

薬事申請に Real World Data を外部対照として利用する際の留意点

日本製薬工業協会 医薬品評価委員会 データサイエンス部会

2020 年度 タスクフォース 1

2022 年 7 月

目次	
略語一覧表	3
1. はじめに	5
1.1. Real world data が注目されている背景	5
1.2. ランダム化比較試験が薬効評価のゴールドスタンダードである所以	5
1.3. 本報告書の目的, 想定読者, 適用範囲, 構成	5
2. 事例調査	6
2.1. BLINCYTO (Blinatumomab)	7
2.2. XPOVIO (Selinexor)	9
2.3. BALVERSA (Erdafitinib)	10
2.4. ROZLYTREK (Entrectinib)	12
2.5. KYMRIAH (Tisagenlecleucel)	14
2.6. BRINEURA (Cerliponase alfa)	17
2.7. プログラフ (タクロリムス水和物)	18
3. RWD 利用事例からの Lessons learned	20
3.1. 事例調査から見てきた課題	20
3.2. 比較可能性の問題	21
3.2.1. 選択基準および除外基準	22
3.2.2. 交絡	22
3.2.3. エンドポイント	23
3.3. サンプルサイズ	23
3.4. 規制当局とのコミュニケーション	24
3.5. 利用する RWD の調査	24
3.6. 結果の透明性の担保	26
3.6.1. Objective Study Design	27
3.7. 事例から得られた留意点のまとめ	29
4. おわりに	30
参考文献	34
Appendix	39
RWD 利用時に発生しうる種々のバイアス	39

略語一覧表

略語	略していない表現（英語）	略していない表現（日本語）
ALL	acute lymphoblastic leukemia	急性リンパ芽球性白血病
ASCT	autologous stem cell transplant	自己造血幹細胞移植
B-ALL	B cell acute lymphoblastic leukemia	B 細胞性急性リンパ芽球性白血病
BiTE	Bi-specific T-cell engaging	二重特異性 T 細胞誘導
BOR	best objective response	最良総合効果
BRIDGE	Bladder Cancer Research Initiative for Drug Targeting in Germany	－
CAR	chimeric antigen receptor	キメラ抗原受容体
CI	confidence interval	信頼区間
CR	complete response	完全奏効
DLBCL	diffuse large B-cell lymphoma	びまん性大細胞型 B 細胞リンパ腫
DOR	duration of overall response	奏効期間
ECOG-PS	Eastern Cooperative Oncology Group Performance Score	ECOG パフォーマンスステータス
EHR	electronic health record	電子健康記録
EMA	European Medicines Agency	欧州医薬品庁
EMR	electronic medical record	-
FDA	Food and Drug Administration	米国食品医薬品局
FGFR	fibroblast growth factor receptor	線維芽細胞成長因子受容体
FMI	Foundation Medicine, Inc.	Foundation Medicine 社
IPI	international prognostic index	国際予後指標
IPTW	inverse probability of treatment weighting	逆確率重み付け
MAIC	matching adjusted indirect comparison	-
ORR	overall response rate	全奏効率
OS	overall survival	全生存期間
PD	progression disease	（腫瘍）悪化
PFS	progression free survival	無増悪生存期間
Ph	3hiladelphia chromosome	フィラデルフィア染色体

PR	partial response	部分奏効
RCT	randomized controlled trial	ランダム化比較試験
RECIST	New Guidelines to Evaluate the Response Evaluation Criteria in Solid Tumors	固形がんの治療効果判定のための新ガイドライン
ROS1	ROS proto-oncogene 1, receptor tyrosine kinase	c-ros がん遺伝子 1
RRMM	relapsed/refractory multiple myeloma	再発または難治性多発性骨髄腫
RWD	real world data	リアルワールドデータ
RWE	real world evidence	リアルワールドエビデンス
TRK	tropomyosin receptor kinase	神経栄養因子受容体／トロポミ オシン受容体キナーゼ
TTD	time to treatment discontinuation	治療中止までの期間
TTR	time to response	奏効までの時間

1. はじめに

1.1. Real world data が注目されている背景

2016年に米国にて制定された 21st Century Cures Act¹により、薬事承認を目的とした Real World Data（以下、RWD）の利活用が推奨されたことや、（超）希少疾患、再生医療、遺伝子治療など、ランダム化比較試験の実施が困難な開発の増加、医薬品の開発コストの高騰、デジタル技術の進歩など様々な理由から、RWD の活用が注目されている²。

2000年に発出された ICH E10³で薬効評価を行う上で対照群を設定する重要性や外部対照の利用がすでに述べられていたが、昨年最終化された ICH E8(R1)⁴では多様なデータソースを前提とした外部対照の利用やその留意点についても述べられている。RWD を外部対照として利用する際の統計学的方法論の研究もなされている。このような中で、各国の規制当局や研究団体からは、医薬品開発に RWD を活用する際のガイダンスや提言が発出されている^{5,6,7,8,9,10,11,12,13}。日本製薬工業協会でも、これまでに医薬品開発に RWD を利活用した事例調査や提言を行うなど、医薬品開発への RWD の利用可能性について検討してきた^{14,15,16,17}。このような背景を踏まえると、今後も RWD の利用は進んでいくと思われる。なお、本成果物では RWD の定義を「多様なデータソースから日常的に収集される患者の健康状態及び／又は医療提供に関するデータ」¹²とした。

1.2. ランダム化比較試験が薬効評価のゴールドスタンダードである所以

ランダム化比較試験（以下、RCT）では、ランダム化によって既知及び未知の交絡因子の群間での偏りを回避することで比較可能性が担保されることから、興味のある治療効果の群間の違いが直ちに薬効として解釈可能となる。したがって、RCT は薬効評価を行う上でゴールドスタンダードとして認識されている。一方、ランダム化を伴わない比較試験／研究（以下、非 RCT）では、一般に既知及び未知の交絡因子が群間で偏っていないことをデザインにより担保することが不可能であり、そのような比較に基づく群間の効果の違いは交絡に起因するバイアスが発生している可能性が否定できない。そのため、ランダム化を伴わずに RWD を比較対照として用いた結果を薬事申請に用いた場合は、その結果の解釈が当局と議論となることがある。このような背景から、現時点では薬事申請での外部対照としての RWD の利用可能性や当局による受け入れ可否の予見性は十分とは言えない。

1.3. 本報告書の目的、想定読者、適用範囲、構成

本報告書では、今後の薬事申請において RWD を適切に使用し、その利活用を促進することを目的として、日本製薬工業協会医薬品評価委員会臨床評価部会 TF2「リアルワールドデータを承認申請へ～活用促進のための提言～」¹⁶で取り上げられた過去の薬事申請で RWD が利用された事例のうち、RWD を外部対照として用いた事例をピックアップし、

主に統計学的観点から審査中の議論について調査した。また、これらの事例からの学びを整理し、今後の RWD 利活用における留意点をまとめた。本報告書では、既存の報告や RWD の要約結果を単純に比較するような RWD の利用ではなく、「RCT の同時対照を代替させる目的での RWD の外部対照としての利用」を主な焦点を当てた。

本報告書の主なスコープは統計学的観点からの留意点ではあるが、統計方法論の詳細については深入りしない。また、薬事申請における外部対照としての RWD の利用に限定しているが、本報告書の内容は非 RCT 全般に該当する項目も含むため、観察研究データを主に扱う部署の担当者や RCT を実施する機会が多くないファーマコビジランス関連の部署の担当者など、薬事申請に携わる統計担当者以外に対しても広く参照可能な内容となっている。

本報告書では、2 節に事例調査の結果、3 節では事例調査結果に基づく RWD 利活用に関する留意事項を整理した。最後に本タスクフォースの提言として、薬事申請で RWD を外部対照として利用する際の留意事項をまとめた。なお、RWD を利用する際には、データ自体の信頼性（データの発生源から分析に利用されるまでの流れの透明性が確保されているか）、疾患やイベントの特定方法（アルゴリズム）の妥当性確保などといったその他多くの課題や論点もあるが、本報告書ではこれらはスコープ外とした。本報告書が薬事申請を目的とした RWD の外部対照としての利活用の参考となれば幸いである。

2. 事例調査

2020 年 4 月に臨床評価部会タスクフォース 2 がまとめた成果物「リアルワールドデータを承認申請へ ～活用促進のための提言～」¹⁶ の「リアルワールドデータを活用した承認申請の事例一覧表」をもとに、外部対照として RWD を利用した事例を特定し、審査報告書を中心に調査を行った。本章では、統計学的観点から当局と議論があった下記の事例（表 1）を中心に、その概要を紹介する。

表 1 調査対象とした承認品目

薬剤名（一般名）	対象疾患	承認年	データソース
BLINCYTO ^{®18,19} (Blinatumomab)	再発又は難治性の Ph 陰性 B-ALL	2014 年 (米国)	レジストリ 文献
XPOVIO ^{®20} (Selinexor)	再発又は難治性の多発性骨髄腫	2019 年 (米国)	EHR
BALVERSA ^{®21} (Erdafitinib)	局所進行性又は転移性尿路上皮癌	2019 年 (米国)	EHR
ROZLYTREK ^{®22} (Entrectinib)	ROS1 融合遺伝子陽性の局所進行性 または転移性非小細胞肺癌	2019 年 (米国)	EHR
KYMRIAHA ^{®23} (Tisagenlecleucel)	再発又は難治性 B 細胞性急性リンパ 芽球性白血病	2018 年 (欧州)	レジストリ
	再発又は難治性びまん性大細胞型 B 細胞リンパ腫	2018 年 (欧州)	レジストリ
BRINEURA ^{®24,25,26} (Cerliponase alfa)	神経セロイドリポフスチン症 2 型	2017 年 (米国)	レジストリ
プログラフ ^{®27} (タクロリムス水和物)	多発性筋炎・皮膚筋炎に合併する間 質性肺炎	2013 年 (日本)	文献

2.1. BLINCYTO (Blinatumomab)

ヒストリカルデータを用いて、単群試験の主要評価項目の解析で設定した閾値の妥当性を評価した事例であり、米国食品医薬品局（以下、FDA）から閾値の妥当性の説明資料として受け入れられている。

開発の経緯

BLINCYTO (Blinatumomab) は、二重特異性 T 細胞誘導 (BiTE) 抗体であり、悪性 B 細胞に T 細胞を誘導することで抗腫瘍効果を発揮する免疫療法剤である。米国では、FDA により画期的治療薬及び希少疾病用医薬品に指定され、2014 年 9 月に第 2 相臨床試験 (MT103-211 試験) を主要な試験成績として、再発又は難治性の Ph 陰性 B-ALL に関する承認申請が行われた。2014 年 12 月、18 歳以上の再発又は難治性の Ph 陰性 B-ALL 患者を対象とした第 III 相試験の結果を提出することを条件に迅速承認された。

MT103-211 試験は、18 歳以上の再発又は難治性の Ph 陰性 B-ALL 患者を対象に多施設非盲検単群第 2 相試験として実施された。主要評価項目は最初の 2 サイクル (4 週間投与後

2週間休薬の計6週間を1サイクルとした) 以内の完全寛解又は造血機能の部分的な回復を伴う完全寛解の達成率であった。主要評価項目の結果は41.6% (95% CI: 34.4%, 49.1%) であり、事前に定められた閾値30%を超えたことが確認された。

申請者によるヒストリカルデータの活用内容

申請者はMT103-211試験で用いた閾値(30%)の妥当性を確認する目的で、事前に計画したヒストリカルデータを用いた以下の解析を実施した。

1) 標準治療での完全寛解率の推定 (201200310 試験)

201200310 試験はMT103-211試験と同様の予後因子(年齢及び前治療の種類)の分布を有する集団における標準治療での完全寛解率を推定する目的で実施された。ヒストリカルデータとして、欧州及び米国の研究グループによる標準治療を受けた再発又は難治性のPh陰性B-ALLの成人患者の血液学的寛解率と生存率に関する1990年から2014年の間で利用可能なデータベースを用いた。解析には、MT103-211試験と同様の前治療歴を有し、完全寛解に関するデータが入手可能な694例を含めた。予後因子による層別解析の完全寛解率に基づき、MT103-211試験と同様の予後因子の分布になるよう重み付け解析を行った結果、標準治療での完全寛解率は24% (95% CI: 20%, 27%) と推定された。

2) 既存療法に対するblinatumomabの効果の予測 (118427 試験及び119834 試験)

118427 試験では1995年から2012年の間に公表された臨床研究の要約データに基づき、既存療法を受けた再発及び難治性ALLの成人患者の血液学的寛解及び全生存率を定量的に評価するための予測モデル(混合効果モデル)を構築した。

119834 試験では上記の予測モデルを用いて、再発及び難治性のALL患者における完全寛解率、完全寛解の期間、及び全生存率における既存の救済療法と比較したblinatumomabの効果を予測した結果、既存の救済療法と比較し、blinatumomabの完全寛解率が高いことが示唆された(12% vs. 33%)。

FDAのヒストリカルデータ利用への見解

主要な試験成績であるMT103-211試験で用いた閾値(30%)の妥当性の説明資料として受け入れられた。

2.2. XPOVIO (Selinexor)

単群試験の外部対照群をヒストリカルデータから構成し、薬事申請時の臨床データパッケージに含めた事例であるが、FDA から外部対照群に対する懸念点が示され、承認審査における意思決定のエビデンスとして受け入れられなかった。

開発の経緯

XPOVIO (Selinexor) は経口核外輸送タンパク質阻害剤であり、4 回以上の治療歴があり、病状としては 2 剤以上のプロテアソーム阻害剤、2 剤以上の免疫調整剤、及び抗 CD-38 モノクロナール抗体の治療に抵抗性が認められた再発または難治性多発性骨髄腫患者 (RRMM) の治療薬である。米国では、優先審査および希少疾病用医薬品に指定されており、ランダム化第 3 相試験で有効性を検証することを市販後の条件として、2019 年 7 月に主要な試験成績である第 2b 相試験 (STORM 試験) の結果に基づき、迅速承認された。

STORM 試験は RRMM 患者を対象とした多施設非盲検単群第 2b 相試験として実施された。主要評価項目は国際骨髄腫ワーキンググループによる多発性骨髄腫の統一効果判定基準に基づく独立評価委員会の判定による全奏効率 (ORR: overall response rate) であり、閾値は 10% と設定された。主要評価項目の結果は 25.4% (95% CI: 18%, 34.1%) であった。

申請者によるヒストリカルデータの活用内容

申請者は STORM 試験と類似した患者集団での全生存時間 (OS: overall survival) と STORM 試験の結果を比較する目的で、EHR (Electronic Health Record) から抽出したデータを用いた後ろ向きの観察研究 (KS-50039 試験) を実施し、STORM 試験の補助データとして薬事申請時の臨床データパッケージに含めた。なお、EHR として Flatiron Health Analytic Database (FHAD) を用いた。

FHAD から 64 例の患者が申請者により規定された選択基準及び除外基準を満たし、解析された。STORM 試験の 122 例の OS と比較した結果、FHAD での OS に対するハザード比は 0.41 (95% CI: 0.26, 0.65) であった。

FDA のヒストリカルデータ利用への見解

FDA は KS-50039 試験について、以下に挙げる問題点を指摘しており、STORM 試験の補助データとして妥当ではないと判断した。

1) プロトコルと統計解析計画書が事前に提出されていないこと

FDA は承認審査の意思決定に観察研究を受け入れる場合、透明性を高め、妥当性の評価を適切に行うために、事前にプロトコル及び統計解析計画書を提出することを要求している。しかし、KS-50039 試験のプロトコル及び統計解析計画書が事前に提出されておらず、本試験の妥当性が判断できない。

2) FHAD からのデータ抽出に用いた選択基準の問題

STORM 試験の選択基準と FHAD からのデータ抽出に用いた選択基準は大きく異なり、選択バイアス、誤分類、及び交絡が生じている可能性が高い。

3) OS の起点の問題

STORM 試験及び KS-50039 試験ともに OS の起点日は、前治療が失敗（終了）した時点としており、同じ定義を用いていた。しかし、STORM 試験の患者は少なくとも XPOVIO の投与を受けており、その時点までは生存しているが、KS-50039 試験の患者は 64 例中 27 例で後治療を受けておらず、早期に脱落しており、STORM 試験の OS の推定にはバイアス（immortal time bias: 不死時間バイアス）が生じている可能性がある。

4) 比較可能性の問題

上記 2) で指摘した選択基準の相違に加え、以下の点から比較可能性に懸念がある。

- ・ KS-50039 試験では後治療を受けられなかった患者が存在し、STORM 試験の患者より予後が悪い患者が含まれていた可能性がある。
- ・ KS-50039 試験ではベースライン時の ECOG-PS や腫瘍ステージ等の重要な予後因子が欠測である患者が少なくなく、両試験の患者背景を十分に比較できない。また、申請者は比較可能性を高めるため、傾向スコアに基づく IPTW（inverse probability of treatment weight）による解析を実施したが、傾向スコアを推定するための予後因子が十分でなかったこと、症例数が十分でないことなどにより、傾向スコアによるアプローチは適切ではないと判断した。

2.3. BALVERSA (Erdafitinib)

単群試験の外部対照群をヒストリカルデータから構成し、薬事申請時の臨床データパッケージに含めた事例であるが、FDA から外部対照群に対する懸念点が示され、承認審査における意思決定のエビデンスとして受け入れられなかった。

開発の経緯

BALVERSA (Erdafitinib) は、経口の線維芽細胞成長因子受容体 (FGFR) キナーゼ阻害剤であり、成人の局所進行性または転移性尿路上皮癌に対する治療法である。米国では、優先審査に指定されており、検証試験 (BLC3001) データを市販後に提出することを条件として 2019 年 4 月に主要な試験成績である第 2 相試験 (BLC2001) の結果に基づき、迅速承認された。

BLC2001 試験は、限られた FGFR 遺伝子変異を有する転移性又は外科的に切除不能な尿路上皮癌患者を対象とした多施設非盲検第 2 相試験として実施された。主要評価項目は RECIST 1.1 に基づく独立評価委員会の判定による最良総合効果であり、閾値は 25% と設定された。主要評価項目の結果は 32.2% (95% CI; 22.4%, 42.0%) であった。

申請者によるヒストリカルデータの活用内容

申請者は FDA との申請前の協議を踏まえ、EHR (Flatiron-FMI 及び BRIDGE データベース) を用いた後ろ向きの観察研究を実施した。研究の目的は①両データベースを用いて FGFR 遺伝子変異と既存治療の効果との関連の評価すること、②Flatiron-FMI から抽出した現在利用可能な治療を受けた患者と BLC2001 試験で BALVERSA を投与された患者の OS を比較することであった。以下では、②の目的に焦点を当てて紹介する。

FGFR 陽性の局所進行性または転移性尿路上皮癌患者について、BCL2001 試験で BALVERSA の投与を受けた患者と Flatiron-FMI データベースで抽出した PD-(L)1 阻害剤もしくは化学療法を受けた患者の OS を比較した結果、BALVERSA は既存治療に比べ、延命効果があると結論付けた。

FDA のヒストリカルデータ利用への見解

FDA は、以下に挙げる問題点を指摘しており、結果の妥当性が統計手法の問題点によって大きく影響を受けるため、BALVERSA の有効性を裏付ける証拠として使用するのに不十分であると結論付けた。

1) 未測定の前因子及び主要な前因子に重大な欠測

外部対照群で生存時間の重要な前因子である ECOG-PS に欠測が多い (28-37%)。また、外部対照群でその他の前因子 (内臓転移など) が収集されていない。

2) 選択バイアス

選択基準及び除外基準に重大な相違がある（例えば、ECOG-PS による除外基準は外部対照群では設定していない）。また、BCL2001 試験は主にヨーロッパの学術医療センターから登録され、外部対照群は米国の地域腫瘍学センターから登録されており、患者がおかれている医療環境は異なる可能性がある。

3) 潜在的な治療の誤分類

Flatiron-FMI に登録されている医療機関以外での治療情報は入手不可能であり、抗がん剤治療歴を正確に把握できない可能性がある。また、統計解析計画書および報告書のいずれにおいても、外部対照群の患者の治療情報の入手方法および妥当性の評価についての記載がない。

4) 外部対照群での死亡情報の収集が不完全である懸念

一般的に地域医療機関では、退院や転院後の生存状況を追跡しないため、外部対照群の死亡情報は不完全である可能性がある。また、統計解析計画書および報告書のいずれにおいても、患者の死亡情報の確認方法についての記載がない。

5) 非常に小さいサンプルサイズ

外部対照群で解析に含まれた患者数は 16～25 例であり、結果を一般化するには非常に小さい。

2.4. ROZLYTREK (Entrectinib)

単群試験の外部対照群をヒストリカルデータから構成し、薬事申請時の臨床データパッケージに含めた事例であるが、FDA から外部対照群に対する懸念点が示され、承認審査における意思決定のエビデンスとして受け入れられなかった。

開発の経緯

ROZLYTREK (Entrectinib) は、ROS1 (c-ros がん遺伝子 1) 及び TRK (神経栄養因子受容体) ファミリーを強力かつ選択的に阻害する経口投与可能なチロシンキナーゼ阻害剤である。米国では、FDA により画期的治療薬及び優先審査に指定され、2019 年 8 月に主要な試験成績である第 2 相試験 (STARTRK-2 試験) 及び第 1 相試験 (ALKA 試験, STARTRK-1 試験) の統合解析の結果に基づき、検証試験により有効性を検証することを

条件として、ROS1 融合遺伝子陽性の局所進行性又は転移性非小細胞肺癌の治療薬として、迅速承認された。

上記 3 試験はすべて多施設非盲検単群試験として実施された。3 試験の統合解析での主要評価項目は RECIST ver1.1 に基づく独立中央判定委員会の判定による奏効率であった。主要評価項目の結果は 78.4% (95% CI: 64.7, 88.7%) であった。

申請者によるヒストリカルデータの活用内容

申請者は ROS1 融合遺伝子陽性非小細胞肺癌患者を対象に本剤（上述の 3 つの単群試験のデータ）と crizotinib（ヒストリカルデータ）を比較することを目的とした WO40977 試験を実施し、承認申請時の臨床データパッケージに含めた。本試験の主要評価項目は治療中止までの期間（TTD: time to treatment discontinuation）、副次評価項目は無増悪生存期間（PFS: progression-free survival）及び全生存期間（OS）であった。なお、ヒストリカルデータは Flatiron Health Analytic Database（米国のオンコロジー施設における電子カルテ情報由来の RWD）を用いた。

傾向スコアに基づく IPTW 法を適用し、年齢、性別、人種、喫煙歴、脳転移及び前治療数を層別因子とした Cox 回帰モデルにより解析した結果、TTD のハザード比は 0.64 (95% CI : 0.40~1.02) であり、crizotinib に比べ、本剤では治療中止のリスクが低いと結論付けた。

FDA のヒストリカルデータ利用への見解

FDA は、以下に挙げる問題点を指摘しており、結果の妥当性が統計手法の問題点によって大きく影響を受けるため、ROZLYTREK の有効性を裏付ける証拠として使用するのに不十分であると結論付けた。

1) プロトコルが事前に提出されていないこと

WO40977 試験のプロトコルは事前に提出されていなかったため、すべての解析は事後解析と位置付けられた。

2) 選択バイアス

比較する群間で選択・除外基準に大きな違いがあり、ベースライン時の集団の特徴に違いがある。

3) 未測定 of 交絡因子及び主要な交絡因子の欠測

ROZLYTREK の臨床試験での選択・除外基準の判定に用いた多くのデータがヒストリカルデータでは未測定であり、患者の抽出に適用できなかった。また、ヒストリカルデータでは、55.1%の患者で ECOG-PS が欠測であった。

4) 統計解析の限界

サンプルサイズが小さく、すべての交絡因子を統計モデルの共変量として含めていないことから、解析上の共変量の調整が不十分である。

5) 評価項目の測定の問題

データベースで利用可能なデータを考慮した上で、一般的に受け入れ可能な評価項目を定義していたが、TTD は増悪後の継続投与により解釈が複雑となり、PFS はデータベース上の X 線画像の欠測によって入手が制限され、OS はベースラインの不均衡によるバイアスの影響を受けやすくなる可能性がある。

2.5. KYMRIAH (Tisagenlecleucel)

単群試験の外部対照群をヒストリカルデータから構成し、欧州医薬品庁（以下、EMA）に対する薬事申請時の臨床データパッケージに含めた事例である。EMA から外部対照群に対して、いくつかの懸念点が示されたが、最終的に KYMRIAH の有効性を示すエビデンスとして受け入れられている。なお、KYMRIAH は「小児および若年成人（3~25 歳）の B 細胞性急性リンパ芽球性白血病（B-ALL: B-cell acute lymphoblastic leukemia）」及び「成人のびまん性大細胞型 B 細胞リンパ腫（DLBCL: diffuse large B-cell lymphoma）」の 2 つの適応を取得したが、EMA の審査報告書において、B-ALL については外部対照の利用に対する記述はほとんどなかった。したがって、本報告書では DLBCL の審査における外部対照の利用について纏めた。

開発の経緯

KYMRIAH (Tisagenlecleucel) は患者由来の T 細胞を遺伝子的に変異させ、CD19 と呼ばれるがん細胞を特異的に認識するキメラ抗原受容体 (CAR: chimeric antigen receptor) を導入後、患者にその変異 T 細胞を輸注する再生医療等製品である。欧州では再発又は難治性の DLBCL の 18 歳以上の成人患者を対象とした第 2 相試験 (C2201 試験) を主要な試験成績として、2018 年 8 月に承認された。

C2201 試験は、再発又は難治性の DLBCL の 18 歳以上の成人患者を対象とした多施設非盲検単群第 2 相試験として実施された。主要評価項目は国際的に用いられている Lugano 基準を用いて独立判定委員会が効果判定を行った全奏効率（ORR: overall response rate）であった。ORR は完全奏功（CR: complete response）又は部分奏功（PR: partial response）が得られた被験者の割合と定義した。副次評価としては奏効までの時間（TTR: time to response）、奏効期間（DOR: duration of overall response）、無増悪生存時間（PFS: キムリア輸注から最初に記録された PD（悪化）又は理由を問わない死亡までの時間）、全生存時間（OS: キムリア輸注から理由を問わない死亡までの時間）等の時間依存の評価項目が含まれた。登録症例 165 例中、主要評価項目の主要解析はすべての被験者がキムリアを輸注後少なくとも 3 ヶ月追跡調査を行ったか又は試験を中止した時点で行われ、奏効率 33.9%（95% CI: 26.8%, 41.7%）で事前に設定した閾値 20%を超えることが検証された。

申請者によるヒストリカルデータの活用内容

申請者は C2201 試験と 3 つのヒストリカルデータ（SCHOLAR-1, the pooled CORAL 拡大試験, PIX301 試験）の間接比較を行い、申請パッケージに含めた。

SCHOLAR-1 との間接比較

SCHOLAR-1 は、DLBCL 患者を対象に救済療法の臨床アウトカムを評価することを目的として 2 つの第 3 相試験と 2 つの観察研究の患者レベルでのデータを併合したものである。本解析では、ORR、CR 及び救済療法の開始からの全生存時間（OS）を評価項目とした。なお、SCHOLAR-1 の患者集団（難治性 DLBCL）と C2201 試験の患者集団（再発又は難治性 DLBCL）で対象患者が異なっていたため、C2201 試験の難治性 DLBCL 患者での間接比較を行った。さらに、DLBCL の予後因子と考えられる初発時の診断、国際予後指標（IPI: international prognostic index）、難治性の区分に基づく傾向スコアで重み付けを行った Matching-adjusted indirect comparison (MAIC) により予後因子の偏りを調整した上での比較を行った。

CORAL 拡大試験との間接比較

CORAL 拡大試験は再発・難治性 DLBCL 患者を対象とした CORAL 試験でプロトコル計画通りに自己造血幹細胞移植（ASCT: autologous stem cell transplantation）を受けられた患者と受けられなかった患者（治療失敗）のデータを後ろ向きにレビューしたデータである。本解析では、ORR と OS を評価項目とした。CORAL 拡大試験の OS 定義は「ASCT 後の再発日又は CORAL 試験治療薬の失敗と判断した日、から理由を問わない死亡までの時間」と定義していたため、C2201 試験の OS を「最近の治療

からの再発日、最近の治療の最後の投与日又は最後に ASCT を受けた日のうち、最も遅い日から、理由を問わない死亡までの時間」と再定義して間接比較に用いた。また、CORAL 拡大試験では 2 つの前治療に失敗した患者のみ含まれていたが、患者数を確保するため、C2201 試験は前治療の数によらず間接比較に含めた。さらに、性別、国際予後指標（IPI）、ASCT 後の再発の有無別を用いて、SCHOLAR-1 との間接比較と同様に、MAIC により、予後因子の偏りの調整した上での比較を行った。

PIX301 試験との間接比較

PIX301 試験は、Pixuvri の海外薬事申請で用いられた第 3 相試験である。本解析では、ORR と OS を評価項目とした。なお、前述の 2 つの間接比較とは異なり、患者背景での対象集団のマッチングは行われていない。

EMA のヒストリカルデータ利用への見解

ヒストリカルデータとの間接比較に関して、以下の点が指摘されたが、探索目的の位置付けであることを考慮すると提出された間接的な比較は臨床試験で得られた効果を支持する追加の説明として用いることは妥当であると結論付けた。

- ベースライン変数での補正が不十分であった。
 - ・ SCHOLAR-1 のベースライン値（共変量）の収集タイミングを考えると、対象疾患である治療に対する「再発又は難治性」よりも早いタイミングでのベースライン値の可能性が高い。
 - ・ 最も比較に適していると考えられる CORAL 試験においても、C2201 試験との間で前治療数に違いがある。
 - ・ OS の Matching のための複数のアプローチでの感度分析は一定の情報を提供したが、バイアス等を除くことはできなかった。

EMA は、一定の奏効及び奏効の維持が観測されたことから、臨床ベネフィットが期待できると結論付けた。しかしながら、承認条件として、（1）更なる有効性評価のため C2201 試験の対象患者とより類似した対象で、キムリア製造から輸注までの時間の詳細なデータを含めたレジストリを用いた前向き観察試験の実施、（2）C2201 試験の長期（24 ヶ月及び 5 年フォローアップの最終解析）有効性及び安全性データの提出、（3）実施中のキムリアと既存治療薬との第 III 相比較試験（H2301 試験）の結果の提出に言及している。

2.6. BRINEURA (Cerliponase alfa)

単群試験の外部対照群をヒストリカルデータから構成し、薬事申請時の臨床データパッケージに含めた事例である。FDA から外部対照群に対して、いくつかの懸念点が示されたが、最終的に BRINEURA の有効性を示すエビデンスとして受け入れられている。

開発の経緯

BRINEURA は遺伝子組換えヒトトリペプチルペプチダーゼ 1 (hTTP1) 酸素前駆体であるセルリポナーゼアルファ (遺伝子組換え) を有効成分とする脳室内投与用の注射剤であり、セロイドリポフスチン症 2 型 (CLN2) に対する治療薬である。CLN2 はトリペプチルペプチダーゼ欠損により引き起こされる遺伝性の疾患で、主に神経系の発達に影響を及ぼし、歩行障害などを引き起こす。重篤な疾患であり一般的に 2 歳から 4 歳の幼少期に発症し、年齢を重ねるとともに多様な徴候が現れ、悪化の一途をたどる。米国では、FDA により画期的治療薬、優先審査及び希少疾病用医薬品に指定され、2017 年 4 月に主要な試験成績である第 I/II 相試験 (190-201 試験) とその継続試験 (190-202 試験) 及びレジストリを用いたヒストリカルコントロール試験 (190-901 試験) により承認された。

190-201 試験は非盲検単群用量漸増試験、190-202 試験はその継続投与試験であり、主要評価項目は CLN2 臨床評価尺度のうち、運動と言語の尺度の合計スコアである ML 尺度であった。

申請者によるヒストリカルデータの活用内容

試験立案時、申請者は対照群を設けず、ベースライン時と投与後での患者内比較を検討していたが、FDA の助言に基づき、レジストリを用いた自然経過観察研究 (190-901 試験) を実施し、臨床試験 (190-201 及び/202 試験) の外部対照群として用いることとした。190-901 試験では DEM-CHILD データベースから自然歴コホートの未治療神経セロイドリポフスチン症 2 型患者 42 例を外部対照として抽出し、190-201 試験の結果と比較した。その結果、BRINEURA 投与群では外部対照群と比較して運動機能や言語機能の低下が抑えられると結論付けた。

FDA のヒストリカルデータ利用への見解

FDA は臨床試験と自然経過観察研究の比較において、いくつかの問題点を指摘している。

- 1) 評価項目である CLN2 臨床評価尺度の評価方法の違い

申請者は主要評価項目として、CLN2 臨床評価尺度のうち、運動と言語の尺度の合計スコアを提案していたが、臨床試験と自然経過観察研究の間で CLN2 臨床評価尺度の評価方法に違いがあった（例えば、臨床試験では前向きに評価されているが、観察研究では後ろ向きに評価されている）。特に、言語尺度の評価方法の不一致は比較可能性および結果の解釈を妨げる可能性があるとして指摘し、FDA は運動尺度のみのスコアを評価することとした。

2) いくつかの患者背景の違い

臨床試験と自然経過観察研究の間で性別、年齢及び CLN2 遺伝子の変異の分布に違いが認められた。性別に関しては、有効性に影響を与える可能性は少ないと判断されたが、年齢及び CLN2 遺伝子の変異の種類については、有効性評価への影響が懸念されたため、FDA は臨床試験と自然経過観察研究の比較の際には、ベースライン時の CLN2 臨床評価尺度の運動尺度スコアに加え、年齢及び CLN2 遺伝子の変異の種類を用いたマッチングを行うことを要求した。

なお、上記の問題を議論するために、審査の間に申請者と FDA との間で多くのミーティングを行い、30 を超える情報提供が要求されていた。

2.7. プログラフ（タクロリムス水和物）

単群試験の外部対照群をヒストリカルデータから構成し、比較する計画であったが、外部対照群の症例集積が進まず、選択基準及び除外基準を変更して追加集積を実施したものの目標症例数に到達しなかった事例である。結果的に PMDA は外部対照群との比較は困難と判断したが、公表文献との比較による臨床的有用性の確認に基づき承認されている。

開発の経緯

プログラフ（タクロリムス）は脱リン酸化酵素のカルシニューリンを阻害することにより T 細胞の活性化を特異的に阻害する免疫抑制剤である。1993 年に「肝移植における拒絶反応の抑制」を適応として承認された後、他臓器への移植領域への適応だけでなく、重症筋無力症など移植以外の領域でも承認されている。

多発性筋炎・皮膚筋炎は膠原病の一つであり、原因不明の慢性炎症性疾患として厚生労働省の「特定疾患治療研究」の対象になっている難病の一つで、患者数が 1 万人前後と推定される希少疾患である。

多発性筋炎・皮膚筋炎に合併する間質性肺炎の治療薬として効能・効果を取得するため、医師主導治験が実施された。この試験デザインは、非盲検単群試験であり、本剤とステロイドを併用した（プログラフ・ステロイド併用群）。また、外部対照してステロイド単独治療が行われた治療群（ステロイド単独群）のデータを後ろ向きに収集することとした。評価はマッチングした両群間で比較することとし、有効性評価項目は OS 及び PFS として計画された。しかしながら、外部対照群のデータが計画した症例数に達しなかったため、マッチングを行った上での比較は実施されなかった。承認審査では、事前に規定した外部対照データに代わり文献データとの比較を行い、その結果に基づき承認された。

医師主導治験について

この試験は非盲検非対照の単群試験として実施され、組み入れられた被験者にはプログラフ及びステロイドが併用された。試験には 26 例が登録され、そのうち 25 例に治験薬が投与された。有効性の主要評価項目である OS は、52 週時点で 88.0% (22/25)、副次評価項目の PFS は 52 週時点で 76.4% (20/25) であった。また、OS の生存率 (88.0%) は、文献で公表されているステロイドと免疫抑制剤の併用による治療データと同程度であり、またステロイド単独治療データより優ると考えられた。

ヒストリカルデータの利用について

疾患の緊急性及び重篤性に基づく倫理的観点から並行対照群は設定されず、外部対照が設定された。医師主導治験の結果（プログラフ・ステロイド併用群）と比較するため、各治験実施医療機関において基準に合致するステロイド単独治療（ステロイド単独群）を対象に、ステロイド剤投与から 52 週間までのデータを抽出することとした。最低目標症例数は医師主導治験における本剤群の症例数の 2 倍以上を目安としたが、5 例の集積にとどまった。その後、選択除外基準を緩和するなどしたが、マッチングを実施した上での本剤群との比較に必要な症例は集まらなかった。症例集積が困難な理由の一つとしては、免疫抑制剤の併用が当初予想していた以上に近年の医療現場で一般化していることが挙げられた。

3. RWD 利用事例からの Lessons learned

前節の事例調査の検討から明らかになった RWD を利用した試験の計画・実施に関する課題について、主に統計学的観点から本タスクフォースで検討した内容をまとめた。

3.1. 事例調査から見えてきた課題

調査事例は、非盲検単群試験で得られたエビデンスを補強する目的で、閾値の妥当性の評価や外部対照群との比較を目的としてヒストリカルデータを利用したものであった。ヒストリカルデータのソースは、レジストリ、EHR、過去の臨床試験データや文献等、利用目的や必要なデータの入手可能性により様々であった。ただし、最終的に規制当局から受け入れられた事例も含め、規制当局から様々な問題点が指摘されており、薬事申請におけるヒストリカルデータの利活用の課題が明らかとなった。各事例で指摘された問題点は、それぞれの品目や対象疾患など特有の状況を踏まえた指摘であったが、それらから共通する課題を抽出すると、以下のように大別できる。

外部対照を用いる際の課題	規制当局から指摘を受けた事例
規制当局との合意形成 (事前の当局との合意等)	XPOVIO, ROZLYTREK, BRINEURA
比較可能性	XPOVIO, BALVERSA, ROZLYTREK, BRINEURA
サンプルサイズ	XPOVIO, BALVERSA, ROZLYTREK, プログラフ

Framework for FDA's Real World Evidence Program⁷等で述べられているとおり、結果の妥当性を保証するためには、試験実施前に試験デザイン及び解析方法についての透明性を高めることが不可欠である。そのためには、プロトコルおよび統計解析計画書を事前に規制当局に提出し、試験開始前に論点を議論し、双方で計画について合意することが重要となる。実際、BRINEURA の事例では、ヒストリカルデータの利用について規制当局と何度も議論を要していた。このことからヒストリカルデータを薬事申請に用いる際には、試験計画の段階で規制当局との議論なども含め、これまでの試験立案以上に時間がかかることも想定される。

比較可能性については、選択除外基準、背景因子、さらにはエンドポイントに関して、臨床試験と外部対照の間で相違が指摘された事例があった。申請者は、重要な予後因子によるマッチングを行うなど、統計解析の段階で比較可能性を高める方策を講じている事例が多かったが、解析手法のみでは限界があることが示唆された。また、可能な限り臨床試験

の選択基準及び除外基準を用いてヒストリカルデータからデータを抽出したり、重要な予後因子でのマッチングの基準に合致した患者のみを解析に用いることは、比較可能性を高めるうえで重要ではあるが、一方で予想以上に解析に含まれる症例数が少なくなってしまう、結果を解釈する上で十分なサンプルサイズには満たない事例があった。この問題を回避するためには、事前の RWD の利用可能性についてのフィージビリティ調査が重要になってくるだろう。

RWD を利用する以上、いくら事前に検討しても回避できない問題も含まれている。一方で、選択除外基準や主要評価項目等を臨床試験とヒストリカルデータで可能な限り揃えたり、様々な理由で生じるバイアスへの対応を事前に考慮するなど、比較可能性を高める取り組みは、承認審査の予見性を高める上でも重要である。今後、ヒストリカルデータを薬事申請に利用する場合には、規制当局と RWD の活用について事前に協議を行い、これまでの事例から経験してわかった課題に向けた方策を規制当局と一緒に考えていくことが重要である。

事例調査の結果を踏まえると、RWD を外部対照に用いる際には特に下記に挙げた点について十分に検討すべきであると考え、各項目の詳細については、3.2 節から 3.6 節で説明する。

【RWD を外部対照に用いる際の検討事項】

1. 比較可能性の問題（3.2 節）
 - a) 選択基準及び除外基準
 - b) 交絡
 - c) エンドポイント
2. サンプルサイズ（3.3 節）
3. 規制当局とのコミュニケーション（3.4 節）
4. 利用する RWD の調査（3.5 節）
5. 結果の透明性の確保（3.6 節）
 - a) Objective Study Design

3.2. 比較可能性の問題

RCT の同時対照を代替する目的で RWD を外部対照として利用する場合は、まず、群間比較における比較可能性を担保する必要がある。RWD を用いた検討では、結果の一般化可能性に興味及ぶことがあるが、これは比較可能性が担保されている前提での議論であるため、まずは比較可能性を高める事を優先して検討する必要がある点に注意する。

RWD を利用する試験により、RCT を実施した場合に得られる結果と同程度の結果を得ようとする場合、RCT を実施した場合に想定される試験を RWD で模倣する概念である「Target trial emulation」の考え方²⁸が参考となる。

3.2.1. 選択基準および除外基準

BALVERSA の事例では、治験参加患者に対しては厳格な除外基準が設けられていたが、RWD から得られた外部対照群は治験参加者と比較してより予後が良い可能性が高いと FDA より指摘を受けた。また、XPOVIO の事例では STORM 試験と FHAD から抽出した外部対照群の選択・除外基準が異なっていたことから、患者集団の背景に差異がある可能性が高いと FDA より指摘を受けた。

選択・除外基準については、RWD を利用する試験を計画する際、規定する選択・除外基準が RWD のデータソースにも適用できるかどうか、また適用した場合にどの程度の例数が基準を満たすかは事前に検討することは可能と思われる。また、その検討結果を事前に規制当局と協議し合意しておくことで、事前に RWD 利用についての予見可能性を高めることは有用だと考える。データソースの構築目的は様々であり、データの二次利用の前提では、臨床試験と同じ選択・除外基準を RWD に適用しようとする、対応するデータソースが存在しない、もしくは存在したとしても、症例数が限られるといったような問題が生じる可能性もあり、治験と同様の基準の当てはめは現実的ではないかもしれない。一方、選択・除外基準は、estimand²⁹と密接な関係があり、安易に RWD に適用可能な基準に合わせるという方策も慎重になるべきである。試験の estimand に基づく選択・除外基準と RWD で特定可能な選択・除外基準を比較した上で、その落としどころを規制当局と事前協議することが重要である。また、前向きなデータソースを構築することが出来れば、ある程度同じような選択・除外基準を設定できる可能性もある。

3.2.2. 交絡

BALVERSA では、測定されていない交絡因子及び主要な交絡因子の重大な欠測（対照群の 30% で ECOG-PS が欠測）があった。また、XPOVIO では、患者背景が異なることに加え、ECOG-PS 等の重要な予後因子が欠測している患者がいた。申請者は傾向スコアを用いてマッチングや IPTW による解析を実施したが、傾向スコアを推定するための予後因子が十分でなかったこと、症例数が十分でないことなどにより、統計的モデルに基づく調整には限界があると FDA から判断された。背景因子の分布に差異が見られた場合には、傾向スコアによるマッチング等を行うことで背景因子を揃えることはできるが、重要な予後因子の欠測の問題などを考慮し、計画時に重要な予後因子の特定とそれらのデータがどの程度収集できるかを事前に確認すべきである。

交絡の調整のために収集する重要な予後因子には、主要な解析で用いる背景因子だけでなく、可能であれば、主要解析に対する感度分析や補足的解析²⁹での使用も踏まえて収集する因子を決めることも有用と思われる。

傾向スコアを用いた方法など、解析時に交絡因子の影響を調整することが可能な手法は存在するものの、未知・未測定の影響因子が存在しないという条件を満たすことが前提となる。ランダム化を行わない状況において、未知・未測定の影響因子が存在しないという状況は非常に強い仮定となる。調整すべき交絡因子の特定方法については、バックドア基準がよく知られているが、その他の考え方もある。その一例として、曝露の原因となる因子で調整する、アウトカムの原因となる因子で調整する、または曝露とアウトカムの両方の原因となる因子で調整する（ただし、操作変数は除外するが、曝露とアウトカムの共通の原因となる未測定変数のプロキシ変数は含める）といった提案³⁰もされているように、交絡因子の適切な特定については複数の立場があることにも留意が必要である。したがって、傾向スコアを用いた解析では、調整しきれないバイアス（残差交絡）が存在するかもしれないことは認識しておく必要があり、感度分析等を通して、未知・未測定の影響因子の影響の程度を評価することも重要である。

3.2.3. エンドポイント

XPOVIO で OS の起点について不死時間バイアス（immortal time bias）が生じていた。ROZLYTREK では、一般的に受け入れ可能な評価項目をエンドポイントと設定したものの、RWD からそれらを収集することが困難であった。BRINEURA では、臨床試験データと自然経過データの間で、有効性の指標である CLN2 臨床評価尺度が異なる（グレードの違い、評価時期の違い、トレーニング有無、前向き・後ろ向き）という問題があった。非 RCT では RCT と比較して様々なバイアスが知られている。これらを事前に理解しておくことで、これらの事例のような課題を予め試験立案時に特定し、対処法を検討しておくことが望ましい。しかしながら試験デザインによってすべてのバイアスが回避できるわけではなく、統計的なアプローチによってバイアスを調整する場合もある。例えば、不死時間バイアスの回避については、ランドマーク解析³¹などが提案されている。このようにエンドポイントについても臨床試験と RWD で同じエンドポイントが設定できるかどうか事前の調査が必要であり、同一のエンドポイントが定義できない場合は、事前に規制当局に相談すべきである。

3.3. サンプルサイズ

XPOVIO, BALVERSA, ROZLYTREK, プログラフにおいて、サンプルサイズ不足が指摘された。XPOVIO では、患者背景が異なっていたことから、傾向スコアマッチングや IPTW による解析を実施したが、症例数が十分でないという問題があった。プログラフで

は、最低目標症例数は医師主導治験における本剤群の症例数の2倍以上を目安としたが5例の集積にとどまった。その後、選択・除外基準の緩和などの対策を講じたが、マッチングを実施した上での本剤群との比較に必要な症例は集まらなかった。症例集積が困難な理由の一つとしては、免疫抑制剤の併用が当初予想していた以上に近年の医療現場で一般化していることが挙げられた。BALVERSAは、対照群のサンプルサイズが非常に小さかった（ $n = 16 \sim 26$ ）。ROZLYTREKでは、選択・除外基準には大きな違いがあり、ベースライン時の集団の特徴に違いがあったが、サンプルサイズが小さいために統計的モデリングにおける共変量の調整が不十分であり、試験群間の治療成績（TTD, PFS, OS）の群間差が、治療効果の違いなのか、もしくは試験集団の違い（すなわち選択バイアス）に起因するのかを判断することは困難であるとコメントされている。結果的にサンプルサイズが小さく十分な検討ができなくなる事態を避けるために、事前にRWDの利用可能性の調査を実施し想定サンプルサイズが確保できるか確認する必要がある。また、想定サンプルサイズについて規制当局にその妥当性を確認することは有用かもしれない。

3.4. 規制当局とのコミュニケーション

規制当局とのコミュニケーションで挙げた課題として、事前にRWDの利用について当局と合意していなかったという事前規定の問題だけではないことがわかった。BRINEURAの事例では、事前にRWDの利用が合意できていても、その後に明らかになった課題に対応するため、当局との議論に時間を要することになり、追加解析や追加データの提供にも多くのリソースを必要としていた。このことはRWDを利用する試験計画において多くの不確実な要因が含まれる中、どこまでプロトコルや統計解析計画書に事前に規定し、その内容を当局と合意するかという問題に加え、事前にすべての問題をクリアにできない限界のため、その後に生じる課題についても規制当局と協議して対応する必要性を示唆していると思われる。つまり、RWDの利用によって臨床試験実施に係るリソースが低減されることが考えられる一方で計画時さらには評価または審査時に規制当局とのやり取りに費やすリソースは決して少なくない可能性があることを十分に留意しておいた方が良いと考える。

3.5. 利用するRWDの調査

RWDを利用した臨床試験を計画する際、多くの課題、限界が生じるが、これらを事前に把握し対処法を検討し、それを規制当局と事前に共有し議論しておくことはRWD利用の予見性を高める上で重要である。そのためには、従来のフィージビリティ調査の対象を広げて検討しておくことが有用となる。

試験立案時は、市場調査や患者数など、試験立案に必要な情報を収集することが一般的である。それに加え、RWDを外部対照として用いる際には、上述したバイアスやその他の問題点を事前に把握しておくことが重要であるため、試験立案時に検討しておくべき事項

は通常の RCT に比べて多くなる可能性がある。それらの事項を含め、試験立案時に必要な情報や調査、その検討などの一連のプロセスを本報告書ではフィージビリティ調査と呼ぶこととする。RWD を利用した臨床試験を実施する際のバイアス（Appendix-表 3）に関連する検討もこのフィージビリティ調査に含めることが望ましい。例えば、フィージビリティ調査を通して潜在的なバイアスを可能な限り特定し、事前にその対処方法や限界を把握しておくことが重要である。試験目的、選択したデータソースによって、生じるバイアスの種類、対処方法、限界が異なると思われるが、事前に把握したバイアス、対処方法、その限界について事前に規制当局と協議しておくことは試験結果の受け入れに対する予見性を高めることにつながると考える。EMA のガイドラインでは、フィージビリティ調査に含まれる項目として以下が挙げられている⁵。

- ・ レジストリ又はレジストリネットワークに関する概要
- ・ 交絡因子の利用可能性、追加データが必要な場合の収集できる可能性の調査
- ・ 利用可能なデータの質と完全性、欠測データの情報と補完方法、その他の確認結果の調査
- ・ 有害事象の特定、解析、報告のためのプロセス
- ・ レジストリの数、レジストリに登録されている患者の数、月／年あたりの新規患者数、フォローアップ期間、追跡不能に関する調査
- ・ 潜在的な情報バイアス、選択バイアス、時間バイアスの調査
- ・ 潜在的な交絡バイアス
- ・ データの特徴、試験デザインに基づく解析的な問題の調査
- ・ データシェアに関連した個人情報の扱い
- ・ レジストリ試験実施の適切性の総合的な評価

フィージビリティ調査結果も踏まえ、プロトコルを作成することが重要である。プロトコルでは、外部ソースを用いる際の試験規模感（症例数、観察期間）を見積もっておくべきである。事例調査（2 節）の中でも、RWD から選択・除外基準を満たす患者に絞り、その上でさらに統計的なアプローチによる交絡因子の調整を行うことで解析に用いる被験者数がさらに限定され、結果としてサンプルサイズが不十分なために解析結果に懸念が示された事例があった。このことから、RWD の患者数と実際に解析に含まれる患者数には乖離があることが一般的であり、その乖離の程度を事前に調査した上で、利用する RWD の特定や試験デザイン、プロトコルや解析計画書を策定することが重要である。

中間事象によるバイアスへの対応については、ICH E9(R1)²⁹ から新たな概念として導入される estimand の観点からバイアスを制御することが重要である。例えば、関心のある治療効果を知りたい対象集団について、RCT のような同時対照の場合は、関心のある同一の

母集団から被験薬群と対照群がランダムに選ばれるが、ランダム化を用いない場合は被験薬群と対照群が試験目的に適した同一の母集団であるかどうかについて注意して対照群を設定する必要がある。また、試験データと外部対照データの間で中間事象を同程度の基準で特定できるのかどうか、および頻度の違いを見ておくことは重要である。

統計解析計画についてもプロトコルと同様に事前に規定しておくことは、事後解析によるバイアスを回避するためにも重要である。また、統計解析計画には外部ソースからデータを抽出する条件も含めることが望ましい³²。試験目的に対応した estimand を定めることになるが、外部対照を用いる際は特にフィージビリティ調査で知り得た情報、データの限界等を踏まえ、必要な統計的な仮定を明示する必要がある、主たる estimand の推定値に対する頑健性についての感度分析を規定することも重要である。また、外部ソースのデータを使う場合、RWD の評価項目の測定時点が臨床試験データの測定時点と合わない場合や、臨床試験の中で収集されるデータに比べると欠測の頻度が多くなることが想定される。事前に欠測の頻度を把握したり、欠測を回避するための手順を規定したりとしても欠測を完全に防ぐことはできない。したがって、欠測に対する対応や必要な仮定についても事前に規定する必要がある。感度分析や補足的解析に必要なデータの入手可能性についても事前に考慮できれば、解析計画を検討する際に有用である。

3.6. 結果の透明性の担保

通常の RCT では、試験開始前に作成したプロトコルに基づき試験データが収集されたのちにデータベースが固定され、データベース固定に先立ち作成されていた統計解析計画書に基づき解析された試験結果が得られる。この一連のプロセスは通常の薬事申請を目的とした薬剤開発ではすでに一般的に実践されており、そのプロセスの妥当性は承認審査の過程で確認される。このようなプロセスに適切に則っている以上は、申請者の都合で好ましい結果を意図的に作り上げることが不可能であり、いわゆる「結果のいいとこどり」による懸念を考える必要はない。一方、既に存在する RWD を用いる際には、上述したような RCT 計画時に行われている前向きなプロセスを実践することが困難となる。これは、データベース固定という概念が RWD には存在しないことから明らかである。言い換えると、RWD を用いた検討では、常に都合の良い結果を提示しているわけではないことを何らかの形で積極的に示していく必要が生じる。結果の透明性の担保のためには、データの抽出手順の詳細（例えば、index date や共変量となる変数を特定するための細かな定義）を記録に残し、第三者による試験の質の評価や結果の再現が可能となる情報を公開することも望ましい。その際、RWD を用いた研究における計画および結果の報告に関するテンプレートも提案されているので参照されたい³³。

3.6.1. Objective Study Design

ランダム化を用いない試験では様々なバイアスが生じる一方で、試験デザインの工夫や統計学的なアプローチによってバイアスへの対処が一部可能となることがある。傾向スコアを用いた検討はその一例である。傾向スコアは測定された重要な交絡因子を用いて推定される。推定された傾向スコアを用いてマッチングやモデル解析、層別解析や重み付け解析などが可能であり、応用範囲が広い。ここで注意すべき点としては傾向スコアの推定に用いるモデルが異なると、得られる傾向スコアの値が異なることがあるため、傾向スコアを用いたマッチングをした際には解析対象となる被験者に違いが生じ、推定される治療効果も変わってくることが挙げられる。また、傾向スコアの推定に用いるデータセットに興味ある評価項目のデータが含まれている場合、統計解析担当者は意図的または恣意的に解析対象集団をコントロールすることが可能であるため、都合の良い結果を導くことが可能となり、解析結果の客観性や信頼性が大きく損なわれることとなる³²。

これに対処するため、試験デザインの検討を2つのステージに分けて行う方法（2-Stage Design）を提案されている³⁴。

最初のデザインステージ（第1デザインステージ）は試験開始前のプロトコル立案時に実施される。第1デザインステージでは、主に外部対照群の選定やサンプルサイズの設定に加えて、それまでに得られている臨床的知見も踏まえ、試験で収集する重要な交絡因子となる共変量の特定を行う。さらには傾向スコアの算出方法および共変量のバランスを評価する基準なども規定しておく。また、ここで重要なことは、独立した統計解析担当者を決めておくことである。この独立した統計解析担当者は、次のステージで興味ある評価項目データにアクセスすることなく、第1デザインステージで規定した方法に従い傾向スコアの推定を行う。

後半のデザインステージ（第2デザインステージ）は、すべての被験者の組み入れが終了し、傾向スコアの推定に必要なベースライン時の共変量のデータの収集が完了した時点で実施する。この後半のデザインステージでは、第1デザインステージで規定した手順に従い、評価項目のデータにアクセスできない独立した統計解析担当者が入手した共変量データを用いて、傾向スコアを推定し、マッチングや層別などを行い、試験治療群と外部対照群での共変量のバランスを評価する。また、必要に応じて、サンプルサイズや検出力等も再検討し、統計解析計画書を最終化する。

上記のプロセスを実践することにより、恣意的に傾向スコアが算出されることを防ぐことができ、傾向スコアを用いたアプローチによって推定された治療効果の客観性、信頼性が

担保される。以下に試験デザインの検討プロセスを2つのステージに分ける Objective Study Design についてまとめた（表 2 および図 1）。

表 2 Objective Study Design における 2つのデザインステージ

Stage	検討事項
第 1 デザインステージ	1) 対象集団，主要評価項目，サンプルサイズの検討 2) 外部データソースの選定 3) 第 2 デザインステージでの統計解析担当者の指名 4) 選択バイアスを防ぐための計画 - 統計的な方法論 - Firewall/Procedure の設定
第 2 デザインステージ	第 1 デザインステージで指名された独立した統計解析担当者が興味ある評価項目データにアクセスしないまま，第 1 デザインステージで規定された解析デザインを実践*する *第 1 デザインステージの統計解析担当者は解析を担当しない

第 1 デザインステージの 1) は通常の臨床試験と同様である。2) の外部データソースの選定では，測定方法，評価方法，時間の違い（temporal differences）などの統計的な手法による調整が難しいバイアスをできる限り排除できるようにオペレーションや臨床的な観点も踏まえデータソースを選ぶべきである。また，4) の選択バイアスを防ぐための計画には 2つのポイントがある。1つ目は，統計的な方法論の規定であり，解析に用いる共変量の特定，群間の共変量のバランスを調整するための統計手法，評価方法，またその基準を規定しておく。共変量のバランスを評価する方法については，いくつかのアプローチが存在し，各共変量について視覚的に判断する方法や検定によって判断する方法や standardized (absolute) mean difference に基づいて判断する方法などがある。この基準については，比較的よく用いられる基準として，0.1 から 0.25 のレンジがある³²。2つ目は，firewall/procedure の設定である。外部データを利用して単群試験の結果と比較を行う場合は，バイアスを最小限に抑える観点から，統計解析担当者は主要評価項目を含む臨床アウトカムにアクセスすべきではない。したがって，興味のある評価項目のデータにアクセスできない独立した統計担当者を決めておく必要があり，そのための仕組み（firewall/procedure）についても試験開始前までに決めておくことが重要となる。

第 2 デザインステージでは，前のステージで指定された独立した統計解析担当者が解析を実践する。また，事前規定及び透明性の観点から，統計解析計画書に加え，フィージビリティ調査結果の報告書の提出⁵，さらには第 2 デザインステージで実施した解析内容等を

規制当局から提出を求められることを想定しておいた方が良いかもしれない。このレポートに含まれる可能性がある内容としては以下が考えられる。

- ・ 交絡の調整に用いた因子
- ・ データ数（症例数）のフローチャート
- ・ 事前に規定した交絡の調整方法や判断基準とその結果
- ・ 効果の推定に用いた解析方法とその結果
- ・ 必要な仮定と感度分析の結果
- ・ 統計解析の手順書

外部データを用いた非無作為化による比較試験の場合

	First Stage	Second Stage
統計解析担当者 A	■プロコルの統計解析パートを作成 対象集団、主要評価項目、症例数設計（外部データの例数も含む） ■SAPドラフトを作成 選択バイアスを減らすアプローチ ・プロペンシティスコアを用いた方法論など ・マッチング、層別、IWTP、診断方法や基準など ・複数の方法論を探索的に実施する場合、デジジョントツリーなど ■独立した統計解析担当者 (B) を指名 ■SAP固定時期を規定	■固定したSAPに従い解析
統計解析担当者 B		■LPI後、患者背景のデータのみ入手（outcomeデータは入手しない） ■SAPドラフトに従いプロペンシティスコアの解析を行う ■規定された時期までにSAPを固定する ■検討結果のレポート作成
	試験立ち上げ	FPI LPI DBL

図 1 外部対照利用時の Objective study design

3.7. 事例から得られた留意点のまとめ

薬事申請において RWD の利用は進んでいるが、今回調査した RWD を外部対照として利用した事例では、外部対照群の適切性について指摘され、薬事申請における意思決定のエビデンスとして受け入れられなかった事例が多かった。一部、統計学的観点からの指摘も示されていたが、多くは選択基準／除外基準の相違および重要な予後因子の欠測による比較可能性への懸念、RWD から抽出した評価項目の定義、サンプルサイズの問題など、利用した RWD に起因する試験デザイン自体への指摘であった。また、RWD を外部対照として用いる計画を事前に規制当局に提出していなかったため、試験結果の妥当性が判断されなかった事例もあった。これらの課題を解決するには、事前にフィージビリティ調査を実施して、利用する外部データの適切性を十分に検討する必要がある。先述のとおり、RWD を外部対照として利用する際には、通常の RCT と比較してより様々な観点から議論すべきポイントが存在する。本邦においては、レジストリに限定されてはいるものの、薬

事承認におけるガイダンスが発出され、各種相談枠も利用可能となっている。事例調査からも研究計画の事前規定の重要性は各規制当局で一貫しており、必要に応じて規制当局と相談のうえ、研究計画を作成し規制当局と認識をすり合わせておくことが重要である。フイージビリティ調査を入念に実施したとしても、研究開始時には想定していなかった計画変更が生じる可能性は否定できないが、このような場合でも規制当局と綿密に連携を取ることで、試験の完全性と結果の透明性の担保につながるかもしれない。

RWD を同時対照の代替として用いる場合は、「RWD を利用しなかった場合に実施していたであろう RCT (target trial)」を特定することが重要である。この target trial を可能な限り模倣しようとすることは、利用する外部データの選択時に参考となる。このような”Target trial emulation”の考え方²⁸は、RWD を用いた研究の計画段階だけでなく、結果の解釈に至るまで、広範にわたり有用となる可能性がある。Target trial を定義することで、ICH E9(R1)における estimand のフレームワークを取り入れる事が容易となり、規制当局とのコミュニケーションや感度分析の切り口、適切な結果の解釈に当たっての指針を示してくれるかもしれない。

RWD を用いた検討では、結果の透明性を可能な限り高める事が重要となる。意図的または恣意的な解析によって得られた結果は、統計学的には α エラーの上昇として解釈することができ、これは規制当局が特に重要視する観点のひとつである。RWD の利用に当たっては、RCT とは対照的に計画段階から α エラーの上昇につながるリスクがあることに注意が必要である。例えば、利用する外部データに興味のある評価項目の情報が含まれている場合は、申請者にとって不都合となるデータを除くことが可能であり、そもそもデータソースの選択時にも申請者にとって都合の良い結果を導くデータを選択することが可能となることから明らかである。このような懸念を防ぐためには、プロトコルや統計解析計画書の最終化段階まで関係者は興味のある評価項目のデータにアクセスすべきではない。

3.6.1 節で紹介した Objective Study Design のように、評価項目のデータへのアクセスを制限した下でデータソースの選択や解析計画の策定の担当者と解析業務の担当者を独立させるなど、結果の透明性を担保するプロセスは有用であろう。また、適切な外部ベンダーに業務を委託することも一案かもしれない。

4. おわりに

RWD の利用はデータの二次利用としての側面が強く影響するが、RCT DUPLICATE³⁵のように、RWD から RCT の結果が再現可能かという観点からの研究も行われており、RWD の利用可能性についての検討が進められている。また、治験の枠組みにおいても、RWD をランダム化比較試験の同時対照群の一部として組み入れる Hybrid control design

が検証試験の試験デザインとして海外規制当局に受け入れられた事例^{36,37}も存在し、RWDの承認申請での利活用は今後も進んでいくと思われる。

開発期間や市販後も考慮に入れた医薬品のライフサイクル全体を考えた際に、観察研究や、今回の対象のようなRWDを外部対照とした試験の結果は、それぞれのエビデンスレベルは異なるものの、当該医薬品に関する情報を与えるという点で一定の意味はあると思われる。しかしながら、薬事申請においてRWDを用いた試験結果のみに基づいて承認可否を下せるかどうかについてはステークホルダーによって様々な論点が生じる可能性があることに注意が必要である。

ランダム化は未知の交絡因子の偏りも含めて、患者背景の分布を揃えることで比較可能性が担保される最も強力な方法であり、科学的に確立された試験デザインである。したがって、ランダム化を実施し、適切に臨床試験が遂行された下で取得されたデータから導かれる結果は、因果的な治療効果としての解釈が可能となる。一方、ランダム化を伴わない試験においては、群間の比較可能性が担保されないため、交絡の制御が極めて重要となる。観察研究や疫学研究では、交絡への対処が重要な研究領域のひとつであり、交絡が存在するデータから、因果的な解釈を可能にする方法論の研究が進んでいる。因果推論的手法を用いることで比較可能性を一定程度高める事が可能になるが、これら手法には観測されたデータからは検証不能な仮定を含むため、これら仮定が結果に及ぼす影響を検討すること（感度分析）が重要である。また、交絡を調整する際に、未測定の影響因子がないという仮定が必要であるが、この仮定が満たされない場合も現実的にはあり得るだろう。この場合、未測定の影響因子があるものの、その影響は小さいと想定される場合や多少の残差交絡が存在した下でも推定された治療効果が大きいと判断できる場合は薬事申請において有効性を裏付ける結論が導けるかもしれない。このように、試験目的、試験デザイン、RWDを含めたデータに応じた統計的なアプローチの選択ができるように統計的因果推論に関する十分な理解に基づく適用が必要となる。

今後もRCTが薬事申請におけるゴールドスタンダードとなり続けられると思われるが、一方でRWDの活用は今後の医薬品開発において重要な役割を持つことが期待される。RWDの利活用に関する議論が近年活発に進んでいるが、RWDの活用をより進めるためには活用場面について更なる論点整理が必要である。これまで述べてきたように、ランダム化を伴わない以上、結果の解釈は議論となりうる。そのため、RWDの利用が合理的であるかについて事前に精査し、患者、規制当局、申請者など様々なステークホルダーに対して共通認識をもって試験を開始する必要がある。申請者の立場としては、治験における同時対照群に組み入れる症例数をRWDで代替できる点ではコストやリソースの削減となり、開発の効率化につながる期待があるが、解釈が難しい結果により承認審査時の議論が難しく

なる可能性は否定できない。仮に適切に RWD を活用した試験を実施したとしても、審査期間が治験データのみから成る従来型の申請パッケージと比べて短縮できるかは現状では不透明であり、RWD の利用が開発計画全体に与える影響に留意すべきである。重要なことは、RCT を実施する場合であっても、RWD を外部対照とする単群試験を実施する場合であっても、それぞれにベネフィットとリスクが存在するということである。それらについて、様々なステークホルダーが十分に議論・理解した上で、当該薬剤の開発において、どのような試験が適しているのかを検討することが必要である。また、RCT や観察研究、RWD を外部対照とした単群試験から得られるエビデンスは相互に補完しあう関係にあるということも認識しておく必要がある。

RWD の活用が有用と考えられる疾患領域としては、希少疾患など RCT の実施が困難と思われる場合が想定される。今回の事例調査からも、希少疾患での RWD の活用について、倫理的観点からの正当化は一つの判断基準となりうるということが推察された。ただし、すべての希少疾患の開発に当てはまることではなく、あくまで規制当局が外部対照として RWD を受け入れるか否かについては、個別の品目で慎重な議論が必要となるというのが現状であろう。これまで希少疾患等で単群試験の結果に公表文献などの既存情報を付加することで承認されている薬剤も多く存在すること、上述したようなランダム化を伴わない試験の限界などを踏まえると、「希少疾患＝RWD を外部対照として用いる」と安易に考えるのではなく、当該薬剤の開発において RWD を単群試験の外部対照として設定する意義について、十分議論すべきである。では、その他の状況ではどのような場合に RWD の利活用につながるのでしょうか。現時点ではコンセンサスは取れていないものの、例えば希少なイベントを観察するために大規模な試験の実施が必要となる場合は、同時対照群の例数を削減するために RWD で代替することは様々なステークホルダーの立場からも受け入れ可能かもしれない。今後は、RWD が活用できる状況がより明確になることでその利用可能性が高まることが期待される。

本報告書の調査でも明らかになったように、RWD の利用に当たってはデータそのものの特性による指摘が多く見受けられた。これは、薬事申請を目的とした時に、RWD がその特性・品質として不十分であったことが原因の一つとして考えられた。RWD の利用がデータの二次利用である限り、薬事申請の目的においては治験と同等のデータではないことはむしろ自然である。薬事申請で RWD を使用していくためには、このような条件の下で薬事申請における意思決定に足るレベルを明確にすることや、RWD 自体のデータをより治験レベルに近づける試み（例：異なるデータソース間でのリンケージ）などが有用であろう。これらについては今後の産官学での議論が必要である。

RWD の薬事申請での外部対照としての利活用に当たっては、RWD の性質や統計学的観点から様々な論点が存在することが明らかとなってきた。申請者は RWD の利活用に当たってこれらの留意点を十分理解し、その適切性について検討することが重要である。RWD の利活用の検討は未だ道半ばであり、様々なステークホルダーを巻き込んだ更なる議論が必要であろう。本報告書が今後のより効率的な医薬品開発に寄与できれば幸いである。

参考文献

1. 21st Century Cures Act. <https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf> (Accessed May 25, 2022)
2. Corrigan-Curay J, Sacks L, Woodcock J. (2018). Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA; 320(9):867-8.
3. ICH. (2000). Choice of control group and related issues in clinical trials E10. https://database.ich.org/sites/default/files/E10_Guideline.pdf (Accessed May 25, 2022)
4. ICH. (2021). General considerations for clinical studies E8(R1). https://database.ich.org/sites/default/files/E8-R1_Guideline_Step4_2022_0204%20%281%29.pdf (Accessed May 25, 2022)
5. European Medicines Agency. (2020). Guideline on registry-based studies Draft. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf (Accessed May 25, 2022)
6. US Food and Drug Administration. (2017). Use of real-world evidence to support regulatory decision-making for medical devices. <https://www.fda.gov/media/99447/download> (Accessed May 25, 2022)
7. US Food and Drug Administration. (2018). Framework for FDA's Real World Evidence Program. <https://www.fda.gov/media/120060/download>. (Accessed May 25, 2022).
8. US Food and Drug Administration. (2019). Submitting documents using real-world data and real-world evidence to FDA for drugs and biologics (Draft guidance). <https://www.fda.gov/media/124795/download> (Accessed May 25, 2022)
9. US Food and Drug Administration. (2021). Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products (Draft guidance). <https://www.fda.gov/media/152503/download> (Accessed May 25, 2022)
10. US Food and Drug Administration. (2021). Real-World Data: Assessing registries to support regulatory decision-making for drug and biological products (Draft guidance). <https://www.fda.gov/media/154449/download> (Accessed May 25, 2022)
11. US Food and Drug Administration. (2021). Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological

- products (Draft guidance). <https://www.fda.gov/media/154714/download> (Accessed May 25, 2022)
12. 厚生労働省. (2021). 「承認申請等におけるレジストリの活用に関する基本的考え方」について. (薬生薬審発 0323 第 1 号・薬生機審発 0323 第 1 号), 令和 3 年 3 月 23 日.
 13. 厚生労働省. (2021). 「レジストリデータを承認申請等に利用する場合の信頼性担保のための留意点」について. (薬生薬審発 0323 第 2 号・薬生機審発 0323 第 2 号), 令和 3 年 3 月 23 日.
 14. 日本製薬工業協会. (2019). 医薬品評価委員会 臨床評価部会 タスクフォース 2「既存の国内リアルワールドデータを医薬品開発にどこまで活用できるか」. <https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc0000005lbf-att/rwd2.pdf> (Accessed May 25, 2022)
 15. 日本製薬工業協会. (2020). 医薬品評価委員会 臨床評価部会 タスクフォース 2「外部対照群を用いた承認申請への第一歩 ―疾患レジストリからの Tips と課題―」. https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc0000005k34-att/bd_rwd_sg1.pdf (Accessed May 25, 2022)
 16. 日本製薬工業協会. (2020). 医薬品評価委員会 臨床評価部会 タスクフォース 2「リアルワールドデータを承認申請へ ～活用促進のための提言～」. https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc0000005k34-att/bd_rwd_sg2.pdf (Accessed May 25, 2022)
 17. 日本製薬工業協会. (2020). 医薬品評価委員会 臨床評価部会 タスクフォース 2「製薬企業における RWD の活用促進に向けて ～現状, 課題, 論点整理, 将来展望～」. https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc0000005k34-att/bd_rwd_sg3.pdf (Accessed May 25, 2022)
 18. US Food and Drug Administration. (2014). MEDICAL REVIEW(S) (BLINCYTO®). https://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125557Orig1s000MedRedt.pdf (Accessed May 25, 2022)
 19. US Food and Drug Administration. (2014). STATISTICAL REVIEW(S) (BLINCYTO®). https://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125557Orig1s000StatR.pdf (Accessed May 25, 2022)
 20. US Food and Drug Administration. (2014). MULTI-DISCIPLINE REVIEW (XPOVIO®).

- https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212306Orig1s000MultidisciplineR.pdf (Accessed May 25, 2022)
21. US Food and Drug Administration. (2019). MULTI-DISCIPLINE REVIEW (BALVERSA®).
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212018Orig1s000MultidisciplineR.pdf (Accessed May 25, 2022)
22. US Food and Drug Administration. (2019). MULTI-DISCIPLINE REVIEW (ROZLYTREK®).
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212725Orig1s000,%20212726Orig1s000MultidisciplineR.pdf (Accessed May 25, 2022)
23. European Medicines Agency. (2018). Assessment Report (KYMRIAH®).
https://www.ema.europa.eu/en/documents/assessment-report/kymriah-epar-public-assessment-report_en.pdf (Accessed May 25, 2022)
24. US Food and Drug Administration. (2017). SUMMARY REVIEW (BRINEURA®).
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761052Orig1s000SumR.pdf (Accessed May 25, 2022)
25. US Food and Drug Administration. (2017). MEDICAL REVIEW(S) (BRINEURA®).
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761052Orig1s000MedR.pdf (Accessed May 25, 2022)
26. US Food and Drug Administration. (2017). STATISTICAL REVIEW(S) (BRINEURA®).
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761052Orig1s000StatR.pdf (Accessed May 25, 2022)
27. 独立行政法人医薬品医療機器総合機構. (2013). プログラフ®審査報告書.
https://www.pmda.go.jp/drugs/2013/P201300087/800126000_20500AMZ00157000_A100_2.pdf (Accessed May 25, 2022)
28. Hernán MA and Robins JM. (2016). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*; 183(8):758-64.
29. ICH. (2019). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1).
https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf (Accessed May 25, 2022)

30. VanderWeele TJ. (2019). Principles of confounder selection. *Eur J Epidemiol*; 34(3):211-9.
31. Gleiss A, Oberbauer R, Heinze G. (2018). An unjustified benefit: immortal time bias in the analysis of time-dependent events. *Transpl Int*; 31(2):125-30.
32. Yue LQ, Campbell G, Lu N, et al. (2016). Utilizing national and international registries to enhance pre-market medical device regulatory evaluation. *J Biopharm Stat*; 26(6):1136-45.
33. Wang SV, Pinheiro S, Hua W, et al. (2021). STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ*; 372:m4856.
34. Yue LQ, Lu N, Xu Y. (2014). Designing premarket observational comparative studies using existing data as controls: challenges and opportunities. *J Biopharm Stat*; 24(5):994-1010.
35. RCT DUPLICATE. <https://www.rctduplicate.org/> (Accessed May 25, 2022)
36. Schmidli H, Häring DA, Thomas M, et al. (2020). Beyond Randomized Clinical Trials: Use of External Controls. *Clin Pharmacol Ther*; 107(4):806-16.
37. Medicenna Therapeutics Corp. (2020). News Release (October 15, 2020). <https://ir.medicenna.com/node/7371/pdf> (Accessed May 25, 2022)
38. Burger HU, Gerlinger C, Harbron C, et al. (2021). The use of external controls: To what extent can it currently be recommended?. *Pharm Stat*; 20(6):1002-16.

日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2020 年度タスクフォース 1
ヒストリカルデータを用いた臨床開発プロセス革新チーム

土屋 悟*	大日本住友製薬(株)
松岡 伸篤*	ファイザーR&D 合同会社
吉田 征太郎*	中外製薬(株)
黒須 真介	EA ファーマ(株) (～2021 年 8 月)
笹川 裕次	Meiji Seika ファルマ(株)
金子 朋美	ノバルティス ファーマ(株)
李 康赫	バイオジェン・ジャパン(株)
小島 将裕	協和キリン(株)
横田 大輔	大塚製薬(株)

*推進委員

Appendix

RWD 利用時に発生しうる種々のバイアス

外部対照を用いるデザインに関連するバイアスの特徴と対応策が Burger (2021) ³⁸ らによってまとめられており、非常に参考になるため、表 3 に示す。

表 3 バイアスとその対応策

Potential bias	Mitigation strategy
Selection bias Patients enrolled in clinical trials are different in some ways compared to patients treated in clinical practice. 選択バイアス 臨床試験の患者が実臨床の患者と異なる	Prospectively set clinical trial inclusion/exclusion criteria (I/E) based on measures collected in routine clinical practice. Apply I/E to external control database. Adjust for remaining population differences using propensity scores methods and/or multivariate analyses. 実臨床で日常的に測定されている項目に基づき、選択除外基準を前向きに設定する。外部対照のデータベースに選択除外基準を適用する。残りの集団の違いは傾向スコアや多変量解析を用いて調整する。
Calendar time bias Patients treated in the past do differently than those treated today. 暦時間バイアス 過去に治療された患者と現在治療を受けた患者の違い	Taking control patients from the same calendar time as clinical trial or having evidence that standard of care and how it is used in this setting did not change Otherwise: Show data indicating acceptable variability. 臨床試験と同じ暦時間からの患者を組み入れる、又は標準治療とその用法用量に変更がなかったことの根拠を示す。もしくは受け入れ可能な変動であることを説明するデータを示す。
Regional bias Patient outcome may vary between regions. 地域バイアス 異なる地域間で患者のアウトカムは異なるかもしれない	Taking patients from the same region or having evidence that standard of care treatment and outcomes does not vary between regions Otherwise: Show data indicating acceptable

	variability 同じ地域の患者を組み入れる，又は標準治療とアウトカムが地域間で異なる根拠を示す．若しくは受け入れ可能な変動であることを説明するデータを示す．
Assessment bias Knowledge of therapy can influence the outcome assessment. 評価バイアス 治療に関する知見がアウトカムの評価に影響を与えうる	Choose objective endpoints, for example, OS, independent review (similar to strategies for open-label randomized studies). Conduct sensitivity analyses. 客観的なエンドポイントを選択する．例えば，OS や独立したレビュー（非盲検ランダム化試験と同じような対応）．感度分析を実施する．
Different endpoint bias Response and progression in clinical trials are measured differently than in routine clinical practice. (This bias not applicable to OS.) 異なるエンドポイントによるバイアス 臨床試験のレスポンスや症状の進行が実臨床とは異なる方法で測定される	Obtain consent for tumor images to allow RECIST-like review. Conduct sensitivity analyses, for example, frequency of tumor assessments. RECIST に準じた評価ができるように腫瘍画像の同意をとる．感度分析を行う（例えば，腫瘍評価の頻度）．
Immortal bias Study start is difficult to determine as an anchor for patient specific time scale Immortal バイアス 患者ごとの時間間隔に対して，評価の起点として試験開始を設定することは難しい．	Need to clearly define time origin. Evaluation of potential bias introduced 時間の起点を明確に定義する必要がある．潜在的なバイアスの評価
Retrospective selection bias Select retrospectively external data and key analysis features. 後ろ向きの選択バイアス 後ろ向きに外部データを選択し重要な解析の特徴を把握する	Pre-specification 事前に規定する

Study bias Patients in clinical trials have different outcomes than in clinical practice. (e.g., due to placebo effect, different care) 試験バイアス 臨床試験の患者は実臨床の患者とは異なるアウトカムになる（プラセボ効果や異なる治療のため）	Show data indicating acceptable variability on outcome 受け入れ可能な変動であることを説明するデータを示す.
Between study variability Reliance on external controls will be more critical if there is a high between study variability that cannot be explained, as this indicates that important factors affecting the outcome of an endpoint in a study cannot be controlled for. 試験間の違い エンドポイントのアウトカムに影響を与える重要な要因をコントロールできないために、もし試験間の変動が大きくなった場合は、外部対照の信頼性に対してより重要な問題が生じる.	In such a situation a RCT may be mandatory or at least strongly recommended. この場合は RCT が必須になるかもしれない、若しくは少なくとも RCT が推奨される.
Intercurrent event bias after study entry Intercurrent events after study entry may be correlated with outcome and jeopardize comparability with external control 試験開始後の中間事象バイアス 中間事象はアウトカムと相関して外部対照との比較可能性に問題を生じさせる可能性がある.	Same bias as for RCTs. Same methods (see ICH E9 addendum) should be applied. RCT と同じようなバイアス. 同じような方法 (ICH E9 addendum 参照) が適用される.