

# 費用対効果評価で 間接比較を実施する際の留意点

令和7年2月

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  
2024年度継続タスクフォース7

## 目次

1. はじめに .....	3
2. 2 試験間での間接比較 .....	4
3. 日本の事例調査 .....	5
3.1 トラスツズマブデルクステカン（エンハーツ：乳がん） .....	6
3.2 トラスツズマブデルクステカン（エンハーツ：胃がん） .....	7
3.3 アンデキサネットアルファ（オンデキサ） .....	8
4. 海外の事例調査 .....	9
4.1 NICE で公表されている事例 .....	9
4.1.1 ザヌブルチニブ（TA833） .....	10
4.1.2 ブリガチニブ（TA670） .....	14
4.1.3 テポチニブ（TA789） .....	17
4.2 各国の HTA 機関による間接比較の利用と受け入れ状況 .....	21
5. 解析手法とシミュレーション研究 .....	22
5.1 解析手法 .....	22
5.1.1 アンカーを有する間接比較 .....	22
5.1.2 アンカーを有さない間接比較 .....	24
5.1.3 その他 .....	24
5.2 シミュレーション研究 .....	26
6. 細胞・遺伝子治療の HTA の課題と対応 .....	27
6.1 細胞・遺伝子治療の HTA における課題 .....	27
6.2 日本での評価事例 .....	29
6.3 海外での評価事例 .....	30
7. ガイドライン（間接比較部分） .....	31
7.1 日本：中央社会保険医療協議会における費用対効果評価の分析ガイドライン 2024 年度版 <sup>1</sup> .....	32
7.2 英国：NICE HEALTH TECHNOLOGY EVALUATIONS: THE MANUAL. 2023 .....	32
7.2.1 TSD18：NICE に提出する際の母集団調整間接比較法 2016 の推奨事項 .....	33
7.3 欧州：エビデンスの統合 直接比較と間接比較 /EU COMMISSION 2024 .....	35
7.4 ドイツ：一般的方法 第 7 版 IQWiG 2023 .....	38
7.5 カナダ：医療技術の経済評価のためのガイダンス CADTH 2018 .....	38
7.6 米国：ICER による価値評価のフレームワーク 2023-2026 .....	39
8. 考察 .....	39

8.1 海外の事例に基づく考察.....	39
8.2 解析手法およびシミュレーション研究に基づく考察 .....	40
8.3 ガイドラインに基づく考察.....	40
<b>9. まとめ.....</b>	<b>41</b>

## 1. はじめに

2019年に費用対効果評価制度が本邦の薬価制度に本格導入されてから約5年が経過し、本邦においても費用対効果評価を実施した品目の事例が蓄積されつつある。それに伴い、評価の過程で一般に直面する統計的課題も浮き彫りになってきた。本邦の費用対効果評価では、費用効果分析を実施する前に、評価対象技術<sup>注1)</sup>の比較対照技術<sup>注2)</sup>に対する相対治療効果を推定し、追加的有用性の有無を評価する必要がある<sup>1)</sup>。しかしながら、この比較対照技術が必ずしも薬事申請におけるピボタル試験（一般に無作為化第3相試験）の対照群と一致するとは限らない。薬事申請と費用対効果評価ではその目的が異なるため、むしろ一致しないケースの方が多いと考えられる。また、近年はがん領域を中心に、単群第2相試験の結果に基づいて薬事承認を得る事例が増えている。この場合は一般に、費用対効果評価を行う段階において、評価対象技術と比較対照技術を比較した無作為化比較試験を利用できない。

このように、関心のある治療群間で直接比較した試験の結果を利用できない場合には、複数の試験を統合した間接比較を行って、追加的有用性の有無を評価する必要がある。しかしながら、直接比較と異なり間接比較では、関心のある治療群間の比較に無作為化を伴わないことから、単純な間接比較では一般に、評価対象技術の比較対照技術に対する相対治療効果の推定にバイアスを伴ってしまうため、何らかの調整を行う必要がある。

データサイエンス部会の費用対効果評価に関するタスクフォースでは、ここ数年 National Institute for Health and Care Excellence (NICE；国立医療技術評価機構)が公表している技術的文書 (Technical Support Document：TSD<sup>2)</sup>)を参考に、間接比較やネットワークメタアナリシス (Network Meta Analysis：NMA)に関する報告書を複数公表してきた<sup>3,4,5)</sup>。特に2022年1月には、近年適用事例が増加傾向にある Matching Adjusted Indirect Comparison (MAIC)や Simulated Treatment Comparison (STC)などの母集団調整法 (Population Adjustment Indirect Comparison：PAIC)に関する報告書「間接比較において母集団調整法は有用か？」<sup>4)</sup>を公表し、本手法の理論的側面について解説した。本手法では特に、評価対象技術の評価した試験と比較対照技術の評価した試験の2試験間で間接比較を行う状況に焦点を当てている。

PAICは理論的には適切な手法であるものの、一般に成立していると見做すことが困難な非常に強い仮定を必要とする。NICE TSD 18<sup>6)</sup>でも、本手法の適用を推奨しているわけではなく、本手法の適用には慎重を要することが述べられている。しかしながら、非常に強い仮定を要するからといって、本手法の適用は不適切と判断するのは早計である。費用対効果評価を実施する段階では、利用可能な試験が限られているケースも多く、PAICを除くと、単純な間接比較（ナイーブな間接比較<sup>1)</sup>）を実施せざるを得ない状況もありうる。従って、実用上は手法の限界を理解しながらPAICを上手く活用していくことも必要である。実際に海外では、PAICの結果を意思決定に用いた事例も多数存在しており、それらの事例から学ぶことは多いと考えられる。

そこで、本報告書では、理論的側面ではなく実用的側面に焦点を当てて、PAICの過去の適用事

注1) 評価対象技術：評価対象として選定された医薬品・医療機器・再生医療等製品

注2) 比較対照技術：評価対象技術が分析対象集団への治療として導入された時点で、臨床的に幅広く使用されており、評価対象技術によって代替されると想定されるものとする。うち、治療効果がより高いものを一つ選定することが原則的な考え方である。

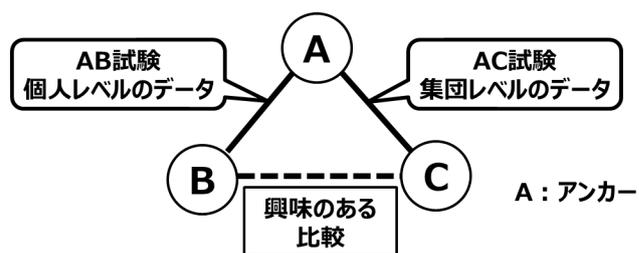
例を通じて、今後こういった場面で PAIC の結果を意思決定に利用しうるかについて考察する。2 章では本報告書で焦点を置く間接比較として、個人レベルのデータ (Individual Patient Data : IPD) を有する試験 (IPD 試験) と、集団レベルのデータ (Aggregated Data : AgD) を有する試験 (AgD 試験) の 2 試験間での間接比較について解説する。3 章では本邦において MAIC が実施された品目を調査し、そのうち 2 つの事例 (エンハーツとオンデキサ) について紹介する。4 章では海外の事例調査を行い、NICE に登録されている技術評価 (Technology Appraisal : TA) のうち、PAIC を用いた特徴的な 3 つの事例 (ザヌブルチニブ、ブリガチニブ、テポチニブ) を取り上げ、各事例における、間接比較の結果に基づく企業からの提案や委員会での議論および最終的な意思決定について紹介する。また、各国の HTA 機関による間接比較の利用と受け入れ状況について紹介する。5 章では MAIC、STC、Multi Level-Network Meta Regression (ML-NMR)、二重頑健推定量を中心に、PAIC の各解析手法の特徴について解説する。また、PAIC の性能を評価したシミュレーション研究の結果についても紹介する。なお、解析手法については過去の報告書<sup>4</sup>でも詳細に解説しているので、そちらも合わせて参照いただきたい。6 章では日本および海外での評価事例を通して、細胞・遺伝子治療における HTA の課題と実際について解説する。7 章では各国の HTA 機関が公表している分析ガイドラインの間接比較に関する記述を紹介し、記述の差異について解説する。8 章では 4 章から 7 章までの内容を踏まえ、タスクフォースによる考察 (海外事例に基づく考察、解析手法およびシミュレーション研究に基づく考察、ガイドラインに基づく考察) を述べる。9 章ではまとめについて述べる。

間接比較とひとえに言っても、実際に間接比較を実施せざるを得ない場合に直面した場合、取りうるオプションは複数あり、その中から、臨床的観点、統計的観点、医療経済的観点等を考慮して最も良いと思われる手法を選択する必要があると考えられる。このように、利用可能な情報が限られている中で、妥当な間接比較を実施するというのは非常に難しい問題であるため、本邦で間接比較を実施する実務担当者にとって、本報告書が少しでも有益な資料となれば幸いである。

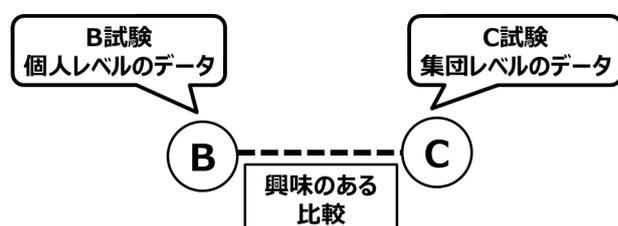
## 2. 2 試験間での間接比較

間接比較にはアンカーがある場合とアンカーがない場合の 2 つに大別される。アンカー (共通の比較対照群 : 治療 A) がある場合の最も単純な状況は、AB 試験 (治療 A と治療 B を比較した無作為化 2 群比較試験) と AC 試験 (治療 A と治療 C を比較した無作為化 2 群比較試験) を用いて、治療 B の治療 C に対する相対治療効果を推定する状況である。一方、アンカーがない場合の最も単純な状況は、B 試験 (治療 B を評価した単群試験) と C 試験 (治療 C を評価した単群試験) を用いて、治療 B の治療 C に対する相対治療効果を推定する状況である。本報告書で焦点を置く MAIC や STC は、2 試験 (AB 試験と AC 試験または B 試験と C 試験が利用可能な状況) で、一方の試験は IPD 試験、もう一方の試験は AgD 試験の場合に適用可能な手法である。このような状況は、一般に評価対象技術は自社の品目のため IPD を有し、比較対照技術は他社の品目のため、他社が実施した臨床試験や観察研究の公開論文の情報しか利用できない場合に想定される。次章以降でも基本的に IPD 試験と AgD 試験の 2 試験間での間接比較に焦点を置く。

### アンカーがある場合



### アンカーがない場合



## 3. 日本の事例調査

本邦では、費用対効果評価の制度化以後（2019年4月から現在まで）、多くの新しい治療薬や特定の疾患を対象とした革新的な医療機器がC2H（CORE 2 Health：Center for Outcomes Research and Economic Evaluation for Health；保健医療経済評価研究センター）により評価されている<sup>7</sup>。2024年3月15日時点での医薬品および医療機器の評価状況<sup>7</sup>を年度ごとにまとめた情報は以下の通りである。

年度	評価終了	評価中	分析中断	合計
2023年	0	8	0	8
2022年	8	5	0	13
2021年	10	0	0	10
2020年	4	0	1	5
2019年	5	0	0	5

これらの年度を通じて、合計50件の医薬品および医療機器が評価され、27件が評価終了、13件が評価中、そして10件がその他のステータス（分析中断の1件も含む）とされている。

評価終了となった27件のうち、NMAが実施された品目は8件、MAICが実施された品目は4件であった。各年度の詳細リストは以下の通りである。

年度	NMA 実施品目	MAIC 実施品目
2022 年	ビンゼレックス、ケレンディア	オンデキサ
2021 年	エムガルティ、レットヴィモ	ポライビー
2020 年	カボメティクス、エンハーツ、リベルサス	エンハーツ
2019 年	トリンテリックス	キムリア

以下では、MAIC が実施された品目のうち、エンハーツとオンデキサの事例を紹介する<sup>8,9</sup>。これらの情報は、C2H のウェブサイト上の公開情報のみに基づく。ただし企業が提出した資料は参照せず、C2H の報告書のみを参照している。そのため費用対効果評価専門組織の最終決定については記載しない。

なお、分析に用いられた各臨床試験の情報は公開資料を参照いただきたい。

### 3.1 トラスツズマブデルクステカン（エンハーツ：乳がん）

項目	詳細内容
分析対象集団	HER2（Human Epidermal Growth Factor Receptor type 2；ヒト上皮増殖因子受容体 2 型）陽性の手術不能または再発乳がん患者で、抗 HER2 療法を行った患者。具体的には、(a) 二次治療および (b) 三次治療以降を受けている。これらの患者は、トラスツズマブ+化学療法を受けている。
評価対象技術名	トラスツズマブデルクステカン
比較対照技術名	トラスツズマブ+体表面積等を考慮して最も安価な化学療法との組み合わせ
費用対効果の評価結果	トラスツズマブデルクステカンの有用性に関するエビデンスが不十分であるため、結論を出すには至らなかった。集団 (a) では、エビデンスが確認できなかったため、分析不能と判断され、追加的有用性の評価はされなかった。集団 (b) では、間接比較の結果、トラスツズマブデルクステカンはトラスツズマブ+化学療法に比べて進行無増悪生存期間（Progression Free Survival：PFS）および治療中止までの時間（Time to Treatment Discontinuation：TTD）が延長していることが示された。
間接比較の手法	企業分析では、アンカー無 MAIC が使用された。この手法では、効果修飾因子（相対的な治療効果を変化させる可能性のある、臨床的にアウトカムと関連する因子）と予後因子（臨床的にアウトカムと関連する因子）の分布が異なる場合でも、両試験で利用可能な共変量の分布を揃えた上でアウトカムを比較することができる。
間 接 比 較 (MAIC) 実施の 詳細	MAIC で、集団 (b) に対しトラスツズマブデルクステカン（U201 試験、第 2 相単群試験、IPD）と比較対照技術（WJOG、KBCSG、HER2CLIMB 試験、AgD）のそれぞれと比較。共変量としては年齢、ECOG（Eastern Cooperative Oncology Group）PS（Performance Status）0 の割合、T-DM1（トラスツズマブエムタン

	シン) 投与前治療ライン数 2 以上の割合、ホルモン受容体陽性の割合、ペル ツズマブ治療歴ありの割合、脳転移ありの割合が使用された。
間接比較の評価 結果に対する当 局見解	公的分析では間接比較の結果に大きな不確実性が伴い、トラスツズマブデル クステカンの治療成績を過大に推計している可能性が否定できず、分析不能 と判断された。しかし、2021 年 9 月時点で進行中の第 3 相ランダム化比較試 験 (Randomized Controlled Trial : RCT) (DESTINY-Breast02 試験) の結果が明 らかになれば、不確実性の小さいよりロバストな分析結果が得られることが 期待されていた。当局は、現時点での間接比較に基づくトラスツズマブデル クステカンの追加的有用性に大きな不確実性があると指摘し、比較対照技術 との間で選択された共変量以外に未確認の潜在的な交絡因子が存在する可能 性や、有効サンプルサイズの小ささ、PFS の評価法の違いなどが原因である とした。

### 3.2 トラスツズマブデルクステカン (エンハーツ：胃がん)

項目	詳細内容
分析対象集団	HER2 陽性 (IHC3+または IHC2+/ISH+) の治癒切除不能な進行または再発の 胃がん患者で、ニボルマブが治療薬として用いられる。
評価対象技術名	トラスツズマブデルクステカン
比較対照技術名	ニボルマブ
費用対効果の評 価結果	トラスツズマブデルクステカンとニボルマブの間で、OS、PFS、治療成功期 間 (Time to Treatment Failure : TTF) を評価。公的分析によると、トラスツズ マブデルクステカンはニボルマブに比べて OS、PFS において有効な結果を示 した。しかしながら、J202 試験 (第 2 相比較試験) の比較対照はイリノテカ ン・パクリタキセルであったが、ニボルマブの有効性と同程度と考えられ、 この試験のハザード比との間に乖離があることが示された。
間接比較の手法	企業分析では、アンカー無 MAIC が使用された。なお、公的分析は、J202 試 験と NMA の論文のデータを用い、Bucher らによる間接比較法 (Bucher 法) <sup>10</sup> にて、化学療法をアンカーとしたトラスツズマブデルクステカンとニボル マブの間接比較を実施した。
間接比較 (MAIC) 実施の詳細	MAIC の実施では、トラスツズマブデルクステカン (J202 試験、IPD) と比較 対照技術 (SCCRE、ATTRACTION-2 試験の HER2 陽性全体のデータ、AgD) のそれぞれと比較。前治療ライン数、胃切除術の治療歴、転移数、ECOG PS 1 の割合などの共変量を調整し、OS、PFS、TTF をアウトカムとして比較し た。
間接比較の評価 結果に対する当 局見解	当局は、製造販売業者による間接比較でニボルマブ群とトラスツズマブデル クステカン群の間に大きな異質性があると指摘し、ニボルマブに対するトラ スツズマブデルクステカンの OS、PFS のハザード比が J202 試験で報告され

	<p>た対イリノテカン・パクリタキセルのハザード比よりも小さいことから、ニボルマブの有効性が過小評価されている可能性がある」と評価している。さらに、HER2 発現レベルの異質性、背景因子の差、試験間の効果の測定値の比較可能性に疑問を呈しており、これらの要因が製造販売業者による間接比較結果と公的分析の再分析結果との間に差異が生じる一因である可能性がある」とされた。</p>
--	--

### 3.3 アンデキサネットアルファ（オンデキサ）

項目	詳細内容
分析対象集団	直接作用型第 Xa 因子阻害剤（アピキサパン、リバーロキサパン、またはエドキサバントシル酸塩水和物）を投与され、生命を脅かす出血または止血困難な出血が認められた患者。この集団は、アンデキサネットアルファを頭蓋内出血および消化管出血に対して異なる投与方法で治療した 4 つのサブグループに細分される。
評価対象技術名	アンデキサネットアルファ+標準的対症療法（保険適用外を除く輸血や輸液等）
比較対照技術名	標準的対症療法
費用対効果の評価結果	追加的有用性の評価において、製造販売業者による MAIC より得られた結果は妥当でないと判断した。システマティックレビューでは、アンデキサネットアルファと標準的対症療法の直接比較を行う研究が同定されなかったため、ANNEXA-4 extension 試験（アンデキサネットアルファ単群試験、IPD）と後向き観察研究である ORANGE 研究（標準的対症療法、IPD）に基づいて間接比較が行われた。
間接比較の手法	傾向スコアマッチングとアンカー無 MAIC が用いられた。
間接比較（MAIC）実施の詳細	企業分析の結果では、アンデキサネットアルファ群と標準的対症療法群の比較が行われ、交絡因子の調整により、出血部位と既往歴を含む因子が考慮された。この分析により、アンデキサネットアルファ群が標準的対症療法群に比べて統計的に有意に死亡率を低下させることが示された。
間接比較の評価結果に対する当局見解	当局は、MAIC による追加的有用性評価に不確実性があると指摘し、結果の妥当性に懸念を示している。特に、ANNEXA-4 extension 試験と ORANGE 研究間での適格基準および除外基準の相違、試験実施期間と場所・試験デザインの違いが不確実性の原因とされた。これらの理由から、MAIC による結果を直接用いるよりも、ANNEXA-4 extension 試験ではなく、ANNEXA-4 試験の IPD と、ORANGE 研究の IPD を傾向スコアマッチングにより背景情報を揃えて比較した Cohen ら <sup>11</sup> の報告の結果をそのまま参照する方がより適切であると判断された。また、Costa ら <sup>12</sup> の報告では、アンデキサネットアルファ群が標準的対症療法群に比べて統計的に有意に低い 30 日死亡率を示した

	<p>ことが確認されており、これがアンデキサネットアルファの追加的有用性のエビデンスとされた。公的分析では、特に Costa らの報告を含む複数の研究結果を評価し、アンデキサネットアルファが標準的対症療法に対して追加的な有用性を持つ可能性があるとし唆された。しかし、研究間での方法論的な違いやデザイン上の不確実性が残るため、これらの結果をどの程度一般化できるかには慎重な検討が必要である。特に重度の消化管出血に関しては、製造販売業者が行った MAIC の評価結果に不確実性があるため、現在のところ追加的有用性を確定的に評価することは困難である。</p>
--	--

## 4. 海外の事例調査

### 4.1 NICE で公表されている事例

英国では NHS (National Health Service; 国民保健サービス) の公的医療制度が用いられており、政府 (保健省) から指定された評価対象技術について NICE により費用対効果評価が実施されている<sup>13</sup>。評価対象技術の選定にあたっては、影響を受ける集団および罹患率・死亡率、費用の影響、政府による政策的な重要性などが考慮される。単一医薬品・単一適応症に対して行われる STA (Single Technology Assessment; 個別技術評価) の分析は、対象となる医療技術に対して企業から提出される評価資料を研究機関がレビューする形式で薬事承認前から開始され、最終的に評価委員会によりその医療技術使用の推奨の有無についての勧告がなされる。その後、この勧告に基づき、NHS により償還の可否が決定される。TA には臨床的有用性 (clinical effectiveness) と医療経済性 (cost effectiveness) が含まれる。臨床的有用性は定められた方法で系統的に文献を分析して評価され、医療経済性は実臨床下の患者または介護者を対象に QALY (Quality-Adjusted Life Year: 質調整生存率) を指標とした公的な医療費および福祉費の範囲における費用効果分析が実施される。

個別技術に関する TA ガイダンスでは、分析結果に関して一部非公開部分を含むものの、企業から提出された評価資料および ERG (Evidence Review Group) からの指摘事項、それらの議論に基づく評価委員会の最終的な意思決定の結果が公表されている。臨床的有用性および医療経済性の評価では、主要な臨床試験が英国の実際の臨床環境を反映していない、関心のある比較対照技術と直接比較した臨床試験が存在しない、また生存時間に関する追跡期間が不十分であるなど、各評価対象技術の開発状況を考慮して評価対象試験を適切に選択して実施する必要があり、既存試験を含む複数の試験間での臨床効果の間接比較による結果も頻繁に用いられている。

間接比較の手法としては、複数の臨床試験の AgD に基づき間接的に比較する TSD1~7 で紹介されている NMA、IPD に基づき間接的に比較する TSD17 で紹介されている傾向スコアによる調整法、一部のみ IPD を用いる間接比較の方法として TSD18 で紹介されている MAIC および STC の方法等が知られている。データサイエンス部会からの報告書「間接比較において母集団調整法は有用か？」<sup>4</sup>では近年での間接比較の適用件数の推移など、適用状況の概要が述べられており、今回、比較的新しい手法である MAIC および STC 等の方法論に基づく解析結果がどのように解釈されて臨床的有用性および医療経済性に反映されているかについて、さらに公表されている評

価技術の TA ガイドラインを調査した。

ここでは、NICE に登録されている TA のうち、特徴的な TA の事例として、主要な臨床試験が無作為化比較試験で、AgD にアクセス可能な他試験との間接比較に MAIC および STC の両方の方法を適用しているザヌブルチニブ (TA833)<sup>14</sup>、主要な臨床試験が無作為化臨床試験で、AgD にアクセス可能な他試験との間接比較にアンカーあり MAIC、アンカーなし MAIC および Bucher 法を適用しているブリガチニブ (TA670)<sup>15</sup>、主要な臨床試験が単群試験で、IPD にアクセス可能な他試験との間接比較に傾向スコアによる調整法および MAIC 法を適用しているテポチニブ (TA789)<sup>16</sup>を取り上げた。

以下では、各事例における臨床的有用性および医療経済性の評価において、間接比較の結果に基づく企業からの提案や委員会での議論および最終的な意思決定について、公開資料で記載されている内容の一部を紹介する。なお、詳細は Web 上の資料<sup>1,15,16</sup>を参照されたい。

#### 4.1.1 ザヌブルチニブ (TA833)

ザヌブルチニブは、英国で「1 つ以上の前治療を受けたことのある、または化学免疫療法が不適当な患者に対する第一選択治療を受けたことのある、ワルデンストレームマクログロブリン血症 (WM) 成人患者の治療」として 2021 年 12 月に販売承認を取得した薬剤である。

##### 4.1.1.1 臨床試験および間接比較

###### (1) 臨床試験

ザヌブルチニブの有効性は、第 3 相の BGB-3111-302 試験 (ASPEN ; NCT03053440) で評価されている。ASPEN は、再発性・難治性疾患、または治療歴がなく化学免疫療法の対象外である患者を対象とした WM の治療の有効性と安全性の評価を目的とした、進行中のザヌブルチニブとイブルチニブの第 3 相非盲検ランダム化比較試験である。この試験では、MYD88 (myeloid differentiation primary response gene 88) の変異状態による 2 つの患者コホートが設定されている。

- コホート1：MYD88変異患者を対象としたザヌブルチニブまたはイブルチニブのランダム割付による治療
- コホート2：野生型MYD88患者を対象としたザヌブルチニブによる治療

ASPEN の主目的は、MYD88 変異の WM 患者 (コホート 1 の再発・難治性群) におけるザヌブルチニブの有効性と安全性をイブルチニブと比較することであった。さらに探索目的として野生型 MYD88 WM 患者 (コホート 2) におけるザヌブルチニブの抗がん活性と安全性が評価された。

###### (2) 間接比較の適用

ASPEN 試験ではザヌブルチニブとイブルチニブの直接比較の結果が得られたが、イブルチニブは費用対効果評価における比較対照には含まれていなかった。比較対照に含まれていた治療のうち、システマティックレビューで特定されたベンダムスチンとリツキシマブ併用 (BR) の単群試験、デキサメタゾン、リツキシマブおよびシクロホスファミドの 3 剤併用 (DRC) の単群試験

については、それぞれ試験結果の AgD が入手可能であったが、ザヌブルチニブと BR および DRC の比較に利用できる共通の対照治療は存在しないため、アンカーを使用した間接治療比較の方法は適用できなかった。そのため、これらの治療を受けた患者の背景を一致させるため、ASPEN におけるザヌブルチニブ群の IPD を重みづけして、BR および DRC により治療された別の集団とのマッチング調整間接比較 (MAIC) が適用された。なお、MAIC は 3 通りの 2 治療間に適用された。それらは、ザヌブルチニブが処方された再発・難治性疾患患者および治療未経験患者と BR が処方された再発・難治性疾患患者、ザヌブルチニブが処方された再発・難治性疾患患者と BR が処方された再発・難治性疾患患者、ザヌブルチニブが処方された再発・難治性疾患患者および治療未経験患者と DRC が処方された治療未経験患者の各 2 治療間比較であった (図 4.1)。

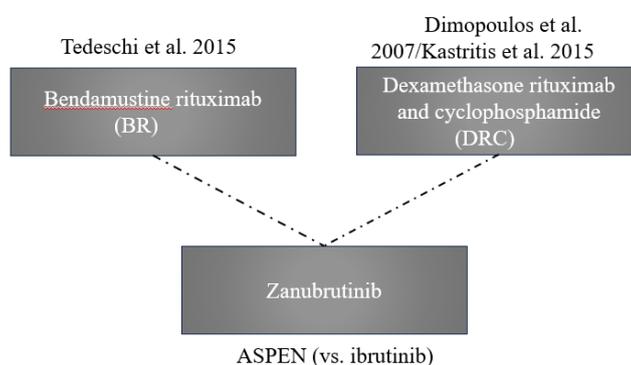


図 4.1 ザヌブルチニブと BR および DRC との間接比較  
(ASPEN 試験は IPD、BR および DRC 試験は AgD にアクセス可能)

MAICによるマッチング調整で考慮されたベースライン特性の分布の範囲内の試験集団では治療結果の比較が可能となる。アンカーありの場合、マッチングは効果修飾因子に基づく調整が必要であり、アンカーなしの比較の場合、効果修飾因子に加えて予後因子も調整する必要がある。ASPEN試験およびその他の公表文献に基づき、以下のベースラインの患者特性は、潜在的な予後因子または効果修飾因子であると考えられ、比較対照試験でのデータの入手可能性に関わらず、MAICのマッチング変数として含める必要がある。

- 年齢 (歳) (75以下 vs. 75超)
- 前治療の個数 (3以下 vs. 3超)
- ECOG PS (1以下 vs. 1超)
- MYD88/CXCR4 (C-X-C Motif Chemokine Receptor 4) 変異状態
- IgM濃度 (g/L) (40以下 vs. 40超)
- $\beta$ 2-マイクログロブリン濃度 (mg/L) (3以下 vs. 3超)
- 血小板数 (/L) ( $100 \times 10^9$ 以下 vs.  $100 \times 10^9$ 超)
- ヘモグロビン濃度 (g/L) (110以下 vs. 110超)
- 髄外疾患の有無
- WM IPSS (Waldenström's Macroglobulinaemia International Prognostic Scoring System)

また、試験間の各患者のベースラインおよび疾患の特徴の違いを調整するために Signorovitch ら<sup>17</sup>によって提案されたマッチングアルゴリズムが使用された。具体的には、ASPEN 試験から入手したザヌブルチニブの IPD について、重み付き平均ベースライン特性が比較対照の試験の論文で報告されたものと一致するように再重み付けされた。マッチング調整の過程で、各患者には、特定の比較対照の試験に参加する確率に対する ASPEN 試験のザヌブルチニブ群に参加する確率のオッズの逆数を表す重みが割り当てられた。患者のベースライン特性に基づき比較対照の試験に参加する可能性が高かったザヌブルチニブ群の患者には解析において高い重みが割り当てられ、各患者が比較対照の試験に参加する可能性に応じた分析への影響が考慮されている。

### (3) 間接比較の妥当性

医師評価による PFS および OS の Kaplan-Meier 曲線に関心のある主要な結果として、マッチング調整の有無の両方で治療間の結果を比較した。報告された Kaplan-Meier 曲線から再構成された患者データが BR および DRC で利用可能な場合、Cox の比例ハザードモデルを使用してハザード比を推定することで生存率を比較した。マッチング調整後、報告されたベースライン特性はそれぞれの比較において非常によく一致していた。MAIC の適用における仮定とその妥当性は表 4.1 のとおりである。

表4.1. MAICの適用における仮定とその妥当性

分類	仮定	仮定に対する考察
患者集団 (DRCにマッチさせたザヌブルチニブ 対 DRC)	ASPEN試験でのザヌブルチニブ群には、再発/難治性患者と化学免疫療法に適さない治療未経験患者を含むため、化学免疫療法に適した治療未経験患者の集団のみを含むDRC群に一致させるように調整した。この際、患者集団間の差異がMAICの結果に与える影響は限定的であると仮定した。	この仮定は、入手可能な臨床のエビデンスが限られていることから設定された。患者集団のこのような違いが、DRCと比較したザヌブルチニブの相対的な生存時間のベネフィットの過小評価につながった可能性がある。これは治療未経験患者の集団では、治療未経験患者と再発/難治性患者の両方を含む集団と比べて、PFSおよびOSの結果が良好であると想定する場合である。
マッチング変数	観察されていない重要な予後因子はザヌブルチニブ群と比較対照群の間でバランスが取れているため、MAICによる結果はバイアスが限定的でロバストであると仮定された。	未観察または未報告のすべてのベースライン特性を完全に調整することはほぼ不可能であり、これがMAICの一般的な限界である。その上で、マッチング調整の有無毎に実施された結果は、一貫して比較対照を上回るザヌブルチニブの延命効果が示さ

		れた。
--	--	-----

#### 4.1.1.2 委員会での議論および最終判断

(委員会) BR および DRC の臨床試験の患者集団が ASPEN 試験と異なる点に注意する必要がある。ASPEN 試験では約 81%が他治療の後にザヌブルチニブが投与されており、残りの患者は化学免疫療法が不相当であったため第一選択治療としてザヌブルチニブが投与されていた。それに対して、DRC の臨床成績は化学免疫療法が適している治療未経験の患者から得られたものである。この集団はザヌブルチニブの適応が承認されている集団に対応しておらず、ASPEN 試験の集団とは異なる。

(企業) 間接的な治療比較により集団を一致させ、結果の偏りを最小限に抑えるための調整を試みた。この試験では、再発/難治性の患者集団におけるザヌブルチニブと BR との比較、および治療未経験の患者集団におけるザヌブルチニブと DRC の結果を示している。初回提出では MAIC を使用した。

(ERG) 比較に利用できる患者データが限られているため、群間で臨床的に関連するリスク因子の分布に調整不可能な差異が生じている可能性がある。

(企業) 技術的な取り組みとして、間接比較の別の方法である STC も使用した。STC は比較により多くのサンプルサイズを利用できるため、より多くのデータを使用して効果的に共変量を調整できる観点からは、STC が好ましいアプローチである。

補足説明：間接比較に STC または MAIC のいずれかの手法を使用したところ、BR および DRC の両方と比較して、ザヌブルチニブは全体の無増悪生存期間が改善していた。また、使用されたアプローチに関係なく、BR と比較したザヌブルチニブの全生存率のハザード比は、DRC と比較したザヌブルチニブのハザード比よりも低かった。

(委員会) これらのハザード比は、DRC が BR よりも効果が高い可能性があることを示唆しているが、これがこれらの比較対照の実際の利点の違いを反映しているのか、それとも各間接比較で比較される母集団の違いに起因しているのか判断できなかった。

(委員会) 無増悪生存率と全生存率のハザード比ががん治療で通常見られるハザード比と比較して低く、ザヌブルチニブが非常に効果的な治療法であることを示唆している。MAIC と STC によって生成されたハザード比の点推定値は異なっていたが、信頼区間の重複が大きいため、2つの手法による結果にある程度の一貫性があることに同意した。

(臨床専門家) ザヌブルチニブと BR および DRC を比較する分析で得られたハザード比が妥当であると思われることを確認した。

(委員会) MAIC と STC の両方に不確実性と限界があると考えたが、MAICの方が透明性が高く、当初の MAIC から STC に切り替える正当な理由は示されていない。そのため、好ましいアプローチは当初の MAIC アプローチであると結論付けたが、これは不確実であることを認めた。また、ザヌブルチニブが化学免疫療法よりも臨床的に有効であることに同意し

た。しかし、間接的な比較には限界があるため、NHS の対象となる患者集団における治療効果の大きさには高度な不確実性があると結論づけた。

#### 4.1.2 ブリガチニブ (TA670)

ブリガチニブは、未分化リンパ腫キナーゼ (Anaplastic Lymphoma Kinase : ALK) 阻害剤による治療歴のない ALK 陽性の進行性非小細胞肺癌患者に対する単剤療法として、2020 年 4 月に製造販売承認を取得した薬剤である。

##### 4.1.2.1 臨床試験および間接比較

###### (1) 臨床試験

ブリガチニブおよび ALK 陽性進行性 NSCLC の一次治療で利用可能な既存の比較対照薬の有効性と安全性に関する情報を特定するためにシステマティックレビューを実施した。その結果、ブリガチニブ試験 1 件 (ALTA-1L)、アレクチニブ試験 3 件 (ALEX、J-ALEX、ALESIA)、セリチニブ試験 1 件 (ASCEND-4)、クリゾチニブ試験 3 件 (PROFILE 1007、PROFILE 1014、PROFILE 1029) の計 8 件のランダム化比較試験が特定された。それらのうち、ALTA-1L 試験および ALEX 試験には、この評価に関連する臨床的有効性のエビデンスが含まれていると考えられた。

###### (2) 間接比較の適用

ブリガチニブと主要な比較対照であるアレクチニブを直接比較した臨床試験は存在しなかった。ALTA-1L 試験はブリガチニブとクリゾチニブのランダム化比較試験で、ALEX 試験はアレクチニブとクリゾチニブのランダム化比較試験であった。そのため、ALTA-1L 試験と ALEX 試験のデータに基づき、ブリガチニブとアレクチニブの相対的な有効性を検討するための間接比較が実施された (図 4.2)。なお、J-ALEX および ALESIA は、試験の対象がアジア人集団のみで英国人を代表していないと考えられるため、間接治療比較から除外された。

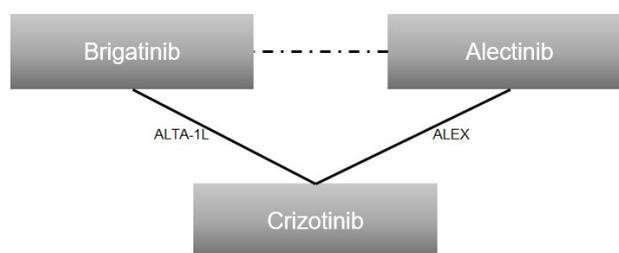


図 4.2 ブリガチニブとアレクチニブの間接比較  
(ALTA-1L 試験は IPD、ALEX 試験は AgD にアクセス可能)

なお、ALTA-1L 試験と ALEX 試験におけるデザイン上の主な相違点には、PFS の定義、治療のクロスオーバー (治療切り替え) の可否、追跡期間、試験前の化学療法歴有の患者の割合、脳転移有の患者の割合などがある。

ALTA-1L 試験と ALEX 試験の母集団の違いを調整する方法として、共通の治療群（クリゾチニブ）をアンカーとした調整済み間接比較（アンカー有 MAIC）と、ブリガチニブ群とアレクチニブ群をそれぞれ単群試験からのものと扱って比較する調整済み間接比較（アンカー無 MAIC）が適用された。参考までに Bucher 法（ALTA-1L データ内の個人に割り当てられたすべての重みが 1 に設定されるアンカー有 MAIC と同等）による比較も適用された。

今回 MAIC 法が適用された理由は、NMA では利用が困難な他の試験の情報が必要となり、回帰分析に基づく STC 法の適用においてはイベント数が十分でないためであった。なお、OS をアウトカムとした回帰分析では有効サンプルサイズおよび自由度がイベント数に依存するが、利用可能なイベント数は少ない状況であった。

- アンカー有のマッチング調整間接比較（アンカー有 MAIC）

臨床医による見解およびデータに基づく潜在的な予後因子および効果修飾因子の検討により、臨床専門家諮問委員会でベースラインの脳転移のみが予後に影響を与える可能性が高いことが示されたため、ベースラインの脳転移のみを調整因子としてアンカー有の MAIC が実施された。

- アンカー無のマッチング調整間接比較（アンカー無 MAIC）

ALTA-1L 試験と ALEX 試験において、ベースラインの脳転移を有する患者の割合と治療切り替えについては両試験のクリゾチニブ群の間でのみ異なっていた。そのため、アンカーとしてのクリゾチニブのリンクを削除し、あたかも 2 つの単群試験から得られた結果であるかのようにブリガチニブとアレクチニブの相対的な有効性を推定するためにアンカー無の MAIC が適用された。ALTA-1L 試験での治療切り替えを説明するために RPSFTM (Rank Preserving Structural Failure Time Model) の使用も行ったが<sup>18</sup>、治療切り替えの影響を完全には排除できず、OS に関するアンカー有 MAIC の結果の解釈が困難になったため、アンカー無の MAIC の解析も重要であると考えられた。

#### 4.1.2.2 委員会での議論および最終判断

- 間接比較に含める試験について

(ERG) 間接比較の対象試験に ALESIA 試験を含めるべきである。それは、ブリガチニブの欧州公的評価報告書で、臨床的有効性データをアジア系家族起源の集団から主にヨーロッパ系家族起源の集団に推定することが可能であると述べているためである。

(臨床専門家) 民族性が臨床結果に影響を与えるとは予想していないが、ALESIA 試験には主に中国出身者が含まれており、彼らはその後英国の NHS とは異なる医療制度にアクセスして異なる治療を受ける可能性がある。

(委員会) ALTA-1L 試験と ALEX 試験が NHS にとってより一般化可能な試験であり、ALESIA 試験を間接的な治療比較から除外し、ALTA-1L 試験と ALEX 試験のデータのみを使用することが適切であることに同意した。

- 間接比較の手法と解釈について

	企業の立場	ERGの立場*
1. アンカー無 MAIC	基本ケースとして設定。ブリガチニブとアレクチニブの相対的有効性を、2つの単群試験からのものであるかのように推定する。クリゾチニブ群の治療切り替えの影響とベースラインの脳転移患者の割合に関する両試験の差異を排除できる。	信頼できる結果は、すべての予後因子および効果修飾因子が考慮されている場合であり、これが満たされていないため、不適切である。
2. アンカー有 MAIC	治療切り替えによりバイアスが生じる可能性がある。	PFS、OSの利用可能な最良の推定値であるが、ALTA-1L試験のIPDにアクセスできず再現できない。
3. Unweighted Bucher	参考として提示。結果はアンカー有のMAICと同様である。	

\*ERGはALESIA試験におけるアレクチニブとクリゾチニブの比較により、ブルガチニブとアレクチニブの有効性を間接的に比較するために利用できる有効性のエビデンスが得られると考え、独自でALTA-1L試験、ALEX試験およびALESIA試験を含めてUnweighted Bucher法を適用した。ALESIA試験からのデータの追加によりブルガチニブとアレクチニブの比較結果の不確実性が増大することを強調した。

(委員会) DSU (Decision Support Unit) TSD18に「共通の比較対照を有し、それが利用可能な場合、『アンカー有』のPAICを検討すべきである。『アンカー無の』母集団調整は、ランダム化試験のネットワークが存在しない場合、または単群試験のみの場合に検討できる可能性がある。」と記載されていることに留意した。

(ERG) アンカー無 MAIC による信頼性の高い結果は、すべての予後因子と効果修飾因子が考慮されている場合であり、これが満たされていない。

(委員会) 最終的に、アンカー無 MAIC の結果は意思決定には受け入れられないと結論付けた。

- コスト最小化

(企業) 臨床試験の結果および間接比較における信頼区間が重複していることから、ブリガチニブとアレクチニブの有効性は同様である可能性が高い。

(ERG) 間接比較で統計的に有意な差がないことは2治療の間に差がないという統計学的なエビデンスを示しているわけではないので、類似性の根拠としては解釈できず、間接比較の結果を意思決定の情報として使用すべきではない。また、コスト最小化分析を行う前に同等性または非劣性を評価しないと、価格だけで判断され、効果の劣った治療が好まれるリスクが生じる。

- 生存時間データが不十分であることについて

(ERG) 臨床試験で全生存期間の中央値が観測されていないこと、ALTA-1L試験のクリゾチニ

ブの結果は治療の切り替えが行われており RPSFTM での調整後の生存時間の推定値は信頼性が低いと考えられる。企業が検討したブリガチニブ対アレクチニブの OS のハザード比 11 件のうち、企業が選択したアンカー無 MAIC のみが、ブリガチニブの OS がアレクチニブよりも数値的に優れた点推定値となった。アンカー無 MAIC は意思決定には不適切と考えており、企業が検討している他の 10 種類の OS のハザード比とは結果が異なるアンカー無 MAIC は、支持して使用できるほど頑健なものではないと考えている。

なお、企業が提示したアンカー有 MAIC、アンカー無 MAIC、Bucher の間接治療比較から得られた全生存期間に対するハザード比の結果が 0.83~1.36 の範囲で信頼区間が広いことについて (ERG) データが十分でなく、全生存期間に関する間接的な治療比較の結果はすべて不確実である。

(企業) ALTA-1L 試験の生存時間データは十分でなく、どちらの治療群でも全生存期間の中央値が観測されていなかったが、試験中に治療法を切り替えたすべての被験者を対象に、再打ち切りなしで RPSFTM を用いたアンカー有 MAIC の結果が、利用可能な最良の全生存期間の結果である。

(臨床専門家) 生存データは十分でないが交絡がなければブリガチニブの投与により時間の経過とともに生存期間が増加することが期待できる。ブリガチニブがアレクチニブと同じ作用機序を持つ第二世代のチロシンキナーゼ阻害剤であることに注目した。ブリガチニブとアレクチニブはいずれも、それぞれ ALTA-1L 試験と ALEX 試験でクリゾチニブと比較して無増悪生存期間により有効性の改善を示し、3 つの方法による間接比較はいずれも、ブリガチニブとアレクチニブが同様の無増悪生存期間の延長につながることを示唆している。アレクチニブとブリガチニブの生物学的および薬理的な類似性、および臨床現場での両技術の経験を考慮して、ブリガチニブによる全生存期間はアレクチニブと同様であることが期待できると確信している。

(委員会) 最終的に、ブリガチニブとアレクチニブが同様の作用機序を有することを考慮し、不確実性は残るものの、無増悪生存期間および中枢神経系無増悪生存期間の延長は全生存期間の利益につながる可能性があることを受け入れた。また、2 つの治療法の類似点を考慮し、ブリガチニブとアレクチニブで同様の全生存期間であると考えられることはもっともらしいことも認めた。

#### 4.1.3 テポチニブ (TA789)

テポチニブは間葉上皮移行因子遺伝子 (MET) ががん原遺伝子によってコードされる MET チロシンキナーゼ (肝細胞増殖因子の受容体) の選択的、可逆的小分子阻害剤で、METex14 スキッピング変異や MET 増殖など、MET の発がん性変化を伴うがんにおいて抗がん活性を示し、MET エクソン 14 (METex14) スキッピング変異を有する進行性非小細胞肺癌 (NSCLC) 成人患者の治療を適応として 2021 年 9 月に承認を取得した薬剤である。

#### 4.1.3.1 臨床試験および間接比較

##### (1) 臨床試験

進行性 NSCLC 患者を対象とした RCT、非ランダム化臨床試験、単群試験、および Real World Data (RWD) 研究を特定するためにシステマティックレビューが実施された。METex14 スキッピング変異を有する NSCLC 患者の治療に関連する 38 件の臨床成績の報告の中で、VISION (NCT02864992) 試験は、METex14 スキッピング変異または MET 増殖を有する進行性（局所進行性または転移性）NSCLC 患者を対象に MET の高選択性小分子阻害剤であるテポチニブ 500 mg の抗がん活性と忍容性を評価するために計画された進行中の第 2 相単群試験で、3 つのコホートで構成されていた。

- コホート A : METex14 スキッピング変異
- コホート B : MET 増殖
- コホート C : METex14 スキッピング変異（確認パート）

なお、コホート C はコホート A での METex14 スキッピング変異の結果を確認するために追加されたものである。

##### (2) 間接比較の適用

今回の評価で考慮された主要なデータはコホート A のものであった。現在テポチニブと比較可能な治療に関して、METex14 スキッピング変異を有する NSCLC 患者を対象とした比較試験のデータは存在していない。野生型 NSCLC における免疫療法および化学療法に関する臨床試験データは入手可能であったものの、患者集団が異なるため比較できていなかった。今回、諮問委員会で臨床専門家により、METex14 スキッピング変異が予後因子であるため、野生型 NSCLC の臨床試験データとの比較が非常に不確実になると示唆された。そのため、METex14 スキッピング変異を有する NSCLC 患者を対象としたレトロスペクティブな RWD がテポチニブとの比較に使用された。

VISION 試験の結果との比較に 4 つの RWD からの IPD が利用可能で、これらは免疫療法または化学療法で治療された患者の特徴や転帰が報告されており、間接的な治療比較を実施する十分な情報を含んでいた。間接比較の手法については、テポチニブと比較対照との患者特性のバランスを改善するために、傾向スコアによる調整法が適用された。観察されていない特性に不均衡がないと仮定し、傾向スコアでバランスを取ることで統計的に偏りのない集団が得られ、NICE DSU 17 のガイダンスに沿ったものになる。傾向スコアに基づいてグループのバランスをとるためにはマッチングや重み付けなどいくつかのアプローチが利用可能である。VISION 試験に類似したグループを取得するために比較用のデータセットに選択/除外基準を適用し、データの損失なくデータセット間で患者特性を一致させるように再調整できるため、マッチングよりも傾向スコアの重み付けが優先された。傾向スコアの推定に含める変数は、臨床専門家による重要性の順序に加え、含めるべき最も重要な要素についての意見が得られた。最終的に得られた共変量を応答との関係の強い順に並べると次のようになった。

- 前治療の経験の有無

- 年齢
- 転移性/ステージ 4 疾患の有無
- 性別
- 組織学
- 喫煙歴の有無

喫煙歴は化学療法では陰性予後因子、免疫療法では陽性予後因子と予想されるため、両方の療法に含められた。その他、ECOG PS も考慮すべき重要な変数であると考えられたが、データが欠落しているため解析に含めることはできなかった。さらに患者数は腺がんの組織型を持つ患者の割合と同程度のみで、VISION 試験と RWD 研究の両方において代替の組織型グループの患者の割合は相対的に小さかった。

なお、公表済の試験結果と比較するために MAIC による間接比較も実施された。この際、IPD を用いた間接比較で検討された重要な予後因子のうち利用可能な全ての患者特性についてバランスがとられた。報告されたすべての変数のマッチングが達成され、患者予後に関するグループ間の差異を可能な限り排除した比較が実施された。ただし、この方法には 2 つの限界があった。1 つ目は群間で METex14 スキッピング NSCLC と野生型 NSCLC の違いをマッチングで考慮できていない点で、2 つ目は METex14 スキッピング変異状態以外の主要な予後因子は一致したが、有効サンプルサイズが大幅に減少しており、一方の試験で未観測のため調整できない因子もあった点である。

#### 4.1.3.2 委員会での議論および最終判断

(委員会) ランダム化比較試験は多くの不確実性の原因を軽減するため、テポチニブの効果を評価するためには単群試験ではなく検証試験として実施または計画されるべきだったと考えられる。企業からの対象患者が少ないためランダム化比較試験の実施が非現実的であるとの見解には同意できない。エビデンスが 1 つの単群試験に基づいており、比較データの欠如により相対的な有効性を評価することが困難であったこと、生存時間データの観測が十分でなかったことにより、テポチニブのデータにはかなりの不確実性があると結論づけた。

(ERG) 間接的な治療比較を調整するには傾向スコアによる調整法が最も適切な方法であることに同意した。

(委員会) 企業のリアルワールドのコホートには英国出身者は含まれていない。

(臨床専門家) 受けた治療法とその後の治療法は英国で行われている治療法と一致していない。

間接的な治療比較の結果、テポチニブには化学療法と免疫療法の両方と比較し、統計的に有意な無増悪生存期間の延長の利点があることが示されたが、統計的に有意な全生存期間の延長は認められなかった。

(臨床専門家) 間接的な治療比較から得られた全生存期間の結果は、臨床現場、特に化学療法で期待されるものは反映されていない。

(委員会) 間接的な治療比較の結果は一貫性がなく、予想に反して、化学療法が免疫療法より効

果的であるように見える場合があることに同意した。間接的な治療比較は小さなサンプルサイズに基づいているためロバストではなかった可能性があり、化学療法や免疫療法と比較して全生存期間がどの程度改善されるかは不確実であると考えられる。

(臨床専門家) 企業は特定の発がん性バイオマーカーを持たない患者を対象とした比較試験のデータに基づいて間接的に治療法を比較することを検討できることを示唆した。

臨床専門家の示唆は比較対照の患者数を増やすことができるため、よりロバストになる可能性があった。

(委員会) 第2回委員会会議でこれらの分析には価値がある可能性があることに同意したが、比較試験の母集団がテポチニブの母集団とは異なるため不確実性があることを認めた。

(企業) 最初の間接的な治療比較は METex14 スキッピング変異を有する NSCLC 集団で行われているため、その結果も考慮されるべきであるとの見解を繰り返したが、第2回委員会会議では特定の発がん性バイオマーカーを持たない患者のみを対象とした比較試験のデータを使用した追加の間接的な治療比較の結果を提供することに合意した。

(委員会) 企業の当初の間接的な治療比較の結果は非常に不確実だが、意思決定において考慮されるべきであると結論付けた。

企業は追加の間接的な治療比較の結果を提供した。この比較は、VISION 試験と、特定の発がん性バイオマーカーを含まない野生型 NSCLC を対象とした試験からのデータとの間で行われ、臨床専門家と協議して、各比較対照治療に最も関連性の高い試験を選択した。

(委員会) 試験の選択には同意したが、ドセタキセル単独療法に対する TAX320 試験の選択には疑問を呈した。それはこの試験が実施されてから NSCLC の治療が大きく変わったからであった。

(企業) ドセタキセルに関するもう1つの潜在的なデータ源が免疫療法とのクロスオーバー治療の割合が高い試験だったため TAX320 試験が選択されたと説明した。TAX320 試験は NSCLC に対する免疫療法治療よりも前から行われているため、そのような治療のクロスオーバーは含まれていなかった。企業は野生型 NSCLC を対象とした試験と VISION 試験のデータを比較するために MAIC を使用した。

(企業・ERG) このアプローチには重要な制限があることに同意した。

(ERG) MAIC ではマッチングされる集団は METex14 スキッピング変異を有する NSCLC ではなく野生型患者であるためバイアスのリスクが生じると説明した。またマッチングプロセスで VISION 試験のコホートのサンプルサイズが大幅に減少しており、特定された予後因子の観点から2つの集団のバランスをとるために、一部の患者の大幅な再加重が必要であることを示唆していると説明した。

(委員会) VISION 試験のサンプルサイズのこの減少は、METex14 スキッピング変異を有する NSCLC の母集団に対する間接比較結果の一般化可能性の欠如を示唆していると理解した。企業の追加の間接的な比較も不確実ではあるが、意思決定には適切だった。

(委員会) 間接治療比較に対する企業の新しいアプローチでは、治療ラインや組織学に応じて、NHS の実臨床に関連する特定の治療法との比較が可能になったが、最初の企業のアプロー

チではテポチニブとグループ化された治療法の比較のみが可能であり、化学免疫療法は除外され、関心のある実臨床のデータが欠如している。

(企業) 実臨床のコホートは METex14 スキッピング変異を有する NSCLC に特有のものであるため、最初のアプローチも考慮されるべきであるとの見解を繰り返した。

(ERG) 企業の見解に同意したが、どちらのアプローチが明らかに優れているわけではなく、複数の不確実性間のトレードオフである。

(企業) MAIC アプローチの限界を受け入れ、基本ケースとしてこの新しい分析を選択した。

(委員会) VISION を比較対照の野生型集団に再加重することによって引き起こされるバイアスのリスクと、METex14 スキッピング変異を有する NSCLC 集団に対する一般化可能性の低下により、追加の MAIC による結果は不確実であることに同意した。各比較の MAIC による間接比較には制限があるものの、意思決定には適切であると結論付けた。

#### 4.2 各国の HTA 機関による間接比較の利用と受け入れ状況

Macabeo ら<sup>19</sup>は、欧州 5 カ国（英国、フランス、ドイツ、イタリア、スペイン）の抗がん剤の HTA 評価における間接比較の受け入れを調査している。2018 年 4 月から 2021 年 4 月に公表された 543 件の評価レポートのうち、120 件（22%）に間接比較が含まれた。国別では、英国が 38/75 件（51%）、フランスが 11/177 件（6%）、ドイツが 21/120 件（18%）、イタリアが 29/123 件（24%）、スペインが 21/48 件（22%）であった。間接比較の手法は、NMA（23%）、Bucher 法（19%）、未調整間接比較（ナイーブな間接比較を含む）（14%）、MAIC（13%）、調整した間接比較（具体的な手法は記載なし）（7%）、逆確率重み付け法（Inverse Probability of Treatment Weighting : IPTW）（4%）、STC（3%）、ネットワークメタ回帰（Network Meta Regression : NMR）（1%）、不明（30%）であった。このうち、RCT に基づく間接比較は全体の 52%、単群試験に基づく間接比較は 17%であった（うち、アンカーのない MAIC または STC（55%）、未調整比較（35%）、IPTW（5%））。（なお、レポート内で複数の手法が用いられた場合はそれぞれカウントしたため合計は 100%を超える。）

間接比較が明示的に受け入れられた評価は、全体の 30%であった。間接比較の受け入れ（レポート中に用いられた間接比較が妥当である旨の記載がある場合に受け入れと評価）が最も多いのは英国の 47%であり、ドイツ（38%）、イタリア（31%）、スペイン（5%）が続き、最も少ないのはフランスの 0%であった。なお、HTA 評価レポートの 49%で、間接比較の受け入れが不明瞭であった。具体的な手法（NMA、Bucher 法、MAIC、IPTW、STC、NMR）のうち、英国では IPTW を除くすべての手法が受け入れられる一方で、ドイツで受け入れられた手法は、Bucher 法のみ（7/12 件）であり、MAIC（0/3 件）は受け入れられておらず、NMA は提出されていない。間接比較が受け入れられなかった理由は、異質性とバイアスリスク（48%）、データの不足（43%）、統計手法が不適切（41%）、不明（32%）などであった。なお、間接比較の受け入れ可能性は各国の経済分析のガイドライン等の規定の影響を受けていることを、結果を解釈する際に留意する必要がある。

van Beekhuizen ら<sup>20</sup>は、EUnetHTA（欧州医療技術評価ネットワーク）が 2006 年から 2021 年に行った 23 件の REA（Joint Relative Efficacy Assessments）をレビューした。REA では 64 件の比較対照との比較が実施され、直接比較は 11 件（17%）、間接比較は 25 件（39%）、比較不能と結論さ

れたものは28件(44%)であった。間接比較の内訳は Bucher 法が8件、MAIC、STC等の PAIC が7件であり、7件の REA には NMA が含まれていた。多くの REA で EUnetHTA による間接比較に基づくエビデンス受け入れ状況は明確ではなかったが、エビデンスの不確実性や限界が指摘されていた。

## 5. 解析手法とシミュレーション研究

### 5.1 解析手法

Bucher 法を含め標準的な間接比較や NMA では、統合するすべての試験間で効果修飾因子の分布が同じであることを前提とする<sup>21</sup>。近年注目を集めている MAIC、STC、ML-NMR などの PAIC は、1つの試験(一般に自社の試験)では IPD を利用できることを利用して、試験間で効果修飾因子の分布が異なる場合でも、その分布の違いを調整して評価対象技術の比較対照技術に対する相対治療効果を推定する方法である<sup>6</sup>。MAIC と STC は基本的に2試験間で比較する状況で利用可能な手法であるが、ML-NMR は3試験以上の間で比較する状況でも利用可能な手法である。近年、PAIC の適用事例は増加傾向にあり、2010年から2018年に NICE に提出された PAIC は、多くががん領域(アウトカムはイベント発生までの期間)で適用されており、そのうちの大部分がアンカーを有さない MAIC であった<sup>22</sup>。

以下では、PAIC の各手法の特徴について説明する。説明には、2章で定義した表記(AB試験、AC試験、B試験、C試験、IPD試験、AgD試験など)を用いる。なお、1章で述べたとおり、データサイエンス部会の費用対効果評価に関するタスクフォースでは、2022年1月に PAIC に関する報告書<sup>4</sup>を公表し、本手法の理論的側面について解説した。この報告書では数式を交えながら、PAIC を適用する際に前提となる仮定の説明や、因果推論の基本的な考え方や基本的な用語の説明をしているので、PAIC の詳細を知りたいければ、是非この報告書を参照いただきたい。本報告書は PAIC の実用的側面に焦点を当てた報告書であるため、PAIC の理論的側面については、その特徴を述べるに留めている。

#### 5.1.1 アンカーを有する間接比較

##### 5.1.1.1 MAIC

MAIC は Signorovitch ら<sup>17</sup>によって提案された、傾向スコアを利用した(傾向スコアの逆数で重み付けした)重み付き解析を拡張した手法である。本手法における傾向スコアは、AgD試験に対して、IPD試験に割り当てられるオッズとして表される。IPD試験の共変量の周辺分布が AgD試験に一致するように、傾向スコアの逆数によって IPD試験の患者を重み付けする。ただし、AgD試験については一般に共変量の周辺分布の要約統計量(連続量の場合は平均や標準偏差、2値変数やカテゴリカル変数の場合は割合)しか利用できないため、モーメント法を用いて重みを推定することになる。

本手法の限界としては、まずは基本的に2試験の状況でしか解釈可能な相対治療効果を推定できない点が挙げられる。次に、傾向スコアを推定するにあたり、すべての効果修飾因子をモデル

に含める必要があるため、両試験ですべての効果修飾因子が測定され且つ報告されていることが前提となる。理論的には効果修飾因子ではない予後因子をモデルに含めると精度が低下する。

さらに、本手法では、AB試験の集団とAC試験の集団を統合した集団における治療効果(Average Treatment Effect: ATE)やAB試験の集団における治療効果(Average Treatment Effect on the Treated: ATT)ではなく、AC試験の集団における治療効果(ATT)を推定する。したがって、同じ2試験(AB試験とAC試験)に対して本手法を適用したとしても、どちらの試験のIPDを有しているかによって、異なる集団における治療効果を推定することになり、異なった結果が得られる。さらに、意思決定を下す際に関心のある集団はAB試験の集団でもAC試験の集団でもなく、また別の集団である可能性もあるため、どの集団に対する治療効果を推定することが適切であるかについてはよく考えておく必要がある。

最後に、本手法は傾向スコアに基づく手法であるため、両試験間で共変量の分布が十分に重なっている必要がある。共変量の分布の重なりが十分でない場合、有効サンプルサイズが減少する。また、PAICの枠組みでは、モーメント法ではなく、重みの推定に制約を加えたその他のエントロピー平衡化も提案されている<sup>23</sup>。

#### 5.1.1.2 STC

STCは回帰モデルに基づく方法を拡張した手法である<sup>24</sup>。IPD試験の個人レベルのデータに対してアウトカムモデルを当てはめ、モデルに含めた予後因子や効果修飾因子のパラメータを推定する。各パラメータを推定した後、個々の共変量の値をAgD試験の各共変量の平均で置き換えることで、AgD試験のアウトカムの平均を推定する(標準的な方法)。しかしながら、アウトカムの平均は各共変量の平均のみではなく共変量の分布自体に依存するため、この各共変量の平均で置き換えるという手法では、多くの状況でバイアスを伴う。そこで、各共変量の平均で置き換えるのではなく、AgD試験の共変量の同時分布から標本を抽出し、その標本を用いてアウトカムを推定するというステップを何回も繰り返すことで、AgD試験のアウトカムの平均を推定する手法(シミュレーションに基づく方法)が提案されている<sup>25</sup>。ただし、この手法では、どのように同時分布を設定するかが課題となる。

また、本手法は回帰モデルに基づく手法であるため、共変量の分布が重なる範囲を超えた部分も外挿してパラメータを推定することになる。外挿自体は、それを行うことが妥当でない場合には欠点となるが、共変量の分布の重なりが十分ではなく、有効サンプルサイズが大きく減少することは追加的有用性の有無の評価のために実施する間接比較ではよく起こりうることであるため、外挿自体が利点となる場合もある。また、アウトカムモデルを当てはめる手法であるため、情報量が少ない場合(例えば、イベント数が少ない場合)には、パラメータの推定が不安定になる可能性があるため、注意を払う必要がある。

#### 5.1.1.3 ML-NMR

ML-NMRはネットワークメタ回帰を拡張した手法である<sup>26,27</sup>。ネットワークメタ回帰は、すべての試験がAgD試験である場合に適用可能な手法であり、試験間の効果修飾因子の分布の不均

衡を調整することができるものの、集団レベルの共変量で調整するため、エコロジカルバイアスが生じる可能性がある。ネットワークメタ回帰を拡張した手法の一つが階層的関連回帰に基づくアプローチである。このアプローチでは、IPD と AgD を正しく関連付けることにより、エコロジカルバイアスを制御することができる。

ML-NMR は 3 試験以上のネットワークに対しても利用可能な手法である。さらに本手法では、追加の仮定 (shared effect modifier assumption) を置くことで、あらゆる集団で認められるはずの相対治療効果を推定できる。

### 5.1.2 アンカーを有さない間接比較

アンカーを有さない間接比較では、一般に治療 B を評価した単群試験 B と、治療 C を評価した単群試験 C の間で比較して、治療 B の治療 C に対する相対治療効果を推定する。最も単純な手法は、無調整で治療 B と治療 C を比較するナイーブな間接比較である。ただし、ナイーブな間接比較では、両試験間ですべての効果修飾因子および予後因子の分布が同じであるという非常に強い仮定を必要とするため、一般にバイアスを伴う可能性が高い。アンカーを有する間接比較と同様に、アンカーを有さない間接比較でも MAIC や STC の適用が可能である。ただし、アンカーを有さない間接比較では、妥当な相対治療効果を推定するためには、すべての効果修飾因子に加えてすべての予後因子も考慮する必要があるため、これらの因子が両試験で測定され且つ報告されていることが前提となる。これは非常に強い仮定であり、成立していると思なすのは一般に困難である。従って、本手法を適用したとしてもバイアスを制御しきることができない可能性があるため、適用には注意を払う必要がある。また、これまで述べたのは、比較対照技術の AgD を利用した手法であるが、過去の臨床試験、レジストリやデータベースから比較対照技術の IPD が利用可能な場合、傾向スコアを用いた方法を用いて、比較可能な外部対照群を構築し、治療 B の治療 C に対する相対治療効果を推定する方法も提案されている<sup>28</sup>。なお、この場合、臨床試験データに基づく群と、異なる臨床試験や観察研究に基づく群間で比較することになるため、品質が異なるデータソースから得られた群間で適切な比較が実施可能かどうか、慎重に検討する必要がある。

### 5.1.3 その他

#### 5.1.3.1 二重頑健推定量

二重頑健推定量では、傾向スコアモデルとアウトカムモデルの両方を考慮することで、モデルの誤特定に伴う影響を軽減することが可能となる。例えば、ある種の二重頑健推定量は共変量で調整したアウトカムモデルに基づくが、同時に AgD 試験に対して IPD 試験に割り当てられるオッズの逆数による重みを利用する<sup>29</sup>。二重頑健推定量では、傾向スコアモデルとアウトカムモデルの両方を正しく特定する必要はなく、2 つのうち何れかが正しく特定されてさえいれば、一致推定量を得ることができる。しかしながら、両方のモデルが誤特定されている場合、二重頑健推定量を用いたとしてもバイアスが生じる。アウトカムがイベント発生までの時間で、両試験で IPD が利用可能な場合、オッズの逆数による重みを用いた重み付きコックス回帰モデルを用いて条件

付きハザードを算出した後、Daniel ら<sup>30</sup>の方法を用いて周辺化することで二重頑健推定量を推定する方法が提案されている。一方、1つの試験でしか IPD が利用できない場合、モーメント法（あるいは、他のエントロピー平衡化の方法）を用いて重みを推定し、重み付けしたデータに対して STC を適用した後、周辺化することで二重頑健推定量を推定することが可能となる。上述のとおり、MAIC はアンカーを有さない状況で利用されることが多いことから、Park ら<sup>29</sup>はこの状況下での MAIC、STC および二重頑健推定量を比較している。

### 5.1.3.2 MAIC が有する二重頑健性の性質

MAIC は傾向スコアを用いた重み付き解析を拡張した手法であるため、傾向スコアを用いた重み付き解析のように、すべての効果修飾因子が測定されていて、かつ統合する2試験のうち一方の試験に割り当てられる確率が正しくモデル化されている（試験の選択モデルが正しく特定されている）場合には一致推定量を得ることができる。一方で、MAIC はエントロピー平衡化の一種であるため、Cheng ら<sup>31</sup>は、MAIC が本質的に線形な二重頑健性の性質を有していること、つまり、試験の選択モデルか、IPD 試験のデータに対して暗黙に仮定した線形な平均アウトカムモデルのいずれかが正しく特定されている場合には一致推定量を得ることができることを示した。さらに、両方のモデルが正しく特定されている場合には有効推定量となり、STC の立場からは、線形な平均アウトカムモデルの下での再重み付き推定量（重み付けした集団に対して線形な平均アウトカムモデルを当てはめた STC）となることを示した。STC は、IPD 試験のデータに対する線形なアウトカムモデルが正しく特定されている場合に限り一致推定量を得ることができるため、MAIC を適用することで、線形なアウトカムモデルを用いた STC よりも頑健な推定量を得ることができる。一方、非線形なアウトカムモデルを用いた STC は、共変量の同時分布に対する追加の仮定を置かない限り、一般にバイアスを伴う。理由としては、モデルによる予測値の平均は必ずしも共変量の平均から推定した予測値とは一致しないからである。

関心のある相対治療効果に対する MAIC 推定量は、ノンパラメトリックモデルの下で局所セミパラメトリック有効推定量となる。つまり、試験の選択モデルと線形な平均アウトカムモデルの両方が正しく特定されている場合に限り、データの発生分布に対して何らパラメトリックな制約を置く必要がないモデル下において有効推定量となる。したがって、試験の選択モデルと線形な平均アウトカムモデルの両方が正しく特定されている場合には MAIC を適用すべきである。同様の二重頑健性とセミパラメトリック有効の性質は、両試験で IPD が利用できる場合に試験間での共変量の分布の均衡を保つための手法である再重み付け手法でも認められている。

MAIC は ATT の推定において、明示的に拡張 (augmentation) しなくても、本質的に二重頑健性とセミパラメトリック有効の性質を有しているが、線形な平均アウトカムモデルであることが条件となるため、通常の AIPW (Augmented Inverse Probability Weighted) 推定量よりも制約が大きい。一部の試験でしか IPD を利用できない状況下での間接比較では、公表された論文等を通じて、AgD 試験の共変量の平均や標準偏差のみが利用可能である。

線形な平均アウトカムモデルに基づく STC は、重み付けしていない集団に対してアウトカムモデルを当てはめる。したがって、このモデルを誤特定した場合、AgD 試験の集団での平均アウトカムモデルにバイアスを伴う可能性がある。しかし、試験の選択モデルを正しく特定した場合

(MAICの重みが試験の選択のオッズで近似できる場合)、再重み付けしたSTCにより、AgD試験の集団での平均アウトカムに対する局所的に線形な近似を可能にし、誤特定に伴うバイアスを減少させることができる。

### 5.1.3.3 条件付きの治療効果と周辺治療効果

MAICはAC試験の集団における周辺治療効果を推定する方法であるが、標準的なSTCはAC試験の集団における条件付き治療効果を推定する方法である(5.1.1.2章)。一般に意思決定に利用したいのは周辺治療効果の方である。上述のとおり、MAICはアウトカムがイベント発生までの時間である場合の間接比較で利用されるケースが多い。アウトカムがイベント発生までの時間の場合、一般にコックス回帰モデルが利用されるが、コックス回帰モデルによって得られる条件付きハザード比は併合可能でないため、周辺治療効果の推定にバイアスを伴う。また、併合可能でない治療効果は共変量の組み合わせによって値が異なるため、間接比較における試験間の比較に用いるのは一般に適切ではない。そこで、条件付き治療効果を推定した後、共変量の同時分布に関して積分を取ることで(平均化することで)、間接比較に利用可能な周辺治療効果を推定することができる。Remiro-Azócarら<sup>32</sup>は、パラメトリックg-computationやモデルに基づく標準化を利用することで、一方の試験ではAgDしか利用できない場合でも、周辺治療効果を推定する方法を提案している。

## 5.2 シミュレーション研究

間接比較の手法を評価した幾つかのシミュレーション研究では、標準的なSTCと比較してMAICの性能が悪いことが示されている<sup>33,34</sup>。しかしながら、これらのシミュレーション研究では、IPD試験とAgD試験の間で共変量の分布はほとんどオーバーラップしておらず、かつアウトカムモデルはおおよそ正しく特定されている状況を想定しているため、STCにとって好ましい状況設定下での検討といえる。共変量のオーバーラップが少ない試験間での外挿は、平均アウトカムに対するパラメトリックモデルに大きく依存する。これらの状況設定下でMAICの性能が悪かった原因は、そのような外挿を支持するデータが不足していたためである。共変量のオーバーラップが十分である状況設定下では、MAICの性能は改善することが示されている<sup>33,34,35</sup>。さらに、オーバーラップが十分である状況下では、MAICが有する二重頑健性の性質から、MAICは線形なアウトカムモデルに基づくSTCよりも頑健な推定量を得ることができる。なお、仮定からの逸脱がない場合、MAICによりバイアスのない相対治療効果を得ることができるものの、有効サンプルサイズが小さい場合にサンドイッチ分散を用いると、標準誤差を過小評価する可能性があることが示されている。また標準的なSTCは条件付き治療効果を推定する手法であるため、周辺治療効果の推定を目的とした場合には、被覆確率が妥当ではなく、バイアスを伴う推定値が得られることが示されている<sup>35</sup>。

## 6. 細胞・遺伝子治療の HTA の課題と対応

### 6.1 細胞・遺伝子治療の HTA における課題

医薬品の HTA において、アンカーのない間接比較が採用される場面、採用せざるを得ない場面のひとつに、CAR-T 療法などの細胞治療や遺伝子治療がある。4 章ではがん領域を中心に間接比較の事例の詳細を説明したが、ここでは、細胞・遺伝子治療について、疾患の希少性、大きな臨床ニーズ、乏しい臨床エビデンス、生涯にわたって臨床利益が得られる可能性と期待、そして、潜在的に高額な費用といった意思決定の背景や文脈を俯瞰することを意図して、HTA における課題と実際に関する議論を紹介する。

英国 York 大学の Drummond らは、こうした細胞・遺伝子治療の HTA における課題を次のように整理している<sup>36</sup>。

#### 臨床的エビデンスに関する主な課題

サロゲートエンドポイント	迅速承認はサロゲートエンドポイントのエビデンスによって承認されることが多い。一方で、サロゲートエンドポイントの妥当性の検証（バリデーション）は同適応症の RCT のメタアナリシスによって評価される <sup>37</sup> 。希少疾病の場合、RCT のメタアナリシスによってサロゲートエンドポイントのバリデーションを行うことは困難である。
比較エビデンスの欠如もしくは単群試験	細胞・遺伝子治療は患者数が少なく、疾患の重篤性、治癒の可能性等により、倫理的な配慮や試験の実施可能性から単群試験が評価時の唯一のエビデンスとなることが多い。単群試験では、ヒストリカル対照、外部対照が必要となる場合があるが、交絡因子を調整するために因果推論の手法が用いられる。これは、標的試験エミュレーション（仮想の RCT をデザインし、その RCT を模倣する形で、いわゆる RWD を用いる観察研究を設計するアプローチ）とみなすことができるが、利用可能な大規模なデータセットがない希少疾患においては課題となり、評価手法として推奨していない HTA 機関も存在する。
臨床試験のサンプルサイズが小さい	臨床試験のサンプルサイズは、遺伝子治療では特に少なく、あるレビューによると、登録患者数の中央値は 213 名、半数の臨床試験は 20 名未満であった。サンプルサイズが小さい場合に、臨床効果のエフェクトサイズの推定の不確実性が大きくなり、また、臨床効果の不均一性の評価（サブグループの評価）が困難になる。また、単一の試験施設のみで臨床試験が行われた場合には、結果の一般化可能性についても懸念が生じる。
長期エンドポイントの外挿	迅速承認による場合、評価を実施する時点で利用可能な臨床試験の観察期間が短く、長期エンドポイントへの外挿が必要となることがある。これは、関心のあるアウトカムとしての生存曲線をどのようにモデル化すべきかについて、かなりの不確実性があることを意味する。この不確実性は、高齢の患者と比較して若い患者を対象とする治療法でいっそう大きくなる。対

	処法として、シナリオ分析と感度分析の充実が提案されている。
--	-------------------------------

#### 経済的エビデンスの主な課題

疾患の重症度	人々は重篤な疾患を持つ患者の健康状態の改善により大きな価値を置くというエビデンスがいくつかあり、重症度によって QALY に重みをつけたり、意思決定の閾値を大きくするアプローチが検討されている。
代替治療法の欠如	希少疾患は重篤な疾患であっても、積極的な代替治療法が存在しない場合も多く、アンメットニーズが大きい。HTA 機関での評価では「臨床的な付加価値 (added clinical value)」(フランス、ドイツ、イタリア) または「増分 QALY」(英国) のいずれかの指標を使用している。新しい治療法が有効であるなら、高い臨床的な付加価値が示唆されたはずである。したがって、治療法が存在しない状況での新しい治療が、なぜそれ自体が付加価値をもたらすのか、費用対効果が悪い場合でも提供されるのか、疑問に思う人もいるかもしれない。その理由は、現在十分なサービスを受けていない個人に治療の選択肢として新しい治療を提供することによって健康の不平等を減らすという、公平性や倫理に基づいている。
生存や QOL の大きな改善	細胞・遺伝子治療は「革新的 (transformational)」と呼ばれることがある。英国 NICE は Highly Specialised Technologies プログラムによって、10 QALY 以上の改善が期待される治療の ICER 閾値を大きな値に設定している。多くの細胞・遺伝子治療は疾患の治癒を意図しており、治癒は継続的な治療よりも価値が高いという主張もあるが、実証的なエビデンスが得られた治療はまだない。
オプションの価値	将来のより良い治療を受ける機会という意味の価値であるが、遺伝子治療で認められたものはない。
介護者や家族にとっての価値	家族が患者と接する際の精神的なストレスに対する価値や家族等によるインフォーマルな介護の価値は広く認識されているが、HTA で定量的に評価されることはほとんどない。
科学的な波及効果	細胞・遺伝子治療が実用化の初期段階にあると考えると、科学的な波及効果は広範囲に及ぶ可能性がある。米国 ICER (The Institute for Clinical and Economic Review) は新しい作用機序を評価の際の考慮事項に含めている。
割引率	一回のみの細胞・遺伝子治療は初期費用が高く、利益は長期に及ぶため割引率が ICER に大きく影響する (将来、長期間にわたって獲得する QALY は現在価値に割引かれるが、時間 0 に発生する費用は時間による割引を受けないため、割引率が高いと費用対効果が悪化する)。細胞・遺伝子治療について標準の割引率を引き下げるという議論もあり、HTA 当局が割引率に関する感度分析を行う可能性がある。
不確実性	細胞・遺伝子治療における主な不確実性は、臨床効果の長期的な持続性であり、費用効果の分析モデルに必要な仮定の信頼性である。対処法には、

	シナリオ分析がある。
市販後のエビデンス	臨床効果の持続性に関する不確実性は、上市後のエビデンスの収集が対処法のひとつである。他方で、高い費用と長期の有効性に関する不確実性に対して、保険償還や価格等による管理について様々な取り組みがあり、数量ベースとアウトカムベースの調整がある。前者は単純な値下げや価格と数量に関する契約、価格の上限、段階的な支払などがある。後者にはエビデンスの収集とリンクした方式やアウトカムに連動した償還方式などがある。

## 6.2 日本での評価事例

日本で評価が行われ結果が公開されている細胞・遺伝子治療にチサゲンレクルユーセル（キムリア）がある。チサゲンレクルユーセルは、キメラ抗原受容体 T 細胞（CAR-T）療法と呼ばれる細胞治療であり、希少疾病用再生医療等製品の指定を受けた。再発または難治性の CD19 陽性の B 細胞性急性リンパ芽球性白血病（B-ALL）、および再発または難治性の CD19 陽性のびまん性大細胞型 B 細胞リンパ腫（DLBCL）を効能・効果として、2019 年 5 月に薬価収載され、同時に H3 区分として費用対効果評価対象に品目指定された。

2021 年 3 月に公表された費用対効果評価報告書<sup>38</sup>によると、MAIC を用いたアンカーなしの間接比較は、15 歳以上 25 歳未満の B-ALL を対象とし、比較対照技術をブリナツモマブ単独療法（±同種 HSCT）およびイノツズマブとオゾガマイシンの併用療法（±同種 HSCT）とした追加的有用性の評価に用いられた。また、70 歳未満および 70 歳以上の DLBCL を対象とし、比較対照技術をサルベージ化学療法（70 歳未満は±同種 HSCT）とした追加的有用性評価にそれぞれ用いられた。

B-ALL および DLBCL のいずれも、システマティックレビューの結果、チサゲンレクルユーセルのランダム化比較試験は確認されなかった。そのため、B-ALL については 3 つの単群試験（B2101J、ELIANA/B2202、ENSIGN/B220）をプールした IPD（15 歳未満、15 歳以上の部分集団）を用いて、DLBCL については単群試験である JULIET 試験の結果を用いて間接比較が行われた（70 歳未満の分析対象集団については 70 歳未満の部分集団、70 歳以上の患者は少ないため、70 歳以上の分析対象集団に関する分析においては JULIET 試験の全体集団が用いられた）。比較対照の臨床データは、15 歳以上 25 歳未満の B-ALL の比較対照ブリナツモマブは Gore 2018 の全体集団、比較対照イノツズマブの臨床データは Bhojwani 2019 の全体集団が用いられた。DLBCL については、CORAL 継続試験（すべての患者が 70 歳以下）の全体集団が用いられた。

それぞれ比較対照の臨床試験で報告されている共変量を用いて、OS を評価指標とした MAIC によるアンカーのない間接比較が実施された。分析結果の推定値等は報告されていないが、各分析対象集団の評価において、OS のハザード比が 1 よりもかなり小さく、95%信頼区間が 1 を含まないことから企業分析の結果が支持され、公的分析班においても追加的有用性ありと評価された。

## 適応症・患者集団ごとの分析手法

対象集団	比較対照	分析手法
小児 B-ALL 15 歳未満	ブリナツモマブ	IPD を用いた間接比較 (Cox 回帰)
小児 B-ALL 15 歳以上	ブリナツモマブ	MAIC による間接比較 (アンカーなし)
小児 B-ALL 15 歳以上	イノツズマブ	MAIC による間接比較 (アンカーなし) 実際には、キムリア IPD は 15 歳以上、対照の AgD は年齢問わず全体集団
成人 DLBCL 70 歳未満	サルベージケモ	MAIC による間接比較 (アンカーなし)
成人 DLBCL 70 歳以上	サルベージケモ	MAIC による間接比較 (アンカーなし) 実際には、キムリア IPD は年齢問わず全例 vs 対照の AgD は全例が 70 歳未満

### ● アンカーのない間接比較に関する論点

企業は、DLBCL の JULIET 試験で 70 歳以上の患者および OS イベント数がごく少なく、統計的検出力の不足や間接比較における比較可能性の検討や調整が困難となることから、全体集団を用いた分析を行った。公的分析でも全体集団を用いて評価を行うことが支持された。その他、費用対効果評価報告書から読み取れる分析上の課題はなかった。チサゲンレクルユーセルの臨床データは単群試験のみではあるが、期待される有効性の程度は相当に大きく、間接比較に伴う不確実性や長期予後の推定に対する不確実性を上回ると解釈された結果と考えられる。

## 6.3 海外での評価事例

チサゲンレクルユーセルについては、日本以外の国々でも医療技術評価が行われている。ここでは、ふたつのレビュー論文に基づいて、CAR-T 療法に代表される細胞・遺伝子治療の HTA 評価の課題と実際について紹介する。

Gye らが 2020 年に実施したレビューによると、小児 B-ALL を対象としたチサゲンレクルユーセルは、オーストラリア (MSAC)、カナダ (CADTH)、英国 (NICE)、ノルウェー (MoMA)、米国 (ICER) で医療技術評価が実施された<sup>39</sup>。比較対照は、ブリナツモマブもしくはクロファラビン (単剤もしくは併用) であった。チサゲンレクルユーセルの単群試験である ELIANA 試験、ENSIGN 試験と比較対照の臨床試験を共変量等の調整を行わないナイーブな間接比較により評価された。米国 ICER は正式な統計的分析を実施しなかった。各評価機関による追加的有用性の解釈はいずれもポジティブであったが、同時に比較対照のない少数例の単群試験で追跡期間が限られるなど、評価の不確実性を指摘していた。成人 DLBCL を対象としたチサゲンレクルユーセルは、CADTH、MSAC、NICE、NoMA による評価が確認された。比較対照はいずれもサルベージ化学療法であり、臨床データはチサゲンレクルユーセルの JULIET 試験および UPENN 試験、サルベージ化学療法の SCHOLAR-1 試験または CORAL 試験が採用された。比較の方法はいずれもナイーブな間接比較であり、B-ALL と同様に、少数例、単群、不十分な追跡期間といった課題が指摘されている。日本では分析評価を伴う費用対効果評価の対象とはなっていないが、成人 DLBCL

を対象としたアキシカブタゲンシロルユーセル（イエスカルタ）の医療技術評価が、CADTH、ICER、MSAC、NICE、NoMA で行われており、CADTH、NICE、NoMA では傾向スコア等を用いて調整した間接比較が採用された。MSAC は調整を行わないナイーブな間接比較を行い、米国 ICER は正式な統計的分析は適当ではないと判断した。疾患モデルを用いた増分 QALY の推定値は、小児 B-ALL（チサゲンレクルユーセル）では成人 DLBCL（チサゲンレクルユーセル、イエスカルタ）よりも評価機関によるばらつきが大きく、試験の追跡期間を超えた外挿が小児疾患ではより大きな不確実性を含むためと考えられた。

Drummond らは、欧米の HTA 機関が細胞治療や遺伝子治療の評価課題にどのように対応しているかについて、8 つの当局（イングランド NICE、スコットランド SMC、カナダ CADTH、イタリア AIFA、フランス HAS、ドイツ G-BA、スペイン AEMPS、米国 ICER）が行った細胞・遺伝子治療 9 製品（Kimryah、Yescarta、Luxturna、Strimvelis、Imlygic、Alofisel、Provenge、Glybera、Zolgensma）の 10 の適応症（Kimryah は DLBCL および ALL）についての 46 件の HTA レポートを分析調査し報告している<sup>36</sup>。全体として、臨床的エビデンスの評価については共通性が高かった。これは、薬事承認に用いられたデータを利用して HTA 評価を行っているためと考えられる。70% の評価でサロゲートエンドポイントが用いられ、85% が単群試験を用いており、78% が長期の有用性を外挿していた。評価に用いられた臨床試験のサンプルサイズは 18 から 436 であり、追跡期間は 12 カ月から 4 年であった。

経済的エビデンスの主な課題について各国 HTA 機関の評価のばらつきは大きかった。評価レポートの多くで言及された点は、不確実性（87%）、疾患の重篤性（76%）、生存期間の大幅な改善（63%）、代替治療の欠如（61%）、割引率（50%、割引率の感度分析は 39%）であった。一方で、言及が少なかった点は、オプションの価値（0%）、科学的波及効果（9%）、家族や介護者にとっての価値（30%）、償還・支払モデル（46%）であった。アウトカムに連動した支払モデル（追加的データ収集を含む）は、Strimvelis（イタリア）、Kymriah（フランス、ドイツ、イタリア、スペイン、英国）、Yescarta（イタリア、スペイン、英国）、Luxturna（スペイン）、Zolgensma（ドイツ、イタリア、スペイン、英国）で用いられた。

## 7. ガイドライン（間接比較部分）

各国の HTA 機関（日本 C2H、英国 NICE、欧州 EUnetHTA、ドイツ IQWiG、カナダ CADTH、米国 ICER）が公表している分析ガイドラインの間接比較に関する記述を紹介する。EUnetHTA は欧州で始まった Joint Clinical Assessment（HTA の文脈における臨床的有用性の共同評価）のために多くのガイドラインを作成・改訂しており、直接・間接比較のガイドラインは包括的で詳細なものである。いずれのガイドラインも直接比較の RCT のメタアナリシスによる統合が最良のエビデンスを提供するとしているものの、比較対照との直接比較の RCT が存在しない場合も多く、間接比較の手法によるエビデンスの統合も容認している。中でも IQWiG の要求が最も厳しく、統合に用いる全ての試験で IPD が利用可能な場合を除き、アンカーのない間接比較を容認していない。EUnetHTA も MAIC や STC によるアンカーのない間接比較を推奨しておらず、アンカーのない状況では IPD の傾向スコア等を用いた比較を推奨している。なお、EU では、2025 年から Joint Clinical Assessment が開始される予定であり、EUnetHTA と同様の間接比較に関するガイドライン

が2024年3月25日付で欧州委員会のウェブサイトにも公示されている。

NICE と EUnetHTA はガイドラインで間接比較の具体的な手法や推奨事項に言及しており、NICE はガイドラインおよび TSD で潜在的に利用可能な間接比較の手法の詳細や推奨事項を示している。日本の C2H のガイドラインは、英国を含め他国のガイドラインと大きく異なる点はないが、IPD が一部の試験で利用可能な場合、MAIC を推奨し、次に NMA を推奨している点が特徴的である。

最近、田中ら<sup>37</sup>は、間接比較に関する各国の HTA 機関、EUnetHTA、FDA/EMA、ISPOR など 10機関の 68本のガイドラインやベストプラクティスのレビューを行い、各ガイドライン等では、バイアスを最小限に抑え、科学的信頼性と透明性を確保する方法が推奨されている。当局によって具体的な推奨事項に違いはあるが、重要な要素には重複が確認されている。個別の事例では間接比較の適合性と受け入れ可能性は個別の状況に依存し、複数の要因に左右されると結論している。

## 7.1 日本：中央社会保険医療協議会における費用対効果評価の分析ガイドライン 2024 年度版<sup>1</sup>

5.7 間接比較を行う場合、以下を原則的な考え方とする。

5.7.1 個人レベルのデータを用いることができる場合、MAIC (Matching adjusted indirect comparison : マッチング調整された間接比較) 等を用いて患者背景の違いなど適切な補正を行う。

5.7.2 個人レベルのデータを用いることができない場合、あるいは手法としてより望ましいと考えられる場合は、複数の RCT を用いたネット ワークメタアナリシスなどを実施する。

5.7.3 「5.7.1」および「5.7.2」においては、アンカーのある分析手法に基づくことが望ましい。

5.7.4 間接比較を行う場合は、間接比較を可能とする仮定や前提条件 (疾患、重症度、患者背景等の異質性や試験の同質性など) についても十分に説明しなければならない。

5.7.5 個人レベルデータも比較試験結果も利用できない場合、その他の代替手段が存在しなければ naïve indirect comparison (単純な間接比較) を用いて評価をすることもできる。この場合、群間の比較可能性について、より詳細な説明が必要である。

## 7.2 英国：NICE health technology evaluations: the manual. 2023

英国では約 20 年前より NICE が医療技術などの保険償還の可否、並びに新薬の薬価に関連した費用対効果評価および総合的評価 (アプレイザル) を実施しており、制度のみならず、評価の事例や技術的側面についても参考になるものが多い。NICE の DSU は、様々な TSD を作成しており、現在までに 24 の TSD が発行されている。我々のタスクフォースでは、NICE が発行している TSD の Evidence Synthesis シリーズ (TSD1~7) をもとに、間接比較を含む NMA を中心としたエビデンス統合の方法論および留意事項を要約した報告書を作成した<sup>3</sup>。また、TSD18 をもとに、比較対照と直接比較を行った試験が存在せず、評価対象についての自社の試験では IPD があり、比較対照の他社の試験では AgD しか利用できない状況で、試験間の効果修飾因子の分布の違いを調整して間接比較を行う手法 (MAIC および STC) についても報告書<sup>4</sup>を作成している。ここ

では、NICE の医療技術評価ガイドライン the manual<sup>41</sup> と過去の製薬協成果物<sup>4</sup>を基に間接比較をテーマとした技術的文書 TSD18 の内容を紹介します。

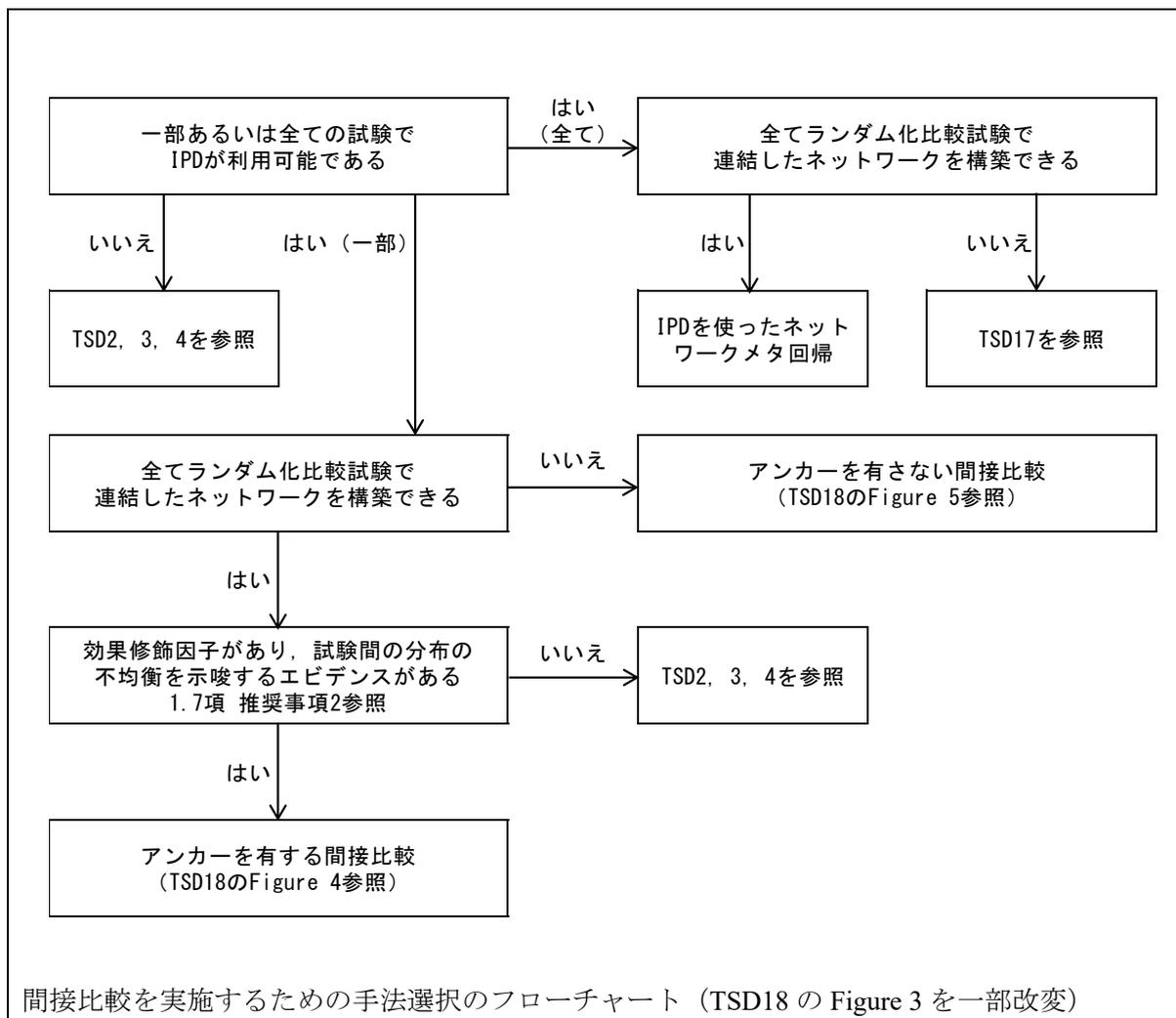
- 3.4.11 単一の RCT のみで評価されていない技術を比較する場合、適切であれば、複数の RCT から得られたデータを NMA とともに提示する。間接比較のみから得られた相対効果の推定値を検討する際、直接比較の欠如に伴う追加的な不確実性を考慮する。NICE は、技術支援文書 (TSD) エビデンス統合シリーズに示された NMA の方法を推奨する。
- 3.4.14 少数の試験から構成されるネットワークでは、間接比較は系統的バイアスに非常に脆弱である。臨床試験間の効果修飾因子が不均衡な場合、ネットワークにおける PAIC を考慮する。PAIC には、ネットワーク内の少なくとも 1 つの臨床試験で個々の患者データ (IPD) が必要である。PAIC の限界を認識し、可能であれば系統的バイアスの大きさを検討する (TSD 18 参照)。
- 3.4.20 調整間接比較または NMA の枠組みを用いてエビデンスを統合する場合は、常に試験のランダム化を維持しなければならない。異なる RCT の単一の治療群同士の結果を比較することは容認できない。このような比較は、観察研究的であり、不確実性が増すことになる。NMA では、直接比較と間接比較を区別し、直接比較によるペアワイズ比較の結果を、完全な NMA に基づく混合比較の結果と合わせて提示する。
- 3.4.21 NMA において、方法論的な質または組み入れ試験の規模に懸念がある場合は、バイアス調整を考慮する (TSD 3 参照)。ペアワイズメタアナリシスや NMA に含めるのに十分な試験がない場合、解析は個々の試験を批判的に評価し、その結果を示す叙述的概要に限定せざるを得ないことがある。このような場合、アプレイザル委員会は相対的な臨床効果について結論を出す際には、特に慎重になる。
- 3.4.22 エビデンス統合の手法は、評価の文脈に適したものでなければならない。基礎となる仮定、目的、選択した手法の長所と限界を説明し、正当化する必要がある。

## 7.2.1 TSD18 : NICE に提出する際の母集団調整間接比較法 2016 の推奨事項

前節までに記載した内容を踏まえ、現時点における MAIC や STC の適用に対する推奨事項を以下に記載する。また、状況に応じて適切な手法を選択するためのフローチャートを示したので、合わせて参照いただきたい。なお、本フローチャートは現時点におけるものであり、今後、シミュレーション研究を始め、母集団調整法のさらなる性能評価の結果なども踏まえて適切な手法を選択する必要があると考える。

推奨事項 1	共通の比較対照群を有する連結したネットワークを構築できる場合には、アンカーを有する間接比較の実施を検討すべきである。また、ランダム化比較試験に基づく連結したネットワークを構築できない場合や、単群試験を伴うネットワークを構築せざるを得ない場合は、アンカーを有さない間接比較の実施が有用な場合がある。
推奨事項 2	連結したネットワークに対して母集団を調整する手法を用いる場合、標準的な手

	<p>法に比べてバイアスの小さい治療効果が得られる可能性が高いと考える根拠を示す必要がある。具体的には、経験的なエビデンスや生物学的妥当性などに基づき、適切に変換した尺度上で、ある因子を効果修飾因子と見做すことにした根拠を示すとともに、試験の集団間の効果修飾因子の分布について、相対治療効果の推定にも大きな影響を及ぼす（バイアスを伴う）ような不均衡があることを定量的に示す必要がある。</p>
推奨事項 3	<p>連結していないネットワークに対して母集団を調整する手法を用いる場合、相対治療効果の推定に関連して、十分な正確度を有して絶対治療効果を予測できる根拠を示すとともに、調整しきれない可能性があるバイアスの程度や範囲を示す必要がある。</p>
推奨事項 4	<p>アンカーを有する間接比較において、MAICのように集団を重み付けする解析では、バイアスを制御するため、試験間での不均衡の有無にかかわらず、すべての効果修飾因子を調整した方が良い。しかしながら、オーバーマッチングにより精度が低下することを避けるため、予後因子は調整しない方が良い。予後因子は、相対治療効果の推定に影響を及ぼさないからである。また、STCのように回帰モデルに基づく方法では、バイアスを制御するため、試験間で不均衡が認められた効果修飾因子を調整した方が良い。それに加え、モデルの適合度の向上のため、不均衡が認められていない効果修飾因子や予後因子も調整した方が良い場合がある。仮に追加した効果修飾因子や予後因子により、アウトカムのばらつきのかなりの部分の説明が可能な場合、治療効果の推定精度を向上させることができる。一方、アンカーを有さない間接比較では、信頼性を有して絶対治療効果を予測するため、MAIC と STC のいずれの手法を用いる場合にも、すべての効果修飾因子と予後因子を調整した方が良い。</p>
推奨事項 5	<p>間接比較は、アウトカムに対して通常用いられるのと同じリンク関数を用いて、線形予測子の尺度上で実施することが推奨される。</p>
推奨事項 6	<p>いかなる治療の比較においても、ターゲット集団は明確に規定しておく必要があり、母集団を調整する手法を用いる場合には、そのターゲット集団における相対治療効果を推定する必要がある。</p>
推奨事項 7	<p>報告要件を厳密に満たすことが推奨される。具体的には、共変量の分布の確認、効果修飾因子であるかどうかに関するエビデンス、(MAIC を適用する場合) 重みの分布、不確実性を表す適切な尺度 (推定値の信頼区間など)、間接比較における標準的な解析手法を適用した際に得られる推定値などである。</p>



### 7.3 欧州：エビデンスの統合 直接比較と間接比較 /EU Commission 2024

欧州医療技術評価ネットワーク (EUnetHTA) が 2022 年に公開した方法論ガイドライン D4.3.2 直接・間接比較<sup>42</sup>は、以前の EUnetHTA ガイドライン (2013 年、2015 年更新) の更新版であり、直接・間接比較のために現在利用可能な方法、その基礎となる仮定、長所と短所を述べ、データ状況に対して適切と思われる手法を示すものである。実務的ガイドライン D4.3.1 直接・間接比較<sup>43</sup>は、評価者／共同評価者を対象とし、実務で使用するための追加的でより詳細な助言を与える。ガイドラインの目的は、提出された直接・間接治療比較が EUnetHTA 加盟国により受理されるべきか否かについて明確な勧告を行うことではなく、各加盟国が、医療技術開発者が提出する JCA 報告書に基づき、直接・間接比較の妥当性を自ら判断できるようにすることである。実務的ガイドラインの各項には、直接または間接比較によって統合されたエビデンスが提出される際に、報告書に含めるべき方法論的詳細に関する要件を示している。これらを含む EUnetHTA のガイダンスは、2024 年 3 月に欧州委員会 (EU Commission) により正式に承認された<sup>44</sup>。

複数のエビデンス源が存在する中で、1 つ以上の既存の介入と比較した新たな介入の相対的な有効性や効果进行评估するためには、エビデンス統合のための適切な方法を用いるべきであ

る。RCTは、治療効果の推定のゴールドスタンダードであり、可能な限りエビデンス統合に用いるべきである。本ガイドラインの目的は、直接的・間接的治療比較に最も一般的に用いられている方法について、その基礎となる仮定、長所、短所を含めて説明することである。エビデンスを統合するすべての方法（直接比較と間接比較）は、交換可能性という基本的な前提に基づいているため、含まれる試験データの十分な類似性と十分な均質性が必要である。間接比較の場合、追加的な前提は十分な一貫性があることである。これらの仮定のいずれかに違反した場合、エビデンス統合の結果が、治療効果について意味のある推定値を与える可能性は低い。エビデンス統合の一般的な選択肢には、固定効果モデルまたはランダム効果モデルの使用、頻度論的手法またはベイズ法の適用などがある。IPDの使用は、適切な統計解析やサブグループ解析を実行できる可能性があるため、推奨される。間接比較は不確実性が大きいいため、一般には直接比較が望ましい。間接比較が必要な場合は、ランダム化を尊重した調整された間接比較のみが一般的には適切であり、これにはエビデンスネットワークが連結されている必要がある。調整された間接比較のための有用なアプローチには、Bucher法、NMAがある。

調整された間接比較が実行不可能な場合、別の方法が利用可能であるが、これらがランダム化の欠落によるバイアスを適切に除去するとは考えにくい。類似性の仮定が満たされない場合、ネットワークが連結され、含まれる試験の一部についてIPDが利用可能であれば、母集団調整間接比較の方法が考慮される。これらの方法では、調整に関連する効果修飾因子がすべて測定されている必要がある。しかし、これは検証不可能で達成できないことが多い。したがって、母集団調整間接比較から得られた結果の頑健性を入念に検討することが不可欠である。調整のためのモデルと共変量の選択戦略は、事前に規定され、透明性のある基準に基づいていなければならない。

観察されたエビデンスや単群試験のような非ランダム化データのある状況、あるいは断絶されたネットワークの場合、交絡を適切に調整できる方法を適用するためには、個々のIPDに完全にアクセスできることが必要である。これらの方法では、調整に関連するすべての共変量（予後因子、交絡因子または効果修飾因子）が測定されていることが必要である。しかし、これは検証不可能であり、達成できないことが多い。したがって、調整のためのモデルおよび共変量選択戦略が事前に規定され、透明な基準に基づいていることが必須である。傾向スコアの使用は頻繁に適用される方法である。必要な仮定は、十分なポジティビティ、十分なオーバーラップ、十分なバランスである。これらの仮定のいずれかが満たされない場合、交絡に対する適切な調整は不可能であり、対応する解析の結果は治療効果の意味のある推定値を提供しそうにならない。傾向スコア法でトリミングを行う場合、最終的な対象集団を詳細に記述しなければならない。推定治療効果の頑健性を評価するために、未測定の交絡因子の潜在的影響を探る感度分析の方法も適用できる。

方法論の選択は、最終的には文脈に固有であり、利用可能なデータに適切でなければならない。多くの場合、相対的有効性の不偏推定を行うために、本ガイドラインに示された方法のいずれを用いても理想的な条件とはならない。そのため、推論を行う際には、基礎となる仮定を慎重に検討する必要がある。使用された方法論的アプローチ、違反する可能性のある仮定や仮定からの逸脱がもたらす結果の不確実性を批判的に評価するために、この分野に特化した専門

知識を持つ統計家の意見を聞くことが推奨される。

#### ランダム化に基づく要約データのメタアナリシスによる間接比較のポイント

- (a) 間接比較を用いて、治療効果を適切に推定するために直接比較を用いるときよりも多くの仮定を必要とする。したがって、可能な限り直接比較を優先すべきである。間接比較を用いる場合は、試験内のランダム化を尊重するため、調整された（すなわち同一の治療をアンカーとする）間接比較のみが適切である。
- (b) 集計データの調整された間接比較は、十分な類似性（similarity）と十分な均質性（homogeneity）に加えて、直接エビデンスと間接エビデンスの十分な一貫性（consistency）を前提とする。これらの前提となる仮定の妥当性を評価し、報告すべきである。
- (c) 類似性、均質性、一貫性を十分に仮定できない場合、治療効果の意味のある推定値が得られる可能性が低いため、調整された間接比較を用いるべきではない。
- (d) 間接比較の有用なアプローチには、Bucher 法、頻度論またはベイズ流の NMA などがある。

#### ランダム化に基づく（アンカーのある）PAIC による間接比較のポイント

- (a) 類似性の性質が成立しない場合、直接比較や間接比較のための通常の方法は無効である。このシナリオでは、ネットワークが連結しており、そのような調整がバイアスを減少させる可能性が高いという十分なエビデンスが先験的にあれば、PAIC を代替アプローチとして考慮することができる。そのためには、モデルや共変量の選択戦略を事前に規定し、透明性のある基準に基づくべきである。
- (b) 試験間の不均衡を調整するために、いくつかの試験で少なくとも 1 つの治療群から IPD を入手することが必要である。
- (c) 相対効果の統合のための PAIC（すなわち、エビデンスの連結ネットワークにおける重み付けモデル（MAIC）またはアウトカム回帰モデル（STC および ML-NMR））は、モデルが正しく特定されていること、特にすべての関連する効果修飾因子がモデルに含まれているという仮定に依存する。
- (d) PAIC から推定される治療効果には、いくつかの原因から生じるさらなる不確実性が伴う。不確実性が大きいと、大きな効果推定値が必要となる。これは、シフト仮説の検定によって正式に達成することができる。
- (e) PAIC によって治療効果が推定される対象集団は、詳細に記述されなければならない。

#### ランダム化に基づかない間接比較のポイント

- (a) 交絡の調整方法は、IPD への完全なアクセスに基づいている。すべての方法は、測定されていない交絡因子がなく、測定されていない効果修飾因子がないことを必要とする。
- (b) STC や MAIC のような PAIC は、切断されたネットワークに適用される場合、全ての IPD を利用する方法よりも強い仮定を必要とし、交絡の調整には一般的に十分ではない。
- (c) 交絡を調整するためのモデルや共変量の選択戦略は、あらかじめ規定され、透明性のある基準に基づくべきである。
- (d) 傾向スコアの適用には、十分なポジティブティ、十分なオーバーラップ、考慮される集団の十分なバランスが必要である。これが達成できない場合、交絡の適切な調整は不可能であり、対応する解析の結果は治療効果の意味のある推定値を提供する可能性は低い。

- (e) 傾向スコア法をトリミングとともに適用する場合、最終的な対象集団を詳細に記述しなければならない。
- (f) 非ランダム化データから推定される治療効果は、多くの原因から生じる追加的な不確実性を伴う。不確実性が大きい場合、大きな効果推定値が必要となる。

#### 7.4 ドイツ：一般的方法 第7版 IQWiG 2023

ドイツ IQWiG のガイドライン<sup>45</sup>は直接比較を強く推奨している(9.3.8 章 Indirect Comparisons)。直接エビデンスと間接エビデンスを組み合わせる手法 (NMA など) には、方法論的に未解決の問題があるため通常は推奨されない。未調整の間接比較 (異なる試験の治療群のナイーブな比較) は許容されない。共通の比較対照群 (アンカー) がない場合、強い仮定に依存する治療効果のモデリング (STC など) や観察研究で用いられる検証不可能な仮定を必要とする因果モデルの手法 (MAIC など) による間接比較は受け入れられない。適切な共通の比較対照群をアンカーとして調整された間接比較のみが受け入れ可能である。Bucher 法や適切な NMA がそれに含まれる。適切な共通の比較対照群がない場合には、完全な IPD 情報が利用できる場合にのみ解釈可能な比較が可能となる。この場合、非ランダム化研究における比較のための方法を使用することができるが、関連するすべての交絡因子を系統的に同定し、それらが収集されているなどの追加の要件も必要となる。ペアワイズメタ解析の十分な類似性と均質性の仮定は、ここでも満たされなければならないが、それに加えて、NMA では、直接エビデンスと間接エビデンスから推定される効果が一貫していることが要求される。基本的な仮定のひとつまたはいくつかを満たされているかどうかにより重大な疑問が存在する場合は、間接的な比較を使用すべきではない。感度分析では、これらの問題を慎重に検討する必要がある。

#### 7.5 カナダ：医療技術の経済評価のためのガイダンス CADTH 2018

カナダ CADTH が公表しているガイダンス<sup>46</sup>によると、RCT のみならず RWD 等の利用可能な全ての情報源からのエビデンスを統合できる方法論的アプローチの開発は、現在進行中の研究分野である。利用可能な全てのデータソースを統合する適切な方法がない場合、メタアナリシスや NMA を使用して、目的に適合し (fitness for purpose)、信頼性 (credibility) があり、一貫性 (consistency) があると評価されたデータを統合することに焦点を当てることができる。ヘルスケアの意思決定に情報を提供するという経済評価の文脈では、意思決定問題には2つ以上の比較対照や直接比較のない複数の比較対照が含まれる可能性がある。そのような場合、標準的なペアワイズのメタアナリシスの枠組みよりも、複数の介入を同時に比較し、間接比較の評価が可能な NMA の枠組みが選択されることができる。他の考慮事項と同様に、分析の目的への適合性、信頼性、一貫性を考慮して判断する必要がある。

## 7.6 米国：ICERによる価値評価のフレームワーク 2023-2026

米国 ICER が公表しているフレームワーク<sup>47</sup>によると、他国の公的 HTA 機関と同様に、直接比較の RCT のメタアナリシスがバイアスの小さいエビデンスを提供するとしている。しかし、承認直後に実薬対照との直接比較の RCT が存在することはまれであり、間接比較が必要になる。また、長期のアウトカムについてはコホート研究や患者/薬剤レジストリも有用な情報を提供する可能性があり、エビデンスについて包括的なアプローチをとると表明している。具体的には、評価対象と比較対照の直接比較試験がない場合、ICER は間接比較（例えば、NMA や MAIC）の手法によりエビデンスを統合することが多い。間接比較は、興味の対象となる治療法の臨床試験のデータを統合するために、共通の比較対照や各試験で収集されている効果修飾因子等の分布に依存することがある。ICER の統合手法の選択は、意思決定課題と利用可能なエビデンスに依存する。全ての報告書において、ICER は使用する統合手法の選択の根拠を示し、その手法を明確に記述している。間接比較による推定値は経済モデルへの入力値としても用いることができる。

## 8. 考察

4章から7章の内容を踏まえ、タスクフォースによる考察を以下に記載する。

### 8.1 海外の事例に基づく考察

ザヌブルチニブ（TA833）の事例では、比較対照との直接比較試験が存在せず、さらにアンカーとなる共通の対照治療も存在しなかったため、アンカーなしの間接比較が実施された。その際、複数の潜在的な予後因子、効果修飾因子の調整が検討された。MAIC は有効サンプルサイズが少なくなり調整不可能な差異が生じる可能性がある一方、STC は相対的により多くのサンプルサイズを用いて共変量調整できる利点があった。最終的に委員会は、MAIC と STC はそれぞれ結果の不確実性および解釈に限界があるが、当初から設定していた MAIC の方が透明性が高く好ましいアプローチであると結論付けられた。また治療群間の生存時間のハザード比について、MAIC 法による信頼区間が STC 法によるものと重なりが大きいことから、2つの手法による結果にある程度の一貫性があると解釈されている。そのため、方法論の特徴や限界を踏まえながら、複数の手法による解析を行い、それらの結果を踏まえて総合的に判断されていると考えられる。

ブリガチニブ（TA670）の事例では、比較対照との直接比較試験が存在せず、アンカーとなる共通の対照治療を有していたものの、一方の試験の対照治療で治療切り替えが行われていたため、通常のアンカーありの間接比較では明らかにバイアスが生じる状況であった。治療切り替えの影響の調整も試みたが、その影響を完全には排除できなかったため、アンカーなしの間接比較も実施したが、予後因子および効果修飾因子を十分に考慮できていないため、委員会では受け入れられなかった。生存時間に関するアンカーあり MAIC 法、アンカーなし MAIC 法、Bucher 法による間接比較の結果は、生存時間の追跡が十分でないことはいずれも信頼区間が広く、不確実性が高い状況であったが、2治療間で無増悪生存時間の間接比較結果および作用機序が類似していることを踏まえ、生存時間についても同様と考えることがもっともらしいとの結論となった。

テポチニブ（TA789）の事例では、単群試験のみを有している状況で、同一疾患の免疫療法または化学療法で治療された患者の IPD を有する RWD との間接比較、および対象疾患がやや異なる

る過去の論文情報に基づく臨床試験結果との間接比較が行われた。それぞれのアプローチによる結果の解釈に限界があり不確実性を有するが、複数のアプローチによる結果を踏まえて議論が行われた。

## 8.2 解析手法およびシミュレーション研究に基づく考察

上述のとおり、近年アンカーを有さない間接比較を実施するケースが増えている。しかしながら、前提となる仮定の厳しさから、アンカーを有さない間接比較では、PAIC を適用したとしてもバイアスを制御しることができない可能性がある。そこで、アンカーを有さない間接比較では、統計的には感度分析や定量的バイアス分析（Quantitative Bias Analysis : QBA）を実施することが重要となる。例えば、MAIC を適用した結果、治療 B の治療 C に対する相対治療効果が認められたものの、未測定の変絡因子の存在が疑われたとする。その場合、未測定の変絡因子の影響を評価する QBA の実施が有用であり、未測定の変絡因子の影響を考慮したとしても、依然として治療 B の治療 C に対する相対治療効果があることが示されれば、追加的有用性があると強く主張する根拠となりうると考えられる。

また、アンカーを有さない間接比較では、過去の臨床試験、レジストリやデータベースの IPD から比較可能な外部対照群を構築し、治療 B の治療 C に対する相対治療効果を推定する方法も提案されている。薬事申請において、試験計画に際し外部対照群との比較を正当化する条件の一つとして、既に完了した前の相の情報等から、外部対照群の治療に比べて大きな相対治療効果が期待される状況にあることが挙げられる。これは費用対効果評価における追加的有用性の有無の評価でも応用可能な条件といえる。例えば、MAIC を適用した結果、得られた相対治療効果が小さく且つ不確実性が大きい場合、前提となる仮定からの逸脱を考慮すると、比較的容易に相対治療効果の方向が逆転してしまうような状況下では、MAIC の結果のみに基づいて追加的有用性を主張するのは困難かもしれない。一方で、得られた相対治療効果が大きい場合や、相対治療効果が小さかったとしてもその精度が高い場合には、意思決定に利用しうると考えられる。

また、アンカーを有する・有さないにかかわらず、PAIC として MAIC が比較的良好に利用されている。MAIC の適用に際し、モデルに含める効果修飾因子や予後因子は、臨床的重要性と有効サンプルサイズの両方を考慮して決定する必要がある。例えば、多くの効果修飾因子や予後因子を考慮して MAIC を適用した結果、因子の重なりが十分ではなく有効サンプルサイズが小さくなった場合には、臨床的重要性を踏まえて、モデルに含める因子を絞り込むことが必要な場合がある。また、MAIC の適用が適切でない場合は、STC の適用を検討することも重要である。ただし、それぞれ一長一短があるので（5.1 章）、状況に応じて適切と思われる手法を選択する必要がある、二重頑健性推定量も有用な選択肢の一つである。

## 8.3 ガイドラインに基づく考察

各国の HTA ガイドライン等で間接比較に関する記載には濃淡があり、英国 NICE や欧州 EUnetHTA は利用可能な手法の数理的解説や前提条件、留意事項が詳細に述べられている。一方で、カナダ CADTH や日本 C2H は比較的簡素な記載にとどまっており、製造販売業者による分析

を求めない米国 ICER の記載はもっとも簡素なものであった。しかしながら、各国の HTA 機関で方法論的な理解、利用可能性や限界について一定の認識の共有があるように思われた。実際の間接比較の利用状況を調査した研究によると (4.2 章)、英国 NICE での利用件数が最も多い。英国 NICE は間接比較の手法の開発と普及に長年尽力してきたことが理由のひとつであり、また、英国では HTA を新規医療技術へのアクセスの可否判断という意思決定の文脈で使用されることも一因と思われる。すなわち、英国 NICE には、NHS での利用を推奨するか否かについて、利用可能なエビデンスに基づき最良な判断をすることが期待されているのである。日本では、価格調整の大きさを決定するための経済モデルへの入力値として間接比較による定量的な推定値を利用せざるを得ない状況があるという点で、英国よりも難しい状況が生じ得る。特に、細胞・遺伝子治療は、種々の迅速承認プロセスの対象となり得るため、承認時に単群試験のエビデンスのみである場合や長期の臨床エビデンスが得られていない場合も多い。こうした不十分な臨床エビデンスは、各国で HTA の新たな課題となっている。間接比較は、利用可能なデータの制約を受けるとともに、データから検証することが必ずしも容易ではない仮定を前提とする手法ではあるものの、利用可能なエビデンスを最大限活用した形で、今行うべき保健医療の意思決定のための有用なツールのひとつである。総合的評価では、単に間接比較による比較結果が有意であるか否か、前提条件の成立が確認できないが故の不確実性といった技術的妥当性のみ注目するのではなく、利用可能なデータと分析手法の限界を踏まえつつ、イノベーションによる医療の進歩や保健医療制度の持続的発展といったより広範囲な価値評価を踏まえた、熟慮による意思決定が期待される<sup>48</sup>。

## 9. まとめ

本報告書では、費用対効果評価における追加的有用性の評価において利用頻度が高まってきている間接比較に注目し、近年での国内および海外での費用対効果評価の複数の事例を取り上げ、適用されている間接比較の方法およびそれらの解析結果が委員会でどのように議論および評価され最終的な結論に至っているかを調査した結果を掲載した。さらに、最新の公表論文に基づく各方法の理論的な特徴および、各国から公開されている最新の HTA ガイドラインの概要を整理した。

間接比較といっても、実際に間接比較を実施せざるを得ない場面に直面した場合、取りうるオプションは複数あり、その中から、臨床的、統計的および医療経済的観点等を考慮して適切な手法を適用する必要があるが、各手法には長所および短所があり、それぞれの解析結果のバイアスや不確実性の程度を考慮することが肝要である。そのため、計画時に最適な手法を特定するのは困難な場合もあり、取り上げた事例では、複数の手法の結果を確認することでバイアスや不確実性による誤った判断を下すリスクを回避しながら総合的に最終結論に至っているものが多い印象であった。

HTA ガイドラインについては、英国 NICE や欧州 EUnetHTA が各国の HTA 機関と比べると相対的に詳細に述べられている状況であるが、方法論的な理解、利用可能性や限界についての考えなど共通している部分が伺えた。

費用対効果評価のために実施する間接比較は、利用可能なデータの制約を受けるとともに、データから検証することが必ずしも容易ではない仮定を前提とする手法ではあるものの、利用可能

なエビデンスを最大限活用した形で、今行うべき保健医療の意思決定のための有用なツールと考えられるため、本邦で間接比較を実施する実務担当者にとって、本報告書が少しでも有益な資料となれば幸いである。

## 参考文献

1. 国立保健医療科学院 保健医療経済評価研究センター (C2H) . 中央社会保険医療協議会における費用対効果評価の分析ガイドライン 2024 年度版. <https://c2h.niph.go.jp/tools/guideline/> (Accessed: 19 July 2024)
2. NICE Decision Support Unit. Technical support documents. <https://www.sheffield.ac.uk/nice-dsu/tsds> (Accessed: 19 July 2024)
3. 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. NICE TSD Evidence Synthesis シリーズの紹介：ネットワークメタアナリシスを中心として. 2020. [https://www.jpma.or.jp/information/evaluation/results/allotment/nice\\_tsd\\_evidence\\_synthesis.html](https://www.jpma.or.jp/information/evaluation/results/allotment/nice_tsd_evidence_synthesis.html) (Accessed: 19 July 2024)
4. 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. 間接比較において母集団調整法は有用か？－Matching Adjusted Indirect Comparison および Simulated Treatment Comparison－. 2022. [https://www.jpma.or.jp/information/evaluation/results/allotment/indirect\\_comparison.html](https://www.jpma.or.jp/information/evaluation/results/allotment/indirect_comparison.html) (Accessed: 19 July 2024)
5. 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. 医療技術評価における多変量メタアナリシスと代替エンドポイント. 2023. [https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_202305\\_TSD\\_20.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202305_TSD_20.html) (Accessed: 19 July 2024)
6. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to Nice. 2016. (Accessed: 19 July 2024)
7. 国立保健医療科学院 保健医療経済評価研究センター (C2H) . 費用対効果評価の制度化以後 (2019.4-) に選定された品目. <https://c2h.niph.go.jp/results/item.html> (Accessed: 19 July 2024)
8. 国立保健医療科学院 保健医療経済評価研究センター (C2H) . [C2H2003] トラスツズマブ デルクステカン (エンハーツ) <https://c2h.niph.go.jp/results/C2H2003.html> (Accessed: 19 July 2024)
9. 国立保健医療科学院 保健医療経済評価研究センター (C2H) . [C2H2207] アンデキサネット アルファ (オンデキサ) . <https://c2h.niph.go.jp/results/C2H2207.html> (Accessed: 19 July 2024)
10. Bucher HC, Cuyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50 (6) :683-91.
11. Cohen AT, Lewis M, Connor A, Connolly SJ, Yue P, Curnutte J, Alikhan R, MacCallum P, Tan J, Green L. Thirty-day mortality with andexanet alfa compared with prothrombin complex concentrate therapy for life-threatening direct oral anticoagulant-related bleeding. J Am Coll Emerg Physicians Open. 2022;3 (2) :e12655.
12. Costa OS, Connolly SJ, Sharma M, Beyer-Westendorf J, Christoph MJ, Lovelace B, Coleman CI. Andexanet alfa versus four-factor prothrombin complex concentrate for the reversal of apixaban- or rivaroxaban-associated intracranial hemorrhage: a propensity score-overlap weighted analysis. Crit Care.

2022;26 (1) :180.

13. NICE guidance. <https://www.nice.org.uk/guidance> (Accessed: 19 July 2024)
14. Zanubrutinib for treating Waldenström's macroglobulinaemia, Technology appraisal guidance, TA833. Published: 19 October 2022. <https://www.nice.org.uk/guidance/ta833> (Accessed: 19 July 2024)
15. Brigatinib for ALK-positive advanced non-small-cell lung cancer that has not been previously treated with an ALK inhibitor, Technology appraisal guidance, TA670. Published: 27 January 2021. <https://www.nice.org.uk/guidance/ta670> (Accessed: 19 July 2024)
16. Tepotinib for treating advanced non-small-cell lung cancer with MET gene alterations, Technology appraisal guidance, TA789. Published: 18 May 2022. <https://www.nice.org.uk/guidance/ta789> (Accessed: 19 July 2024)
17. Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, Betts KA, Wu EQ. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health*. 2012;15 (6) :940-47.
18. 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. 費用対効果評価における生存時間解析の諸問題. 2022. [https://www.jpma.or.jp/information/evaluation/results/allotment/c\\_e\\_a\\_problems.html](https://www.jpma.or.jp/information/evaluation/results/allotment/c_e_a_problems.html) (Accessed: 19 July 2024)
19. Macabeo B, Rotrou T, Millier A, François C, Laramée P. The Acceptance of Indirect Treatment Comparison Methods in Oncology by Health Technology Assessment Agencies in England, France, Germany, Italy, and Spain. *Pharmacoecoon Open*. 2024;8 (1) :5-18.
20. van Beekhuizen S, Che M, Monfort L, Hashim M, Azough A, Kubitz N, Griffin A, Price M. Analyzing Indirect Treatment Comparisons in Eunetha Assessments: Lessons Learned for the Implementation of EU Joint Clinical Assessments? *Value in Health*. 2024;27 (6) :S1.
21. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis. 2011; last updated April 2012. <http://www.nicedsu.org.uk> (Accessed: 19 July 2024)
22. Phillippo DM, Dias S, Elstada A, Ades AE, Welton NJ. Population Adjustment Methods for Indirect Comparisons: A Review of National Institute for Health and Care Excellence Technology Appraisals. *International Journal of Technology Assessment in Health Care*. 2019;35 (3) :221-8.
23. Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*. 2012;20 (1) :25-46.
24. Caro JJ, Ishak KJ. No head-to-head trial? Simulate the missing arms. *PharmacoEconomics*. 2010;28 (10) :957-67.
25. Ishak KJ, Proskorovsky I, Benedict A. Simulation and Matching-Based Approaches for Indirect Comparison of Treatments. *PharmacoEconomics*. 2015;33 (6) :537-49.
26. Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data. *Stat Med*. 2006;25 (12) :2136-59.
27. Jackson C, Best AN, Richardson S. Hierarchical related regression for combining aggregate and

- individual data in studies of socio-economic disease risk factors. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2008;171 (1) :159-78.
28. NICE. NICE real-world evidence framework (ECD9) 2016.  
<https://www.nice.org.uk/corporate/ecd9/chapter/overview> (Accessed: 19 July 2024)
  29. Park JE, Campbell H, Towle K, Yuan Y, Jansen JP, Phillippo D, Cope S. Unanchored Population-Adjusted Indirect Comparison Methods for Time-to-Event Outcomes Using Inverse Odds Weighting, Regression Adjustment, and Doubly Robust Methods With Either Individual Patient or Aggregate Data. *Value Health*. 2024;27 (3) :278-86.
  30. Daniel R, Zhang J, Farewell D. Making apples from oranges: comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biom J*. 2021;63 (3) :528-57.
  31. Cheng D, Tchetgen ET, Signorovitch J. On the double-robustness and semiparametric efficiency of matching-adjusted indirect comparisons. *Res Synth Methods*. 2022;14 (3) :438-42.
  32. Remiro-Azócar A, Heath A, Baio G. Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data. *Res Synth Methods*. 2022;13 (6) :716-44.
  33. Abrams K. *2020 Sources and Synthesis of Evidence; Update to Evidence Synthesis Methods*. NICE Decision Support Unit. 2020.
  34. Phillippo DM, Dias S, Ades AE, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study. *Stat Med*. 2020;39 (30) :4885-911.
  35. Remiro-Azocar A, Heath A, Baio G. Methods for population adjustment with limited access to individual patient data: a review and simulation study. *Res Synth Methods*. 2021;12 (6) :750-75.
  36. Drummond M, Ciani O, Fornaro G, Jommi C, Dietrich ES, Espin J, Mossman J, de Pouvourville G. How are health technology assessment bodies responding to the assessment challenges posed by cell and gene therapy? *BMC Health Serv Res*. 2023;23 (1) :484.
  37. 田中司朗, 大庭幸治, 吉村健一, 手良向聡. 代替エンドポイントの評価のための 統計的基準とその適用事例. *計量生物学* Vol. 31, No. 1, 23-48 (2010) .
  38. 国立保健医療科学院 保健医療経済評価研究センター (C2H) . 費用対効果評価報告書 チサゲシレクルユーセル/B-ALL (キムリア R) 2021.  
<https://c2h.niph.go.jp/results/C2H1902.html> (Accessed: 19 July 2024)
  39. Gye A, Goodall S, De Abreu Lourenco R. A Systematic Review of Health Technology Assessments of Chimeric Antigen Receptor T-Cell Therapies in Young Compared With Older Patients. *Value Health*. 2022;25 (1) :47-58.
  40. Tanaka S, Igarashi A, Moor RD, Li N, Hirozane M, Hong LW, Wu DBC, Yu DY, Hashim M, Hutton B, Tantakoun K, Olsen C, Fashami FM, Samjoo IA, Cameron C. A Targeted Review of Worldwide Indirect Treatment Comparison Guidelines and Best Practices. *Value Health* 2024:S1098-3015.
  41. NICE health technology evaluations: the manual. Published: 31 January 2022.  
<https://www.nice.org.uk/process/pmg36/chapter/evidence-2> (Accessed: 19 July 2024)
  42. EUnetHTA 21. Individual Practical Guideline Document, D4.3.1: Direct and Indirect Comparisons. 2022.

- <https://www.eunetha.eu/wp-content/uploads/2022/12/EUnetHTA-21-D4.3.1-Direct-and-indirect-comparisons-v1.0.pdf> (Accessed: 19 July 2024)
43. EUnetHTA 21. Methods Guideline, D4.3.2: Direct and Indirect Comparisons Version 1.0. 2022.  
<https://www.eunetha.eu/wp-content/uploads/2022/08/EUnetHTA-21-Deliverable-D4.3.2-Methodological-Guideline-on-Direct-and-indirect-comparisons-V1.0.pdf> (Accessed: 19 July 2024)
44. Directorate-General for Health and Food Safety. Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons Version 1.0. 2024.  
[https://health.ec.europa.eu/latest-updates/methodological-guideline-quantitative-evidence-synthesis-direct-and-indirect-comparisons-2024-03-25\\_en](https://health.ec.europa.eu/latest-updates/methodological-guideline-quantitative-evidence-synthesis-direct-and-indirect-comparisons-2024-03-25_en) (Accessed: 19 July 2024)
45. IQWiG. General Methods Version 7.0. 2023.  
[https://www.iqwig.de/methoden/general-methods\\_version-7-0.pdf](https://www.iqwig.de/methoden/general-methods_version-7-0.pdf) (Accessed: 19 July 2024)
46. Canada's Drug Agency. Guidelines for the Economic Evaluation of Health Technologies: Canada - 4th Edition. Last Updated: May 13, 2021.  
<https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition>  
(Accessed: 19 July 2024)
47. ICER. Value Assessment Framework. 2023.  
[https://icer.org/wp-content/uploads/2024/02/ICER\\_2023\\_2026\\_VAF\\_For-Publication\\_021324.pdf](https://icer.org/wp-content/uploads/2024/02/ICER_2023_2026_VAF_For-Publication_021324.pdf)  
(Accessed: 19 July 2024)
48. 日経チャンネル. 第35回製薬協政策セミナー. 2024.  
<https://channel.nikkei.co.jp/jpma2024/> (Accessed: 19 July 2024)

## 2024 年度データサイエンス部会 KT7

### 作成担当者

河田 祐一	グラクソ・スミスクライン株式会社*
直井 一郎	住友ファーマ株式会社
町田 光陽	塩野義製薬株式会社
吉田 瑞樹	ファイザーR&D 合同会社*

### レビュー担当者

秋澤 優太	ノバルティスファーマ株式会社
荒西 利彦	日本イーライリリー株式会社
東 美恵	エーザイ株式会社
蔡 志紅	日本イーライリリー株式会社
中島 章博	帝人ファーマ株式会社*

### \*TF リーダー

### 監事

酒井 弘憲	エーザイ株式会社
-------	----------