイアスのリスクが低いのが一般的である。

サロゲートアウトカムについては、サロゲートアウトカムの変化が、対象となる最終的な患者のアウトカムの変化と因果関係があることを示す十分な証拠が必要である（[Ciani et al.](#)）。

観察データではアウトカムの確認は盲検化されないが、解析者は解析計画を確定する前にアウトカムを盲検化することができる（[解析のセクション](#)を参照）。

#### 5-4-1-6. 関心のある因果効果

研究者は、関心のある因果効果を記述すべきである。試験は通常 2 つの因果効果のうち 1 つを推定するようにデザインされる：介入への割り付けの効果（[intention-to-treat](#)）または治療プロトコルを遵守することの効果（[per-protocol](#)）。観察データを用いて治療割り付けの効果を推定することは通常不可能である。しかし、治療開始（as-started 効果）を用いて代理推定することができる。プロトコルごとの効果に相当するものは、on-treatment effect と呼ばれることもある。

通常、NICE が最も関心を持つのは開始時効果である。しかし、治療中止（または切り替え）が相当なものであったり、NHS における日常診療やアウトカムを反映しないと予想される場合には、治療中解析の結果を提示することが重要である。また、安全性と有害事象の解析には、治療中解析が最も適切であろう。On-treatment effect は、治療シーケンスや NICE が関心を持つ他の複雑な介入などの動的な治療戦略をカバーするために拡張することもできる。

#### 5-4-1-7. 分析計画

解析計画には、[介入事象](#)を考慮した上で、対象となる因果効果をどのように推定するかを記述する。同時発生事象とは、治療開始後に発生し（治療切り替えやアドヒアランスの欠如など）、対象となるアウトカムの解釈に影響を与える事象のことである。これは [estimand framework](#) によってサポートされる（詳細は[臨床試験における estimand と感度分析に関する ICH E9 \[R1\]の補遺](#)を参照）。

同時併発事象の関連性は、関心のある因果効果に依存する。開始時の解析では、治療の中止、切り替え、増強は通常無視できる。しかし、これらの変化が大きい場合、経時的に曝露の誤分類が増加するリスクがある。ほとんどの場合、これは効果の推定値を無い方向に偏らせる。

On-treatment 分析や動的治療戦略をモデル化する場合、患者が治療計画に従わなくなった時点で追跡を打ち切ることが多い。医薬品（およびいくつかの instrument）については、継続的な曝露は、処方日および予想される使用期間（例えば、供給日数から導き出される）によって代用され、観察間のある程度の猶予期間が許容される。打ち切りがアウトカムに依存し、介入間で異なる場合、バイアスを引き起こす情動的打ち切り（informative censoring）の可能性、および時変交絡に特に注意を払う必要がある。

交絡、情動的打ち切り、欠損データ、測定誤差への対処を含む統計解析に関するさらなる内容は、[解析セクション](#)に示されている。

[パネル 1](#) は、ターゲット・トライアルの使用例を示している：

#### パネル 1：ターゲット・トライアル・アプローチの例

##### 例 1：閉経後女性の冠動脈性心疾患に対する HRT 開始の効果は？

Women's Health Initiative ランダム化比較試験では、ホルモン補充療法を開始すると閉経後女性の冠動脈性心疾患のリスクが増加することが示された。これは、冠動脈性心疾患

のリスクを減少させたとする以前の観察研究と矛盾するものであった。[Hernánら \(2008年\)](#) は、Nurses' Health Study のデータを用いて、Women's Health Initiative 試験を可能な限り再現し、標的試験のアプローチをとった。彼らは、試験と観察研究の結果の違いは、観察コホートにホルモン補充療法の有病者を組み入れたことに起因することを示すことができた。これらの女性たちは、すでに治療期間中、アウトカムを経験することなく生存していたのである。新規使用者デザイン（ターゲットトリアルアプローチの他の原則も同様）に従って、彼らは試験と一致した効果推定値を出すことができた。

#### **例 2：進行した慢性腎臓病患者において、透析治療を開始する最適な推算糸球体濾過量（eGFR）はどのくらいか？**

IDEAL ランダム化比較試験では、透析開始時期が早い群と遅い群では、死亡率と心血管イベントの減少が緩やかであることが示された。早期治療群の平均 eGFR スコアは 9.0mL/min/1.73m<sup>2</sup>、後期治療群の平均 eGFR スコアは 7.2mL/min/1.73 m<sup>2</sup> であった。従って、透析を開始する最適な時期については依然としてかなりの不確実性がある。[Fuら 2021](#) 人は、National Swedish Renal Registry のデータを用いて IDEAL 試験をエミュレートし、試験で達成された eGFR の狭い区切りで同様の結果を得ることができた。そして、より広い範囲の eGFR 値に解析を拡張し、透析療法を開始する最適な時点を特定することができた。

#### **例 3：フルチカゾンプロピオン酸エステル+サルメテロール（FP-SAL）の治療開始と、1) FP-SAL なし、2) サルメテロールのみの治療の COPD 増悪に対する効果は？**

TORCH 試験では、FP-SAL の投与は、FP-SAL の投与なしまたはサルメテロールのみの投与と比較して、COPD 増悪リスクの減少に関連することが示された。しかし、この試験では 80 歳以上の成人、喘息または軽度の COPD 患者は除外されている。TORCH 試験の結果がこれらの患者にどの程度当てはまるかについては不確実である。[Wingら 2021](#) 人は、COPD 増悪に関する TORCH 試験の所見を、イングランドの Clinical Practice Research Datalink のプライマリケアデータを用いて、サルメテロールのみとの比較では再現できたが、FP-SAL なしとの比較では再現できなかった。これは、プラセボ対照試験を模倣することの難しさを反映している。より広い対象集団に解析を拡大することで、COPD 重症度による治療効果の不均一性のエビデンスを示すことができたが、年齢や喘息診断によるものではなかった。

## **5-5. 分析**

### **5-5-1. 交絡バイアスのリスクへの対応**

### 5-5-1-1. 交絡因子の同定と選択

潜在的な交絡因子は、透明性があり、体系的で再現性のあるプロセスに基づいて、解析の前に同定されるべきである。主な証拠源は公表文献と専門家の意見である。時変交絡因子の存在を考慮すべきである。これらの交絡因子は、アウトカムと将来の曝露レベルに影響し、以前の曝露レベルの影響を受ける可能性がある。これらの交絡因子は、時変介入や動的治療戦略をモデル化する場合や、情動的打ち切りに対処する場合に特に関連する。

開発者は、介入、[共変量](#)、関心のあるアウトカム間の因果関係についての仮定を概説すべきである。理想的には、これは directed acyclic graphs(DAG) ([Shrier and Platt 2008](#)) として知られる因果関係図を用いて行われるであろう。

共変量の不適切な調整は避けるべきである。これは、曝露とアウトカムの間の因果経路上の変数（過度の調整）、コライダー（合流点）、または instruments で管理することに起因する可能性がある。経時的に値が変化する可能性のある交絡因子は、時変交絡因子に適切に対処する統計的方法を用いる場合を除き、指標日以前に記録すべきである。

共変量の選択は、例えば共変量候補の数が非常に多い場合、十分な共変量セットを同定するために、機械学習などの高度な計算アプローチを用いることができる ([Ali et al .](#)) これらの方法の使用は明確に正当化され、因果関係の仮定との整合性が検証されるべきである。統計的有意性に基づく共変量の選択は避けるべきである。

### 5-5-1-2. 交絡に対処する方法を選択する

明確な因果関係の仮定に基づく調整された比較が、naive な（または調整されていない）比較よりも望ましい。交絡に対処し、ランダム化を近似するために統計的アプローチを用いるべきである ([割付け手順のセクション](#)を参照)。

層別化、マッチング、多変量回帰、[傾向スコア](#)法、またはこれらの組み合わせなど、観察された交絡因子を調整するために様々なアプローチを用いることができる。これらの方法は未測定の影響因子がないことを前提としている。層別化、制限、完全マッチングなどの単純な調整法は、交絡がよく理解され、よく記録されている交絡因子が少数である研究課題に適切であろう。

潜在的交絡因子が多い場合は、多変量回帰や[傾向スコア](#)（または疾患リスクスコア）法などのより複雑な方法が望ましい。傾向スコアは、観察された共変量に基づいて介入を受ける確率を与える。傾向スコアを用いる方法には、マッチング、層別化、重み付け、回帰（またはこれらの組み合わせ）などいくつかある。これらの異なるアプローチの長所と短

所に関する一般的な議論は、[Ali et al.](#)手法の選択は正当化されるべきであり、関心のある因果効果に沿ったものでなければならない。

交絡バイアスに対処するための回帰法と傾向スコア法の相対的な性能に関する証拠はまちまちである ([Stürmer et al.](#)) しかし、傾向スコア法を用いることは、研究実施の透明性という点で利点があるかもしれない：

- 傾向スコアはアウトカムデータを参照することなく作成されるため、強力な研究がバランスプロセスと組み合わせれば、結果の選択的報告のリスクを減らすことができる。
- ある種の傾向スコア法では、観察された共変量の観点から介入群の類似性を調べることが可能であり、比較可能性が達成された程度に関する証拠を提供する。0.1未満の絶対標準化差は一般的に良好なバランスを示すと考えられるが、その変数が結果に強い影響を及ぼす場合は、絶対差が小さくても重要な場合がある。

回帰法や傾向スコア法は、介入群間または介入レベル間の人々の類似性を高めるために、一部の参加者を除外することもある。このような方法を用いる場合、内部妥当性、検出力、一般化可能性のトレードオフを考慮すべきである。比較効果の研究では、一般的に内部妥当性を優先すべきである。

時間的に変動する交絡因子は、通常、上記の方法では調整すべきではない。経時的に変化する交絡因子が以前の介入レベルの影響を受けない場合は、治療中分析でも許容される場合があるが、これはまれである。重み付けを伴う限界構造モデルを含む G 法が望ましい ([Pazzagli et al.](#)) 時変交絡因子の調整には、追跡期間全体にわたる質の高いデータが必要である。

残留交絡によるバイアスを調整したり、その影響の可能性を探るために、様々な感度解析やバイアス解析を用いることができる ([研究の頑健性の評価の項を参照](#))。これは、交絡因子と結果の関係に関する外部データ、または解析データベースのデータが豊富なサブサンプルのデータがあれば、そのデータから情報を得ることができる ([Ali et al.](#)) ネガティブコントロール (すなわち、介入に関連しないと予想されるアウトカム) も有用である ([Lipsitch et al.2010](#))。

試験データの外部対照となる適切なリアルワールドデータの情報源が複数存在する可能性がある場合、開発者は各データ源について個別に効果を推定するか、データ源をプールすることで検出力を高めるかを検討すべきである。データ源のプールは、対象範囲やデータ

の質に関してデータ源間の異質性が限られている場合にのみ行うべきである。各データソースの効果の個別推定値は常に提供されるべきである。

外部対照は、ランダム化対照試験における内部（または同時）対照を補足するために用いることもできる。内部対照と外部対照を組み合わせるにはいくつかの方法があり、外部データの重みが異なる ([NICE Decision Support Unit report on sources and synthesis of evidence](#))。

観測されない交絡に対処するために、instrumentに基づく方法（または準実験計画）を用いることができる ([Matthay et al.](#)) 個々の患者レベルのデータを用いて、観察された特性や観察されていない特性によるベースラインの交絡に対処する方法に関するさらなる技術的ガイダンスは、[NICE の意思決定支援ユニットの技術支援文書 17](#) に示されている。

## 5-5-2. 情報の偏りへの対応

欠損データ、測定エラー、誤分類を含むデータの質の限界は、バイアスと精度の損失を引き起こす可能性がある。ここでは情報バイアスに対処するための分析的アプローチについて述べる。データ適合性を理解するために必要な情報は、情報バイアスの重要性についての洞察を提供する ([データ適合性の評価のセクション](#)を参照)。

### 5-5-2-1. 情報検閲

縦断的研究では、アウトカムが完全に観察される前にフォローアップが終了すると打ち切りが起こる。これは、データ収集期間の終了（管理上の打ち切り）、追跡不能、治療切り替え、非遵守、死亡などのイベントの発生によって起こり得る。アクティブな比較対照がない研究で、時間に関連したバイアスを避けるためのクローニングのような分析戦略によって誘発されることもある ([Hernán and Robins 2016](#))。

打ち切りが有益な場合（つまり、結果や治療割り付けに関連している場合）には、バイアスが生じる可能性がある。例えば、on-treatment 分析では、実験的薬剤を使用している人が、有益性がないと認識されたために治療プロトコルを遵守する可能性が低い場合、有益な打ち切りにつながる可能性がある。治療中効果や動的治療戦略をモデル化する場合、治療切り替えによる打ち切りは有益である可能性が高い。有益な打ち切りに対処する方法は、重み付けを用いた限界構造モデルやその他の G-手法など、時変交絡に対する方法と同様である ([Pazzagli et al.](#)) 欠損データに対処する方法も使用できる ([欠損データのセクション](#)を参照)。

### 5-5-2-2. 欠損データ

欠損データの影響は、欠損データの量、欠損データのある変数、欠損データのメカニズムに依存する。開発者は、関連性があれば、曝露群間や経時的な欠測のパターンを比較し、欠測の原因やそれが対象のアウトカムに関連しているかどうかを考慮すべきである。アウトカムに関するデータの欠落は、アンケートへの無回答や打ち切りなど、様々な理由で生じる可能性がある。

欠損データの量が少なく、完全にランダムに欠損している可能性が高い場合は、完全記録分析で十分である。欠損データを扱う高度な方法には、インプューテーション、逆確率重み付け、最尤推定などがある。これらの方法のほとんどは、利用可能なデータを用いて欠損データのメカニズムが適切にモデル化できる（つまり、ランダムに欠損する）ことを前提としている。そうでない場合は、感度分析またはバイアス分析が望ましいかもしれない（[頑健性の評価のセクション](#)を参照）。欠測データの取り扱いに関する枠組みは、[Carpenter and Smuk 2021](#)に記載されている。

### 5-5-2-3. 測定誤差と誤分類

測定誤差は、調査変数の測定値が真実からどの程度ずれているかを表す。カテゴリ変数の場合、これは誤分類として知られている。測定誤差の影響は、誤差の大きさと方向、誤差のある測定変数、誤差が介入グループ間で異なるかどうかによって依存する。測定誤差は、バイアスを引き起こしたり、推定値の精度を低下させたりする。

曝露のランダム測定誤差は、治療効果の推定値を無い方向に偏らせる傾向がある（ただし、常に偏らせるわけではない）（[van Smeden et al.](#)）連続アウトカムにおけるランダム測定誤差は、推定値の精度を低下させるが、比較効果の推定値にはバイアスがかからない。リスク比や割合比の場合、感度が低くても特異度が100%であれば、カテゴリ的アウトカムの非差別的誤分類によって比較効果の不偏推定値が得られる。そのため、高い特異度を達成するためにアウトカム変数を定義することがしばしば推奨される。

一般に、曝露、共変量、またはアウトカムにおける測定誤差の差は、比較効果の推定値に偏りを生じさせるが、その偏りの方向を予測することは困難である。測定誤差の構造や大きさに関するデータがあれば（例えば、内部または外部の検証研究によって）、この情報をキャリブレーションやその他の高度な手法を用いて分析に組み込むことができる（[van Smeden et al.](#)）

## 5-6. 研究の頑健性の評価

リアルワールドのデータを用いた比較効果の研究は複雑であるため、開発者はデータの収集や解析の際に多くの不確実な決定や仮定を行わなければならない。このような決定は、個々に、あるいは総体として、比較効果の推定値に大きな影響を与える可能性がある。従って、これらの仮定の逸脱に対する結果の頑健性を実証することが不可欠である。[表 3](#)では、いくつかの領域にわたる主要な感度分析について説明する。どの感度分析に重点を置くかは、データの長所と短所、およびバイアス、研究の仮定、不確実性の影響が最も大きい分野によって、事例によって異なる。これらのアプローチは、臨床効果の測定に直接適用することもできるし、費用対効果分析に反映させることもできる。

バイアスの主要なリスク（例えば、未測定の交絡、欠損データ、主要変数の測定誤差のために生じるもの）については、定量的バイアス分析が有用である。定量的バイアス分析とは、以下のような一連の技法である：

- 結果を変えたり、意思決定の閾値に影響を与えたりするために、どの程度のバイアスが存在しなければならないかを検討する。
- 効果測定に関連するバイアスの方向、大きさ、不確実性を推定する。

研究の結論を変えるためにどの程度のバイアスが存在しなければならないかを調べる方法は、より単純な傾向があり、e-value アプローチなどがある。これらのアプローチは、単一の測定不能なバイアス源を探索する場合に最も有用であるが、バイアス源は多くの場合複数存在し、相互作用する可能性がある。開発者は、これらの方法を適用する前に、パラメータのバイアスのもっとも妥当なレベルを検討し、事前に指定すべきである。より洗練されたアプローチでは、バイアスをモデル化し、効果の推定に組み込むことを検討している([Lash et al. 2014](#))。バイアスパラメータは、外部情報や分析データ源のデータが豊富なサブサンプルによって知ることができる。外部バイアスの同定と妥当性は、明確に記述され、正当化されるべきである。

バイアス解析は、データ収集、設定、時間の違いによってデータの比較可能性が低下する可能性がある場合、リアルワールドデータを外部対照に用いた研究で特に価値がある。[パネル 2](#)にリアルワールドのバイアス分析の例を、[表 3](#)に感度分析の例を示す。

## パネル 2：バイアス分析の例

クリゾチニブ不応の ALK 陽性非小細胞肺癌に対する ALK 阻害薬アレクチニブの有効性は、セリチニブと比較してどうか？

ALK 陽性非小細胞肺癌患者の全生存期間に対するアレクチニブとセリチニブの比較有効性は、head-to-head 試験がないため不明である。[Wilkinson ら \(2021 年\)](#) は、Flatiron

Health データベース（米国の電子カルテに由来）から得られたセリチニブのリアルワールドデータを用いて、第2相試験のアレクチニブ群に対する外部対照を形成した。著者らは、アレクチニブ投与開始群に有意な生存期間の改善を認めた。しかしながら、セリチニブを投与された患者の Eastern Cooperative Oncology Group Performance Status (ECOGPS) のベースラインデータが欠落していたこと（47%の患者でデータが欠落していた）、未測定の間接によるバイアスが残存する危険性があった。

これらのリスクを探索するためにバイアス分析法が用いられた。治療効果を除去するために必要な介入と死亡の間の観察されない間接因子の相対リスクを推定するために、e-value 法が用いられた。2.2 という推定相対リスクは、観察された間接因子よりもかなり高く、重要であるが捕捉が不十分な間接因子に関する推定不均衡を考慮すると、可能性は低いと考えられた。

ECOG データの欠落については、欠落データの原因が非ランダムであり、セリチニブ群における欠落データ値は、マルチプルインピュテーションに基づく予想値よりも悪い可能性が高いと仮定した。彼らは、欠損データに関するもっともらしい仮定では、介入と死亡率の間に観察された [関連](#) を説明できないと主張した。

### 表 3

#### データキュレーション、試験デザイン、解析の決定に対する結果の頑健性を調べる感度解析の例

ドメイン	感度分析またはバイアス分析の例
曝露の誤分類	<ul style="list-style-type: none"> <li>治療中の分析</li> <li>関連性がある場合は、供給日数、猶予期間、洗浄期間、暴露-影響ウィンドウ、潜伏期間を含む暴露の定義を変える。</li> </ul>
結果の誤分類	<ul style="list-style-type: none"> <li>既知のパフォーマンス指標を調整する</li> <li>定量的バイアス分析</li> </ul>
人口	<ul style="list-style-type: none"> <li>別の患者適格基準</li> </ul>
検出バイアス	<ul style="list-style-type: none"> <li>医療利用の測定値を共変量として含める</li> </ul>

データキュレーション、試験デザイン、解析の決定に対する結果の頑健性を調べる感度解析の例

ドメイン	感度分析またはバイアス分析の例
	<ul style="list-style-type: none"> <li>• ベースライン以前に医療システムと定期的な接触があった者に限る</li> </ul>
フォローアップ時間	<ul style="list-style-type: none"> <li>• 開始時および治療中の分析</li> <li>• インフォーマティブな打ち切りのために、アウトカム期間をグループ間で同じになるように制限する。</li> <li>• 一般ユーザーと新規ユーザーの分析</li> </ul>
逆の因果関係	<ul style="list-style-type: none"> <li>• 曝露終了から追跡調査開始までのタイムラグを導入または変更する。</li> </ul>
交絡	<ul style="list-style-type: none"> <li>• 選択した交絡因子を追加または削除する</li> <li>• 共変量が特定されるルックバック期間の延長</li> <li>• ネガティブコントロール（改竄エンドポイントまたはプローブ変数とも呼ばれる）を用いて、治療とは無関係であるはずのアウトカムについて、同じモデルを用いて比較効果を推定する（これらの結果は、効果推定値の校正にも用いることができる）。</li> <li>• 検証研究で観測された変数を用いて、観測された効果推定値を未測定バイアスのために調整する傾向スコアキャリブレーション</li> <li>• 定量的バイアス分析</li> </ul>
欠損データ	<ul style="list-style-type: none"> <li>• さまざまな方法を使う</li> <li>• 統計モデルに共変量の欠損変数指標を含める</li> <li>• 定量的バイアス分析（例えば、ランダムメカニズムではない欠損を仮定する）</li> </ul>
モデル仕様	<ul style="list-style-type: none"> <li>• モデルの仕様を変える</li> <li>• 異なる前提を持つ分析アプローチを使用する（三角測量）</li> </ul>

## データキュレーション、試験デザイン、解析の決定に対する結果の頑健性を調べる感度解析の例

ドメイン	感度分析またはバイアス分析の例
データキュレーション	<ul style="list-style-type: none"><li>連続変数の代替分類または除外データの調整</li></ul>

### 5-7. 報告

[定量的リアルワールドエビデンス研究の実施に関するセクション](#)で、リアルワールドエビデンス研究の透明性のある報告と優れた実施に関する一般原則を示す。以下の報告に関する考慮事項は、比較効果試験において特に重要である：

- リアルワールドのエビデンスの使用を正当化する。これには、ランダム化エビデンスがない理由、既存の臨床試験の限界、特定の研究課題に対して意味のあるリアルワールドエビデンスを作成する能力などを、関連するものとして含める。
- 分析が行われる前に、一般にアクセス可能なプラットフォーム上で研究プロトコル（[統計分析計画](#)を含む）を公表すること。
- 独立した研究者が研究を再現できるよう、研究を十分に詳細に報告すること。
- 研究デザイン図を提示する。
- 各データソースについて、データの出所と目的適合性を理解するために必要な情報を提供する（[データの目的適合性の評価に関するセクション](#)を参照）。
- 交絡に対処するための統計的手法の使用を正当化し、その方法を明確に報告する（[付録3](#)参照）。
- 最初のデータから最終解析までの患者の除外を、患者フロー（または減少）図を用いて、除外の理由も含めて明確に記述する。
- 可能であれば、統計的調整前後の治療群間の患者の特徴を提示すること。外部対照研究については、変数の定義やデータ収集の違いを明確に記述する。
- 調整後解析と非調整後解析、すべてのサブグループ解析と感度解析、バイアス解析の結果を示す。

### 5-8. 品質評価

エビデンス作成者は、[研究計画](#)段階でバイアスのリスクを特定すべきである。また、感度分析またはバイアス分析を用いて、主解析の仮定からの逸脱に対して結果がどの程度頑健であるかを記述する。これは、[付録2](#)の報告方法を用いて、特定のバイアス領域について

行うことができる。この情報は、リスクオブバイアスツールを完成させる（または批評的に評価する）際に役立つ。非ランダム化研究のバイアスリスク評価ツールとして推奨されるのは ROBINS-I ([Sterne et al.](#) 非ランダム化研究の不確実性は、通常、推定介入効果の統計的不確実性では完全には捉えられないことを認識すべきである ([Deeks et al. 2003](#)))。

開発者は、NHS の対象集団に対する試験結果の一般化可能性についてコメントすべきである。これは、患者、医療環境、治療経路、時間の違いから導き出され、[データ適合性評価から](#)提供された情報によって裏付けられる。分析標本と対象集団の患者の特性の違いは、必ずしも結果が一般化されないことを意味しないことに留意すべきである。これは、推定の規模（例えば、相対効果か絶対効果か）、治療効果における異質性の程度、これが適切にモデル化されているかどうかなど、他の要因に依存する。

## 付録 1 - データ適合性評価ツール (DataSAT)

- [DataSAT 評価テンプレート](#)
- [DataSAT - ケーススタディ](#)

ダウンロード可能な DataSAT 評価テンプレートについては、[ツールとリソース](#)をご覧ください。

### DataSAT 評価テンプレート

#### 研究課題

ここにリサーチクエスチョンを追加する。

#### データの実証性

項目	応答
データソース	各データ提供元について、その名称、バージョン、データ収集日を記入すること。ウェブサイトがある場合は、そちらへのリンクも記入すること。
データ・リンケージとデータ・プール	リンクされたデータセット、リンク方法、リンクのパフォーマンス特性を報告する。リンケージが第三者（NHS デジタルなど）によって行われたかどうかに留意する。  どのデータソースがプールされたかを明確に記述する。
データ・ソースのタイプ	データ源の種類（例えば、電子カルテ、登録、監査、調査）を記述する。
データ収集の目的	データ収集の主な目的（例えば、臨床治療、償還、instrument の安全性、調査研究）を記述する。
データ収集	収集した主なデータの種類（臨床診断、処方、処置、患者経験データなど）、データの記録方法（臨床コーディングシステム、自由記述、遠隔モニタリング、調査回答など）、データの収集者（医療従事者、自己申告、デジタルヘルス技術など）を記述する。データ収集の性質がデータ収集期間中に変更された場合（例えば、コーディングシステムまたは 実践、データ収集システムの変更）、その変更点を明確に記

## データの実証性

項目	応答
	<p>述する。データ提供者間で、データの収集方法、収集内容、およびその質に違いがある場合は、それを記述する。</p> <p>調査研究のために追加的なデータ収集が行われた場合は、データ収集の妥当性と一貫性がどのように保証されたかを含めて記述してください（例えば、研修など）。</p>
ケア設定	<p>使用した各データセットのケア設定（例えば、プライマリケア、セカンダリケア、専門医療センター、ソーシャルサービス、自宅での使用（ウェアラブルデバイスの場合、またはアプリやウェブサイトでの自己申告データ））を明記する。</p>
地理的設定	<p>データソースの地理的範囲を明記すること。</p>
人口カバー率	<p>対象集団がデータセットによってどの程度表現されているかを述べる（例えば、集団の代表性や患者数）。</p>
データの期間	<p>データの対象期間を明記すること。</p>
データ準備	<p>分析のために生データにアクセスしたかどうか、あるいはデータ所有者がクレンジングや変換などのデータ調整ステップを実施したかどうかの詳細を示す。共通データモデルへの集中的な変換が行われたかどうか。共通データモデルのタイプ、バージョン番号、マッピングの詳細など、関連情報へのリンクを含む。</p> <p>研究課題への取り組みに特化したデータ準備の完全な詳細は、<a href="#">データキュレーションに関する報告のセクション</a>で取り上げている。</p>
データガバナンス	<p>各ソースについて、データ管理者と資金提供者の詳細を記載する。データへのアクセスと使用のための情報ガバナンスプロセスを記述する。</p>
データ仕様	<p>データ仕様書があるかどうかに注意。これにはデータモデル、<a href="#">データディクショナリ</a>、またはその両方が含まれる。</p>
データ管理計画と品質保証方法	<p>データ管理計画、ソース品質保証方法の文書化、関連文書へのリンクの有無に留意する。</p>

## データの実証性

項目	応答
その他の資料	<p>その他の文書があるかどうかを記す。もしあれば、主要な出版物へのハイパーリンクまたは引用を示す。</p> <p>データセットが Health Data Research UK (HDRUK) のイノベーションゲートウェイから入手できる場合は、HDRUK のウェブサイト上のプロフィールへのハイパーリンクを提供すること。</p>

## データの質

データの質に関する詳細は、集団の適格基準、アウトカム、介入または曝露、共変量などの主要な研究変数について提供されるべきである。

研究変数	ターゲット・コンセプト	運営上の定義	品質次元	評価方法	評価結果
どのような変数か（例えば、集団の適格性、結果）。	対象概念を定義する（例えば、心筋梗塞 [MI]）。	運用上の定義を定義する。例えば、MI は主診断位置の ICD-10 コード I21 で定義される。	選択：正確か完全か	質をどのように評価したかを記述する。該当する場合は、過去の検証研究の参考文献を示すこと。	可能であれば、質の定量的評価を記載する。例えば、「陽性的中率 85%（75%～95%）」など。

## データの関連性

データ関連性の報告に関する推奨事項を参照のこと。

項目	応答
人口	分析標本が対象集団をどの程度反映しているかを記述する。これは、（例えば、主要な予後変数に関するデータの欠落による）データの除外を考慮する必要がある。

データ関連性の報告に関する推奨事項を参照のこと。

項目	応答
ケア設定	ケアの設定が、NHSにおける日常的なケアをどの程度反映しているかを記述する。
治療経路	データ中の人々が経験した治療経路が、NHSにおける日常的な治療経路（診断検査を含む）をどのように反映しているかを記述する。
主要な研究要素の入手可能性	主要な集団の適格性基準、アウトカム、介入、共変量（交絡因子や効果修飾因子を含む）を含む必要なデータ変数が利用可能であるという点で、データセットが研究課題の要件をどのように満たしているかに留意する。
研究期間	データがカバーする期間が、意思決定にどの程度関連した情報を提供するかを述べる。これは、ケアパス（検査を含む）の重要な変更やアウトカム率の背景的な変化をカバーするものでなければならない。
測定のタイミング	測定のタイミングが研究課題のニーズに合っているかどうかを記述する。
フォローアップ	データセットで利用可能なフォローアップ期間が、アウトカムを評価するのに十分であることを留意されたい。
サンプルサイズ	データセットに含まれる対象集団のサンプルサイズを提示し、それが頑健な結果を生み出すのに十分であることを証明する。

### DataSAT - ケーススタディ

なお、このケーススタディの報告は、[Wingら 2021年の公開情報](#)に基づいている。

#### 研究課題

慢性閉塞性肺疾患（COPD）の患者において、長時間作用性  $\beta 2$  アゴニストと吸入コルチコステロイドの配合剤であるフルチカゾンプロピオン酸エステルとサルメテロールの併用は、サルメテロールのみの併用や曝露なしと比較してどのような効果があるか？

## データの実証性

項目	応答
データソース	<p><a href="#">臨床実践研究データリンク (CPRD) GOLD</a></p> <p><a href="#">病院エピソード統計 (HES) 入院患者ケアデータ</a>。</p>
データ・リンケージとデータ・プール	<p>CPRD と HES はリンクされている。患者は NHS digital が行う集中的な連結アルゴリズムで特定される。これは 4 つの識別子に基づく 8 段階の決定論的連結アルゴリズムを用いている：NHS 番号、性別、生年月日、郵便番号である。</p> <p>登録患者の 75% について HES データとのリンクが可能である。</p> <p><a href="#">CPRD のリンクデータに関する</a>情報を参照。</p>
データ・ソースのタイプ	<p>HES = 管理記録</p> <p>CPRD = 電子カルテ</p>
データ収集の目的	<p>Hospital Episode Statistics (HES) は、医療提供者が NHS デジタルに提出した情報に基づき、Secondary Uses Service (SUS) のデータから作成される。データ収集の主な目的は、イングランドにおけるサービス提供に対する病院の診療報酬をサポートすることである。</p> <p>CPRD は、英国内の GP 診療所のネットワークから匿名化された患者データを収集する。当初、このデータは患者がプライマリ・ケア・サービスを受けている間に収集される。</p>
データ収集	<p>CPRD = 人口統計、臨床診断 (Read v2 または SNOMED-CT)、検査 (medcode または SNOMED-CT)、処方 (prodcode) (投与量、投与経路、投与期間を含む)。CPRD GOLD は、Vision ソフトウェアシステムを使用して、GP 診療所から完全にコード化された患者の電子カルテを収集する。データは、Vision ソフトウェア内で働く医療・介護スタッフによって記録される。</p> <p>HES = 診断 (ICD-10)、処置 (OPCS-4)、入院、退院、ケアの種類、基本的な人口統計。HES データは、患者が病院にいる間に収集され、病院内の医療・介護スタッフとのやり取りの中で記録され、臨床コーダーのチームによって組み立てられる。</p>
ケア設定	<p>HES = 二次医療</p>

## データの実証性

項目	応答
	CPRD = プライマリーケア
地理的設定	HES = イングランド  CPRD = Vision ソフトウェアを使用した英国の一般診療所の代表的サンプル。HES にリンクされた CPRD データはイングランドでのみ利用可能。
人口カバー率	CPRD GOLD には、現在登録されている約 300 万人分（英国人口の約 4.74%）のデータがある。 <a href="#">CPRD データのハイライトを見る</a>  HES のデータはイングランドのすべての NHS 臨床委員会グループをカバーしている。
データの期間	CPRD とリンクした HES データセットは、2000 年 1 月から 2017 年 1 月までをカバーしている。
データ準備	CPRD の詳細は不明。ただし、一般診療所は、その記録が研究的な質を持つことを証明した後にのみ組み入れられる。  HES は、データが研究用に公開される前に集中処理を適用する：  HES データセットの処理中に実行されるルール。データの価値と質を向上させるために設けられており、特定のフィールド内のデータの検証、追加フィールドと値の導出、無効または HES データセットの範囲外のレコードの削除などのルールが含まれる。
データガバナンス	CPRD は、保健社会医療省（DHSC）の執行機関である MHRA のセンターである。したがって、DHSC は CPRD データのデータ管理者である。  HES データは Health and Social Care Information Centre（NHS Digital としても知られる）によって管理されている。  CPRD は、MHRA、Wellcome Trust、Medical Research Council、NIHR Health Technology Assessment program、Innovative Medicines Initiative、英国保健省、Technology Strategy Board、Seventh Framework Programme EU、および様々な大学、委託研究機関、製薬会社から資金提供を受けている。

## データの実証性

項目	応答
	<p>HES のデータ収集は英国政府によって義務付けられ、資金援助を受けている。</p> <p><a href="#">CPRD のデータ保護および処理に関する通知</a>。</p> <p><a href="#">病院エピソード統計 GDPR ウェブページ</a></p>
データ仕様	<p>HES のフィールドは <a href="#">NHS データモデル</a>と <a href="#">NHS データ辞書</a>に由来する。</p> <p><a href="#">CPRD GOLD データ仕様書</a>。</p>
データ管理計画と品質保証方法	<p>HES は処理とデータ品質チェックを行う：<a href="#">処理サイクルと HES のデータ品質</a></p> <p>CPRD GOLD のデータ品質保証情報は確認されなかった。しかし、個々の一般診療所からの記録は評価され、研究品質と判断された後にのみ CPRD に収録される。</p>
その他の資料	いない。

## データの質

研究変数	ターゲット・コンセプト	運営上の定義	品質次元	評価方法	評価結果
人口	慢性閉塞性肺疾患	COPD の CPRD 診断 (Read v2) コード ( <a href="#">Quint et al.</a> )	精度	COPD 患者を特定するアルゴリズムを、医師による質問票をゴールドスタンダードとして比較した検証試験 ( <a href="#">Quint et al.</a> )	陽性適中率 (PPV) : 87% (95%信頼区間 [CI]78% ~92)
人口	疾患の重症度	スパイロメトリー測定による慢性閉塞性	完全性	スパイロメトリーデータが欠落	20%

## データの実証性

項目	応答	
	<p>肺疾患 (GOLD) ステージ (<a href="#">コードリスト</a>参照)</p>	<p>している患者の割合</p>
<p>介入</p>	<p>フルチカゾン プロピオン酸 エステル+サルメテロール</p> <p>CPRD 処方記録と<a href="#">コードリスト</a>で決定された薬物治療の定義との一致</p>	<p>精度</p> <p>CPRD の処方データは高精度が期待される</p> <p>該当なし</p>
<p>成果</p>	<p>COPD 増悪</p> <p>以下のいずれか：</p> <p>下気道感染症または COPD 急性増悪の CPRD 診断 (読み) コード</p> <p>COPD に特異的な抗生物質と経口コルチコステロイド (OCS) を 5 ~ 14 日間併用する。</p> <p>AECOPD の呼吸器症状が 2 つ以上あり、COPD に特異的な抗生物質および/または OCS が同日に処方された記録 (Read code) 。</p>	<p>精度</p> <p>COPD 増悪患者を特定するアルゴリズムを、医師による質問票をゴールドスタンダードとして比較した検証試験 (<a href="#">Rothnie et al.</a>)</p> <p>PPV: 86% (95% CI 83% to 88%)</p> <p>感度: 63% (95%CI 55%~70)</p>

## データの実証性

項目	応答				
		<a href="#">コードリストを参照</a>			
成果	全死因死亡率	国家統計局（ONS）の死亡率統計に記録（CPRD データに一元的にリンクされている）	精度	ONS の死亡記録は、死亡に関する最も標準的なデータである。	該当なし
共変量体（交絡因子）	アルコール摂取量	CPRD で直接報告された（最も近い指標日）	完全性	アルコール摂取に関するデータが欠落している患者の割合	30%

## データの関連性

項目	応答
人口	<p>CPRD に登録された患者の人口統計学的特徴は、より広い英国人口と類似している。CPRD の結果は、一般的に、より広い対象集団に一般化されると予想される。</p> <p>社会経済的状態、飲酒量、BMI のデータが欠落している記録を除いて完全な記録解析を行った。これらの変数はすべて、データの欠落が 5%未満であった。</p> <p>患者の約 5 分の 1 は、CPRD にスパイロメトリーの測定値が記録されていないため除外された。測定をしていない患者は、医療サービスとの接触が少ない傾向があり、結果の一般性に影響を与える可能性がある。</p>
ケア設定	<p>適切である。COPD 治療薬は通常プライマリケア（CPRD）で投与され、関連するイベントはプライマリケアまたはセカンダリケア（CPRD または HES）で観察される。</p>
治療経路	<p>このデータは、NHS における日常的な診療を表している。</p>

## データの関連性

項目	応答
主要な研究要素の入手可能性	<p>曝露とアウトカムに関する十分なデータが入手可能である。CPRD からは処方データのみ入手可能で、調剤データは入手できないが、これは調剤の良い代用品となることが期待される。</p> <p>重要な交絡因子である可能性のある陰性可逆性スパイロメトリーの結果に関する情報は得られなかった。</p> <p>CPRD では用量に関する情報は限られている。</p>
研究期間	調査期間以降、COPD 管理に関する英国の臨床実践に大きな変化はない。
測定のタイミング	縦断的な分析の性質により、研究課題に答えることができる。エントリー日は、リアルワールドの臨床イベントのタイミングをよく反映することが期待される。
フォローアップ	COPD 増悪という主要アウトカムが発生するには、平均 2 年の追跡で十分である。
サンプルサイズ	COPD 増悪について必要なサンプルサイズは、80%、有意差 5%で各群 600 と推定された（詳細は <a href="#">Wing ら 2021</a> を参照）。リアルワールドのサンプルサイズは各群約 2,500 で、これをはるかに超えている。

## 付録 2-バイアスのリスクを最小化するために用いた方法の報告

- [メソッド報告テンプレート](#)
- [メソッド報告 - ケーススタディ 1](#)
- [メソッド報告 - ケーススタディ 2](#)

ダウンロード可能な偏見報告テンプレートについては、[ツールとリソース](#)を参照のこと。

### メソッド報告テンプレート

#### バイアスのリスクを最小化するために使用した方法を報告するためのフォーム

バイアスの種類	バイアスはどのように対処または評価されたか
試験参加時の選択バイアス	試験参加時の選択バイアスは、曝露とアウトカムに関連する適格基準に基づく患者の選択、または患者が適格基準を満たした日、治療が割り当てられた日、および追跡調査の開始の間のずれなど、いくつかの理由で生じる可能性がある。時間に関連したバイアスの一般的なタイプは、有病者バイアス、リードタイムバイアス、不死時間バイアス、感受性物質の枯渇である。試験開始時の選択バイアスの可能性と、試験デザイン、統計解析、感度解析を通じてどのように対処または調査されたかを説明する。
試験終了時の選択バイアス	個人の研究からの退出方法による選択バイアスの一般的な原因は、情動的打ち切り (informative censoring) である。これは追跡不能や打ち切り事象の発生が原因かもしれません。有益な打ち切りの可能性と、分析でどのように対処したかについて議論してください。
交絡への対応	<p>測定されていない（または未知の）交絡因子、測定が不十分な交絡因子、または時変交絡による交絡のリスクを記述する。これは、潜在的な交絡因子の体系的な同定、時変交絡の可能性を含む明確な因果関係の仮定、比較群間のベースライン特性の違いによって知らされるべきである。</p> <p>研究デザイン（適切なアクティブコンパレータの選択など）と解析（適切な統計モデルを用い、時変交絡を考慮する）を通して、特定された交絡リスクにどのように対処したかを示す。可能であれば、調整後のベースライン特性のバランスに関する経験的データを提供する。</p> <p>残存交絡について懸念が残る場合は、感度分析またはバイアス分析を用いて結果への影響を評価したことを示す。</p> <p>バイアスを誘発するような不適切な調整を行った共変量がないことを確認する。例えば、介入とアウトカム間の因果経路上の共変量が調整されていないこと（過剰調整）を示す。これは、指標日以降に測定された共変量の使用から生じる可能</p>

## バイアスのリスクを最小化するために使用した方法を報告するためのフォーム

バイアスの種類	バイアスはどのように対処または評価されたか
	性がある。コライダーや器具の調整は避ける。これは因果関係図によって知ることができる。
検出バイアス	<p>比較群間の医療行為の違い（例えば、フォローアップの頻度や強度、検査の違いなど）やフォローアップの期間の違いから生じる検出バイアスの可能性を記述する。</p> <p>試験デザイン（例えば、類似のフォローアップを行う比較対照の使用）や解析（例えば、指標日以前の医療利用の調整）により、どのように対処したかを記述する。</p>
測定誤差と誤分類	<p>測定誤差や誤分類によるバイアスの可能性を記述する（これはデータの適合性の評価によって知らされるべき）。どの変数が不正確なのか、それがランダムなのか組織的なのか、比較群間でどのように異なるのかを検討する。</p> <p>統計分析（外部データやキャリブレーションを取り入れるなど）を通じてバイアスのリスクに対処したこと、または感度分析やバイアス分析を用いて結果への影響を評価したことを示す。</p>
欠損データ	<p>欠損データによるバイアスの可能性を記述する（これはデータの適切性の評価によって知らされるべき）。どの変数に欠損データがあるか、それがランダムか系統的か、比較群間でどのように異なるかを検討する。</p> <p>統計的手法（マルチプルインピュテーションなど）を用いてバイアスのリスクにどのように対処したかを示し、その妥当性を証明する。欠測が観測された変数で説明できない場合やメカニズムが不明な場合は、感度分析やバイアス分析を用いて、「ランダムではない」欠測の仮定による影響を調べることができます。</p>
逆の因果関係	<p>変数間の因果関係、介入とアウトカムのデータ記録間のタイムラグ、またはケア経路から生じる介入とアウトカム間の逆因果のリスクを記述する。</p> <p>逆因果のリスクが研究デザイン（例えば、導入期間や縦断的追跡調査）、解析（例えば、instrument 変数）、感度分析による評価を通じてどのように対処されたかを記述する。</p>

### メソッド報告 - ケーススタディ 1

なお、このケーススタディの報告は、[Fu et al.2021](#) の公開情報に基づいている。

この研究では、進行した慢性腎臓病患者において、異なる推定糸球体濾過量（eGFR）での透析開始が心血管イベントと生存に及ぼす影響を評価している。この研究では Swedish Renal Registry のデータが用いられた。

### Fu らの 2021 年版に基づく方法報告ツールの完成例。

バイアスの種類	バイアスはどのように対処または評価されたか
試験参加時の選択バイアス	<p>透析開始時期の効果に関するこれまでの観察研究では、適格基準を満たした時期、治療割り付け時期、追跡開始時期が一致していないために生じるリードタイム・バイアスや不死時間バイアスのリスクが高い。この試験は IDEAL 試験から情報を得たターゲット試験を模倣した。時間ゼロの誤指定の問題を避けるために、この研究ではクローニング、打ち切り、重み付け法を用いた。患者はクローン化され、eGFR に従って各治療に割り付けられ（ベースケースでは 15 の治療戦略の一つ）、所定の治療戦略から逸脱した時点で打ち切られる。この方法は、IDEAL 試験の結果を同試験でみられた eGFR 値の範囲で再現することによって検証された。</p> <p>この集団ベースの研究では、集団の選択による選択バイアスは問題にならなかった。</p>
試験終了時の選択バイアス	<p>選択バイアスは、患者が「治療戦略」に従わなくなったときの打ち切りによって誘発される可能性がある。打ち切りによって誘発される選択バイアスに対処するために、ベースラインおよび時変交絡因子を用いて打ち切り重みの逆確率を推定した。</p> <p>フォローアップまでの損失は非常に少ない。</p>
交絡への対応	<p>アウトカムモデルでは、人口統計、検査値、前治療、入院などのベースライン測定値を調整した。時系列的に変化する交絡因子は、eGFR の現在および過去の測定値を含む打ち切り重みで調整した。</p> <p>筋量貯蔵、尿毒症症状、体積状態、QOL、身体活動など、他の重要な交絡因子の可能性のあるデータについては入手できず、尿中アルブミン-クレアチニン比と血漿カリウムについてはコホートの一部についてのみ入手可能であった。残存交絡の可能性を評価するために、本研究では以下の感度分析を行った：</p>

**Fu らの 2021 年版に基づく方法報告ツールの完成例。**

バイアスの種類	バイアスはどのように対処または評価されたか
	<ul style="list-style-type: none"> <li>• adjusted for urine albumin-creatinine ratio and plasma potassium in the subset of patients with measurements and observed no impact on results</li> <li>• replicated the results of the IDEAL trial over the eGFR separation observed in the trial.</li> </ul>
Detection bias	<p>Outcomes included 5-year all-cause mortality and major adverse cardiovascular events (composite of cardiovascular death, non-fatal myocardial infarction, or non-fatal stroke). These are likely to be accurately observed regardless of small differences in level of surveillance, for example, resulting from earlier dialysis treatment.</p>
Measurement error and misclassification	<p>Timeliness and accuracy of variables extracted from the Swedish Renal Registry have previously been demonstrated. In particular, cardiovascular comorbidities have a very high positive predictive value, generally between 85% to 95%.</p> <p>eGFR was calculated with the Chronic Kidney Disease Epidemiology equation from routine plasma creatinine measurements. This has been shown to be accurate to within 30% of measured glomerular filtration rate 85% of the time.</p>
Missing data	<p>Data on initiation of dialysis and key outcomes are thought to be complete. Data on mandatory items such as eGFR is also very high.</p> <p>For non-mandatory data items in the registry, missingness was greater. For example, body mass index was missing in 26% of patients, urinary albumin to creatinine ratio in 44%, and potassium in 29%. This was assumed to be missing completely at random and determined by the preferences of the attending physician. Sensitivity analysis in the subset of people with data available showed had no impact on results.</p>

**Fu らの 2021 年版に基づく方法報告ツールの完成例。**

バイアスの種類	バイアスはどのように対処または評価されたか
Reverse causation	Reverse causation is not expected to be a problem in this analysis.

**メソッド報告 - ケーススタディ 2**

このケーススタディの報告は、[Wilkinson et al.の 2021](#) 年に公表された情報に基づいていることに留意されたい。

本試験は、ALK 陽性非小細胞肺癌患者の生存期間に対するアレクチニブとセリチニブの比較有効性を推定するものである。本試験では、Flatiron Health 社のセリチニブに関するリアルワールドデータを使用し、第 2 相試験でアレクチニブを投与された患者に対する外部対照を形成する。

**Wilkinson らの 2021 年に基づく方法報告ツールの完成例。**

バイアスの種類	バイアスはどのように対処または評価されたか
試験参加時の選択バイアス	<p>本試験では、アレクチニブを投与された第 2 相試験に登録された患者と、セリチニブを投与された米国の日常診療の患者を比較した。選択バイアスのリスクを最小化するためにいくつかのステップがとられた：</p> <p>リアルワールドのデータにおける組み入れ基準を、試験に組み入れられた集団と一致させる。</p> <p>リアルワールドデータで観察されなかった前治療歴のある患者を試験から除外すること</p> <p>試験期間と類似した期間のリアルワールドデータを使用する。</p> <p>新規ユーザー向けのアクティブコンパレータ設計を採用</p> <p>このアプローチの有効性を実証するために、リアルワールドのデータのみを用いて比較を繰り返したところ、同様の結果が得られた。</p>

**Wilkinson らの 2021 年に基づく方法報告ツールの完成例。**

バイアスの種類	バイアスはどのように対処または評価されたか
試験終了時の選択バイアス	これは、フォローアップの損失が限定的な、開始時の分析である。打ち切りは有益ではないと思われる。
交絡への対応	<p>主要な予後予測変数は、システマティックレビューによって前向きに特定された。</p> <p>既知の主な交絡因子は、限界はあるものの、データとして把握されていた。データの欠損や主要交絡因子の誤分類への対処については以下を参照。</p> <p>ベースライン時またはそれ以前に測定された観察された交絡因子を用いて傾向スコアを推定した。推定には治療重みの逆確率法を用いた。調整後、治療群間の共変量パターンに大きな差があるという証拠はなかった（標準化平均差はすべての変数で 0.1 未満）。</p> <p>感度分析では、追加の変数で調整しても結果は変わらなかった。</p> <p>定量的バイアス分析を用いて、未知の交絡因子が推定治療効果を排除するためにどの程度強い交絡効果を持つ必要があるかを評価した。推定された e 値は 2.4 であり、これは、交絡因子-死亡率および交絡因子-治療関連について、測定された交絡因子で見られたレベルよりもかなり高いレベルを必要とする。</p>
検出バイアス	死亡率という結果には検出バイアスはないと考えられた。
Measurement error and misclassification	<p>Data on mortality is sufficiently well captured in the real-world data with sensitivity of 91% and specificity of 96%.</p> <p>サーベイランスが限られているため、リアルワールドデータでは中枢神経系転移が誤って分類されている（過小報告）ことが懸念された。感度分析の結果、推定治療効果を排除するためには、リアルワールドデータにおける有病率を 40%大きくする必要があったことがわかった。</p>
欠損データ	ベースラインのパフォーマンスステータス（European Cooperative Oncology Group [ECOG]スコア）のデータ

**Wilkinson らの 2021 年に基づく方法報告ツールの完成例。**

バイアスの種類	バイアスはどのように対処または評価されたか
	<p>欠落は、リアルワールドのデータで多く（32%）、これは重要な予後変数である。主解析では、完全症例解析においてデータが完全にランダムに欠落していると仮定した。</p> <p>この仮定は無効であることが予想されたため、データがランダムに欠落していると仮定して多重代入（multiple imputation）を用いて感度分析を行った。結果は完全症例解析と一致した。</p> <p>定量的バイアス解析は、ECOG スコアがインピュテーションモデルで予想されるよりも悪い場合に、ランダムでないデータの欠落に関する懸念に対処するために行われた。閾値分析を用いたところ、ECOG スコアの欠落がある場合のどのような合理的な仮定においても、研究の結論は同様であった。</p>
逆の因果関係	この分析では、逆の因果関係は問題にならないと思われる。

付録 3-特定の分析法に関する報告情報

選択された分析法に関する報告の手引き

方法	説明	報告情報
直接的または間接的な標準化	選択された共変量に関して暴露群の比較可能性を高める方法	<ul style="list-style-type: none"> <li>標準参照集団（記述）</li> <li>標準化に使用した共変量</li> </ul>
階層化	データを部分集合（層）に分けて分析する。	<ul style="list-style-type: none"> <li>層の共変量定義</li> <li>各層のオブザベーション数</li> <li>各層における記述統計と結果</li> </ul>
マッチング	同一または類似の特徴を持つ個人のマッチング	<ul style="list-style-type: none"> <li>マッチングに使用される変数</li> <li>マッチング・アルゴリズム</li> <li>適合キャリパー（該当する場合）</li> <li>マッチング率</li> <li>一致数と除外数</li> </ul>
傾向スコア（一般）	特定の介入を受ける確率の推定；利用可能な方法の範囲（下記）	<ul style="list-style-type: none"> <li>傾向スコアの推定に用いたモデル（ロジスティックや多項式など）</li> <li>使用した共変量とモデルへの組み込み方法</li> <li>調整前後の傾向スコア分布（マッチング前後など）</li> <li>マッチ分析、トリミング分析、</li> </ul>

選択された分析法に関する報告の手引き

方法	説明	報告情報
		<p>切り捨て分析、重み付け分析に寄与した N/%。</p> <ul style="list-style-type: none"> <li>統計分析の診断チェック</li> <li>高次元の傾向スコアモデルの報告については、<a href="#">Tazareら 2022</a> を参照のこと。</li> </ul>
傾向スコア（層別化）	傾向スコアに基づいて患者を層（例えば十分位）に分類し、層特異的効果を集計する。	<ul style="list-style-type: none"> <li>地層の定義</li> <li>トリミングと、層定義の前か後か</li> <li>Tables for stratified population characteristics</li> </ul>
Propensity score (weighting)	Weights attached to individuals based on inverse of propensity scores	<ul style="list-style-type: none"> <li>How weights are calculated</li> <li>Whether and how weights are trimmed, truncated or stabilised</li> <li>非加重および加重人口特性の表</li> <li>ウェイトの平均値と分布</li> </ul>
傾向スコア（マッチング）	傾向スコアが類似している個人をマッチング	<ul style="list-style-type: none"> <li>キャリパーとスケールを含むマッチングアルゴリズムを使用</li> </ul>

## 選択された分析法に関する報告の手引き

方法	説明	報告情報
		<ul style="list-style-type: none"> <li>• マッチング比率 (固定 1 : 1 や 変動 1 : 5 な ど)</li> <li>• 非マッチ集団と マッチ集団の特 徴の表</li> </ul>
<p>多変量回帰調整 (傾向スコアの使 用を含む)</p>	<p>介入と共変量の関数としてアウト カムを比較する統計モデル</p>	<ul style="list-style-type: none"> <li>• モデルの種類 (線形回帰やポ アソンなど)</li> <li>• 使用した共変量 とその組み入れ 方</li> <li>• 診断チェック</li> </ul>
<p>instrument 変数 分析</p>	<p>instrument」を用いて、人全体 または経時的な曝露の外的変動を 探索する。instrument 変数は介 入と関連するが、それ以外は結果 とは無関係である。</p>	<ul style="list-style-type: none"> <li>• モデルの種類 (2 段階最小二 乗法など) と診 断チェック</li> <li>• instrument と 介入との関連性 の強さ(例え ば、オッズ比、 リスク差)。</li> <li>• 理論的正当化: 介入を通さない 限り、 instrument は 結果に影響を与 えず、 instrument は 結果と原因を共 有しない。</li> <li>• instrument と 介入のレベル間 の集団特性の分 布を示した表</li> </ul>

## 選択された分析法に関する報告の手引き

方法	説明	報告情報
		<ul style="list-style-type: none"> <li>• バイナリーアウトカム、暴露、instrumentについては、instrument、治療、アウトカムの各組み合わせの度数の表</li> <li>• instrument 変数分析における具体的な因果効果とその従属前提（例えば、単調性）による報告については、<a href="#">Swanson and Hernán 2013</a> を参照のこと。</li> <li>• 改竄テストの結果：具体例については <a href="#">Labrecque and Swanson 2018</a> を参照のこと。</li> </ul>
<p>中断された時系列</p>	<p>個人または集団を独自の対照とし、複数の時点にわたって観察する。効果は、介入前後の期間における結果の傾向を比較することによって観察される。</p>	<ul style="list-style-type: none"> <li>• モデルの種類（通常の最小 2 乗回帰による区分線形回帰など）</li> <li>• Study time period and time intervals</li> <li>• 介入効果の時点の事前特定（例えば、分析時点が介入実施時点でない場合は説明が必要など）</li> </ul>

## 選択された分析法に関する報告の手引き

方法	説明	報告情報
		<ul style="list-style-type: none"> <li>• 介入前、介入後、介入間のデータポイント数（時間）、および予測に寄与するデータポイント数。</li> <li>• 分析した各グループ（例えば、介入前後および定義したサブグループ）について、参加者の特性および欠測データを比較した表。</li> <li>• 経時的な結果（すなわち、介入前後の傾向）を示す表とグラフ。</li> <li>• 診断的チェックの結果（例えば、自己相関、定常性、季節性、モデル仕様のチェック）、および行った調整。</li> <li>• 改ざんテストの結果（例えば、介入実施前の疑似開始期間の使用など）</li> </ul>

## 枠組みの開発経緯

- [背景](#)
- [NICE 開発チーム](#)

### 背景

私たちは、研究機関や専門機関、その他の規制機関や医療技術評価機関による研究や既存のベストプラクティス・ガイダンスを照合して、この枠組みを開発した。

私たちは、この枠組みの策定にあたり、一連のワークショップと公開コンサルを通じてフィードバックを求めた。

私たちは、以下を含む多くのステークホルダーと関わり、フィードバックを得た：

- 患者および患者団体
- ヘルス・チャリティーズ
- 医療従事者
- 製薬・医療技術業界
- データ管理者および委託研究機関
- アカデミー
- 国際医療技術評価機関
- 英国の医療システム・パートナー
- NICE 委員会のメンバー。

いただいたご意見をもとに、枠組みを修正しました。

リアルワールドのエビデンス枠組みの開発およびレビューに参加して下さったすべての方々に感謝したい。

### NICE 開発チーム

シェイマス・ケント、リン・キンケード、マヌジ・シャルマ、ショーン・ロワルク、ステイーブン・ダフィールド、ヴァンダナ・アイヤー・グプタ、ジョアン・グロソップ、ポール・ヨンソン