間接比較において母集団調整法は有用か？

－Matching Adjusted Indirect Comparison 及び Simulated Treatment Comparison－

日本製薬工業協会

医薬品評価委員会　データサイエンス部会

2021 年度継続タスクフォース 3

医療経済評価に関連する統計的手法の整理と紹介

Ver 1.0

2022 年 1 月

本報告書の中で Technical Support Document 18（Decision Support Unit, National Institute for Health and Care Excellence）を引用／参考にした記載は，本タスクフォースの解釈であり，NICE には一切の責任はない。また TSD18 の全てを紹介したわけではなく，一部を抜粋し紹介している。それらの紹介にあたり，タスクフォースによる解釈や追加の説明を加えている。これらは本タスクが独自にまとめたものであり，NICE から見解を得たものではない。

目次

はじめに..... 1

1. 母集団調整法の内容とその理論的背景..... 2

1.1 状況設定.....2
1.2 解析手法を用いる際に前提となる仮定（全ての試験で IPD が利用可能な場合）.....5
1.2.1 標準化や一般化を用いる際に前提となる仮定.....5
1.2.2 較正法を用いる際に前提となる仮定.....7
1.3 MAIC.....8
1.4 STC.....9
1.5 その他の手法.....11
1.5.1 ネットワークメタ回帰を拡張した手法.....11
1.5.2 その他の再重み付け解析.....12
1.6 MAIC や STC の性質.....13
1.6.1 MAIC やSTC を用いる際に前提となる仮定.....13
1.6.2 尺度の重要性と効果修飾因子との関連.....16
1.6.3 適切なターゲット集団における治療効果の推定.....17
1.6.4 不確実性.....17
1.7 推奨事項.....18

2. シミュレーション研究の紹介.....21

2.1 アンカーを有する間接比較.....21
2.1.1 シミュレーションの設定.....21
2.1.2 シミュレーション結果の要約.....25
2.2 アンカーを有さない間接比較.....26
2.2.1 シミュレーションの設定.....27
2.2.2 シミュレーション結果の要約.....27

3. 母集団調整法の適用事例.....29

3.1 MAIC と STC の文献での適用例.....29
3.1.1 MAIC の文献での適用例.....29
3.1.2 アンカーを有する間接比較とアンカーを有さない間接比較.....32
3.1.3 複数の研究の利用可能性.....32
3.1.4 より大きな治療ネットワーク.....33
3.1.5 有効サンプルサイズと重み分布.....34
3.1.6 マッチング変数の選択.....34
3.1.7 尺度の選択.....35
3.1.8 STC の文献での適用例.....35
3.2 NICE 技術評価（TA）での適用例.....36
3.2.1 レビューの概要.....36

3.2.2 レビューの結果……37
3.2.3 臨床領域別の使用状況……41
3.2.4 アウトカムの種類……42
3.2.5 母集団調整法……42
3.2.6 アンカーを有する間接比較とアンカーを有さない間接比較……43
3.2.7 調整された共変量……43
3.2.8 より大きな治療ネットワーク……44
4. ディスカッション……45
おわりに……48
APPENDIX : Rによる実装例……49
参考文献……62

はじめに

本邦では令和元年より医薬品及び医療機器の価格調整を目的として，費用対効果評価制度が本格導入された(1)。英国では約 20 年前より NICE（National Institute for Health and Care Excellence）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及び総合的評価（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(2)。NICE の DSU（Decision Support Unit）は，様々な技術的文書（Technical Support Document：TSD）を作成しており，現在までに 21 の TSD が発行されている(3)。本タスクフォースでは NICE の TSD で扱われる主題を整理して，国内で費用対効果評価を実践する際に有用と思われる情報を提供する活動を行ってきた。

費用対効果評価を実施する場合，興味のあるターゲット集団（例えば，ある疾患に関して本邦のレジストリに登録される患者集団）における新薬の既存薬に対する相対治療効果を推定する必要がある。しかしながら，費用対効果評価を実施する段階で，新薬と（比較検討すべき全ての）既存薬を直接比較した試験（head to head 試験）のデータが利用可能な状況は稀である。そこで，head to head 試験がない場合でも，複数の試験の結果を統合して，興味のある新薬の既存薬に対する相対治療効果を推定するため，間接比較やネットワークメタアナリス（Network Meta Analysis：NMA）がよく実施されている。

標準的な間接比較や NMA では，各試験の公表論文などから得られる集団レベルあるいは群レベルのデータ（Aggregate Data：AgD），例えば，アウトカムやベースライン特性の各群の平均や標準偏差（連続変数の場合）または割合（2 値変数の場合）などを用いて相対治療効果の推定を行う。間接比較や NMA は非常に有用な手法であるものの，これらの手法を適用するにはいくつかの仮定が必要となる。その一つが，間接比較や NMA に用いる全ての試験（更には興味のあるターゲット集団）間で，効果修飾因子の分布が同じであるという仮定である。この仮定が成立しない場合，標準的な間接比較や NMA を行っても妥当な治療効果を推定することはできない。対処法の一つとして，AgD に対するネットワークメタ回帰が用いられる場合があるものの，エコロジカルバイアスが生じる可能性がある(4)。

近年，間接比較や NMA に用いる複数の試験のうち，一部の試験（例えば，自社の試験）では個別患者データ（Individual Patient Data：IPD）を利用できることを利用して，試験間の効果修飾因子の分布の違いを調整して治療効果を推定する手法が提案されている［母集団調整法（Population adjustment method）と呼ばれる］。本手法は，近年適用事例が増えており，本邦における費用対効果評価制度の本格導入後の評価対象品目でも適用事例が認められていることから，現時点で本手法の理解を深めておくことは非常に有益であると考える。

そこで，本報告書では母集団調整法についての解説を行う。1 章では母集団調整法の内容とその理論的背景について，2 章では母集団調整法の性能を評価したシミュレーション研究について，3 章では母集団調整法の適用事例について解説する。また，Appendix には R による実装例を示す。本報告書は，TSD18(5)の内容を基礎にしているものの，TSD18 では述べられていない母集団調整法のシミュレーション研究による性能評価や，TSD18 が発行された 2016 年以降の適用事例などについても纏めている。

なお，本報告書では一般的な間接比較や NMA に対する基本的な知識，及び因果推論に対する基本的な知識を前提とする。それらについては，日本製薬工業協会の過去の報告書やシンポジウム資料を参照いただきたい(4,6,7)。

1. 母集団調整法の内容とその理論的背景

「はじめに」で述べたとおり，間接比較や NMA により，複数の試験を統合して興味のある薬剤間の治療効果を推定できるものの，試験間で効果修飾因子の分布が異なる場合には，その影響を調整する必要がある。全ての試験で IPD が利用可能な場合，古典的な標準化（standardization）や一般化（generalization），較正（calibration）法，あるいは，それらを基にした傾向スコアに基づく重み付き解析，回帰モデルに基づく方法，さらには 2 重ロバスト法による調整が可能である。しかしながら，企業が実施する間接比較や NMA では，全ての試験で IPD が利用可能な状況というのは稀である。通常は，自社の試験では IPD の利用が可能であるものの，競合他社の試験では AgD しか利用できないという状況が多いと考えられる。このように，一部の試験では IPD を利用できるものの，残りの試験では AgD しか利用できない場合，試験間における効果修飾因子の分布の違いを調整して治療効果を推定する方法として，Matching Adjusted Indirect Comparison（MAIC）(8–10)，Simulated Treatment Comparison（STC）(10,11)，ネットワークメタ回帰に基づく方法(12–15)，エントロピー平衡化（entropy balancing）(16)などの様々な母集団調整法が近年提案されている。そのうち，MAIC と STC はここ数年適用事例が増加傾向にあり，注目されている手法の一つである。そこで，本報告書ではこれらの 2 つの手法を中心に解説する。なお，全ての試験で IPD が利用可能な場合に適用可能な手法の詳細は本報告書では割愛する。日本製薬工業協会の過去の報告書(6)や TSD18 を参照いただきたい。

MAIC や STC の基本的な考え方は，古典的な標準化や一般化，較正法と同様である。これらは，ある集団において観察された治療効果から，別の集団において観察されるはずの治療効果を推定する手法である。一般的な標準化や一般化，較正法では，対象とする全ての試験で IPD が利用可能な状況を想定しているのに対し，MAIC や STC では，利用可能な試験が 2 試験の場合，AB 試験（治療 A と治療 B を比較する試験）では IPD が利用可能であるものの，AC 試験（治療 A と治療 C を比較する試験）では AgD しか利用できない状況を想定している（AB 試験や AC 試験の説明は次節で行う）。また，共変量 X についても，AgD では全ての共変量の完全な同時分布を利用することができず，各共変量の周辺分布（平均や標準偏差など）しか利用できない状況を想定している。このように，AC 試験の IPD が利用できない状況では，通常の標準化や一般化，較正法，あるいは上述したような，それらを基にした手法を利用できない。そこで，一部の試験でしか IPD を利用できない状況でも，その一部の試験の IPD を利用して母集団を調整した治療効果を推定するため，通常の手法を拡張した手法（MAIC 及び STC）が提案されている。MAIC とは傾向スコアに基づく重み付き解析を，STC とは回帰モデルに基づく方法をそれぞれ一部の試験でしか IPD を利用できない状況に拡張した手法である。また，MAIC と STC は，アンカー（共通の比較対照群）を有する間接比較と，アンカーを有さない間接比較の両方で適用可能である。なお，MAIC や STC は，観察された効果修飾因子の分布の試験間の相違を調整するための手法であり，治療に関連する試験間の相違，例えば，治療遵守の割合の差，併用治療の割合の差，治療スイッチングの割合の差などは調整することができない。

1.1 状況設定

IPD が利用可能な AB 試験（自社の試験）と，AgD のみが利用可能な AC 試験（競合他社の試験）がそれぞれ 1 試験ずつある状況を想定する。ターゲット集団 P（一般には，AB 試験の集団や AC 試験の集団とは異なる）において，適切な尺度上での治療 B に対する治療 C の相対治療効果（パラメータ：$d_{BC(p)}$）の推定に関心がある。なお，カッコ付きの下付き文字により，ある特定の集団であることを示す。AB 試験の集団における各治療に対するアウトカムの期待値を表すパラメータをそれぞれ$\mu_{A(AB)}$，$\mu_{B(AB)}$，$\mu_{C(AB)}$とする［AB 試験で評価していない治療（本状況設定の場合は治療 C）に対するパラメータも含む］。AB 試験の集団における$\mu_{A(AB)}$,$\mu_{B(AB)}$に対する推定量をそれぞれ$\bar{Y}_{A(AB)}$，$\bar{Y}_{B(AB)}$とする。$\bar{Y}_{A(AB)}$，$\bar{Y}_{B(AB)}$は各治療に対するアウトカムの平均（連続変数の場合）や割合（2 値変数の場合）などである（AB 試験では$\mu_{C(AB)}$は推定できない）。AC 試験でも AB 試験と同様に，3 つのパラメータ（$\mu_{A(AC)}$，$\mu_{B(AC)}$，$\mu_{C(AC)}$）と 2 つの推定量（$\bar{Y}_{A(AC)}$，$\bar{Y}_{C(AC)}$）を定義する（AC 試験では$\mu_{B(AC)}$は推定できない）。

適切な尺度（ロジット尺度，対数尺度など）を選択した後，対応する適切なリンク関数$g(.)$を用いることで，各試験の集団における周辺相対治療効果$d_{AB(AB)}$，$d_{AC(AC)}$に対する推定量$\hat{\Delta}_{AB(AB)}$，$\hat{\Delta}_{AC(AC)}$は以下の式で表される。

$$\hat{\Delta}_{AB(AB)} = g\left(\bar{Y}_{B(AB)}\right) - g\left(\bar{Y}_{A(AB)}\right)$$

$$\hat{\Delta}_{AC(AC)} = g\left(\bar{Y}_{C(AC)}\right) - g\left(\bar{Y}_{A(AC)}\right)$$

例えば，ロジット尺度に対するリンク関数は$g(y) = \log\left(y/(1-y)\right)$，対数尺度に対するリンク関数は$g(y) = \log(y)$となる。また，アウトカムの尺度のままであれば，恒等リンク関数（$g(y) = y$）を用いる。

【効果修飾因子の分布が試験の集団間で同じ場合】

間接比較を行うための標準的な手法では，選択した尺度上で，試験の集団間（AB 試験の集団，AC 試験の集団及びターゲット集団 P）で効果修飾因子の分布に相違はないことを前提とする。従って，以下のとおり，周辺相対治療効果は集団間で同じと仮定する（相対治療効果の恒常性の仮定，1.2.2 節参照）。

$$d_{AB(AB)} = d_{AB(AC)} = d_{AB(P)}$$

$$d_{AC(AB)} = d_{AC(AC)} = d_{AC(P)}$$

この仮定と，各試験内では対象者はランダム化されていることを考慮した下で，相対治療効果$d_{BC(P)}$に対する推定量は，以下の式で表される。

$$\hat{\Delta}_{BC(P)} = \hat{\Delta}_{AC(AC)} - \hat{\Delta}_{AB(AB)}$$

NICE の医療技術評価では一般に絶対治療効果（各群のベースラインからの平均変化量や反応率など）が必要であるため，これらの相対治療効果から，ターゲット集団 P における絶対治療効果を推定する。ターゲット集団 P における，比較対照群である治療 A の絶対治療効果（$\mu_{A(P)}$）の推定量を$\bar{Y}_{A(P)}$とすると，ターゲット集団 P における治療 B 及び治療 C の絶対治療効果（$\mu_{B(P)}$，$\mu_{C(P)}$）に対する推定量はそれぞれ以下の式で表される。

$$\hat{Y}_{B(P)} = g^{-1}\left(g\left(\bar{Y}_{A(P)}\right) + \hat{\Delta}_{AB(P)}\right)$$

$$\hat{Y}_{C(P)} = g^{-1}\left(g\left(\bar{Y}_{A(P)}\right) + \hat{\Delta}_{AC(P)}\right)$$

なお，ここで重要な点として，2 つの推定量を区別して表示する。一つはある集団における統計量として導出される推定量であり，バーを用いて表示する。上式では，集団 P で認められる$\bar{Y}_{A(P)}$がそれに当たる。もう一つは，ある種の外部情報や仮定を追加することで導出される推定量であり，ハットを用いて表示する。上式では，$d_{AB(P)} = d_{AB(AB)}$，$d_{AC(P)} = d_{AC(AC)}$とした仮定した下で，AB 試験の集団，AC 試験の集団でそれぞれ認められる相対治療効果$d_{AB(AB)}$，$d_{AC(AC)}$から導出した$\hat{Y}_{B(P)}$，$\hat{Y}_{C(P)}$がそれにあたる。

【効果修飾因子の分布が試験の集団間で異なる場合】

効果修飾因子の分布が試験（AB 試験と AC 試験）の集団間で異なるため，母集団を調整することにより，AC 試験と同一の集団で認められるはずの治療効果を推定する場合について考える。なお，予後因子については，たとえ試験間でそれらの分布が異なったとしても，結果に影響を及ぼさない。なぜなら，各試験内で対象者はランダム化されているため，（対象者数が十分に多く，適切にランダム化されているという状況下では）これらの因子は相対治療効果に影響を及ぼさないからである。効果修飾因子 X は，変換した尺度上で，線形な効果を有すると仮定する。例えば，X を条件付けた下での相対治療効果は，切片項（X=0 の場合の相対治療効果$d_{AB}(0)$）と交互作用項γXを用いて，以下のように表される。

$$d_{AB}(X) = d_{AB}(0) + \gamma X$$

ここで，γは効果修飾因子に対するパラメータとする。効果修飾因子については，試験間でそれらの分布が異なる場合，各試験のデータから直接推定した相対治療効果$d_{AB(AB)}$，$d_{AC(AC)}$は，それぞれ当該試験で認められた効果修飾因子の分布と同様の分布を有する集団でのみ妥当な推定値となる。なお，効果修飾因子の調整はバイアスを減らすことだけを目的に行うのではない。効果修飾因子がある場合，間接比較に基づく意思決定の判断にも影響を及ぼす可能性があるからである。例えば，年齢が効果修飾因子と考えられる場合，ある年齢層で有効性を有する治療が別の年齢層で有効性を有さない可能性が生じる。

詳細は 1.3 節及び 1.4 節で説明するが，MAIC や STC により効果修飾因子の調整が可能である。MAIC や STC では，AB 試験における IPD を用いて，$\hat{Y}_{A(AC)}$，$\hat{Y}_{B(AC)}$を予測する。なお，$\hat{Y}_{A(AC)}$，$\hat{Y}_{B(AC)}$は AB 試験の集団が AC 試験の集団と同じであった場合に，AC 試験において治療 A，治療 B を受けた場合にそれぞれ観察されるはずのアウトカムの平均や割合などの予測値である。これらのアウトカムの予測値を求める方法には，IPD を有する AB 試験において，AB 試験の共変量の分布が AC 試験の共変量の分布と一致するように，AB 試験の個々の対象者を重み付けする方法（MAIC）と，AB 試験のアウトカムに対して共変量を用いた回帰モデルを当てはめ，回帰係数を推定した後，それを AC 試験における共変量の分布に用いる方法（STC）がある。

このアウトカムの予測値$\hat{Y}_{A(AC)}$，$\hat{Y}_{B(AC)}$を用いて間接比較を行う。まずアンカーを有する間接比較は，相対治療効果の条件付き恒常性（1.2.2 節参照）を仮定した下で，以下の式で表される。

$$\hat{\Delta}_{BC\,(AC)} = \left(g\left(\bar{Y}_{C(AC)}\right) - g\left(\bar{Y}_{A(AC)}\right)\right) - \left(g\left(\hat{Y}_{B(AC)}\right) - g\left(\hat{Y}_{A(AC)}\right)\right) \quad (1\text{-}1)$$

一方，アンカーを有さない間接比較は，絶対治療効果の条件付き恒常性（1.2.2 節参照）を仮定した下で，以下の式で表される。

$$\hat{\Delta}_{BC\,(AC)} = g\left(\bar{Y}_{C(AC)}\right) - g\left(\hat{Y}_{B(AC)}\right) \quad (1\text{-}2)$$

標準的な間接比較や NMA では通常，尺度の変換が行われるものの，MAIC や STC の文献では，尺度の変換を行わずに（恒等リンク関数を用いて），アウトカムの自然尺度上での間接比較をよく例示している。アウトカムの自然尺度上での治療 A，B，C に対するアウトカムの平均や割合などの推定量は，アンカーを有する間接比較ではそれぞれ，$\bar{Y}_{A(AC)}$，$\hat{Y}_{B(AC)} - \hat{Y}_{A(AC)} + \bar{Y}_{A(AC)}$，$\bar{Y}_{C(AC)}$で表される。一方，アンカーを有さない間接比較ではそれぞれ，$\bar{Y}_{A(AC)}$，$\hat{Y}_{B(AC)}$，$\bar{Y}_{C(AC)}$で表される。また，通常よく行われる変換した尺度上でのアンカーを有する間接比較の場合，治療 A，B，C に対するアウトカムの平均や割合などの推定量はそれぞれ以下の式で表される。

$$\bar{Y}_{A(AC)}$$

$$g^{-1}\left(g\left(\hat{Y}_{B(AC)}\right) - g\left(\hat{Y}_{A(AC)}\right) + g\left(\bar{Y}_{A(AC)}\right)\right)$$

$$\bar{Y}_{C(AC)}$$

上述したとおり，MAIC や STC の潜在的な利点は，ある 1 試験（一般に自社の AB 試験）の IPD があれば適用可能である点であり，競合他社の試験（AC 試験）については，IPD は必ずしも必要なく，アウトカムの AgD（$\bar{Y}_A$，$\bar{Y}_C$）と共変量の分布さえ利用できれば問題はない（1.3 節，1.4 節参照）。しかしながら，AB 試験の IPD のみが利用可能な場合，MAIC や STC に基づく治療 A，B，C に関する適切な比較は，AC 試験と同一の集団でのみ可能となる。従って，何らかの意思決定を下すにあたり，自社の AB 試験の集団ではなく，競合他社の AC 試験の集団をターゲット集団と見做して良いか（すなわち，AC 試験の集団をターゲット集団 P と見做して良いか），あるいは，UK コホート（NICE への提出の場合）や本邦のレジストリで代表される集団と見做して良いか，という本質的な問題が生じる。

1.2 解析手法を用いる際に前提となる仮定（全ての試験で IPD が利用可能な場合）

1.2.1 標準化や一般化を用いる際に前提となる仮定

本節では，ある集団で認められた治療効果から別の集団（一般にはターゲット集団）で認められるはずの治療効果を推定するにあたり，標準化や一般化のように，全ての試験で IPD が利用可能な場合に用いる手法において前提となる仮定について説明する。これらの仮定は，MAIC や STC のように，一部の試験でしか IPD が利用できない場合に用いる手法においても必要となる（1.6.1 節参照）。

これらの手法を適用するにあたり，まず関心があるのはターゲット集団における平均治療効果

（Population Average Treatment Effect：PATE）の推定にある。PATE は，ある集団における平均治療効果，すなわち，標本における平均治療効果（Sample Average Treatment Effect：SATE）をターゲット集団に一般化することにより推定できる。また，状況によっては（例えば，政策を決定する場合），ある集団における治療群の平均治療効果（Sample Average Treatment effect on the Treated：SATT）から，ターゲット集団における治療群の平均治療効果（Population Average Treatment effect on the Treated：PATT）の推定に関心がある(17)。PATT や SATT に基づく治療効果は，実際に治療群に割り当てられた対象者のみに基づく。なお，ランダム化比較試験（対象者数が十分に多く，適切にランダム化されているという状況下）では，SATE と SATT は期待的に等しくなる。

ある集団（標本）において認められた平均治療効果の一般化，すなわち，PATE の推定に必要な仮定として，主に以下の 4 つが挙げられる(17,18)。

【仮定 1】各治療に対するアウトカムの同質性（homogeneity of outcomes on each treatment）：どの対象者がその試験に割り当てられたか否かにかかわらず，各治療に対するアウトカムは変わらない。

【仮定 2】SUTVA（Stable Unit Treatment Value Assumption）：ある対象者のアウトカムは他の対象者の影響を受けない。さらに，どの治療においても，受ける治療の詳細（用量や種類など）は対象者によって異ならない。

【仮定 3】強く無視できる治療の割り当て（strongly ignorable treatment assignment）：観測された共変量で条件付けた下では，治療の割り当てはランダムであり，ターゲット集団からの標本抽出とは独立である。すなわち，ある試験において，測定の有無にかかわらず予後因子や効果修飾因子の群間の不均衡は生じない。

【仮定 4】強く無視できる標本の割り当て（strongly ignorable sample assignment）：観測された共変量で条件付けた下では，標本抽出とアウトカムの両方と関連する未測定の因子はなく，ターゲット集団の各対象者が標本として抽出される確率は全てゼロより大きく 1 未満となる。

【仮定 1】について，例えば，試験実施計画書で規定した選択・除外基準が試験間で異なることにより，本仮定からの逸脱が生じる可能性がある。【仮定 2】について，適切な試験デザインを用いることで，本仮定が成立していると見做すことができる。本仮定は，因果推論に基づく手法を適用する際に通常置かれる。【仮定 3】について，適切なランダム化を実施することで，本仮定が成立していると見做すことができる。【仮定 4】について，未測定の効果修飾因子や予後因子の分布が試験間で異なる場合に本仮定からの逸脱が生じる可能性がある。このように，【仮定 1】，【仮定 2】，【仮定 3】については，適切に計画されたランダム化比較試験では，通常成立していると見做すことができる仮定であり，ペアワイズのメタアナリシスを始め，間接比較や NMA などの統合解析において通常置かれる仮定である。

これらの仮定の成否の判断には placebo test(18)が有用な場合がある。一般的な placebo test では，共通の比較対照群（プラセボ群とは限らない）について，ターゲット集団で認められたアウトカムの観測値が，一般化に基づくアウトカムの予測値と一致するかどうかを確認する。しかしながら，placebo test では，【仮定 1】，【仮定 2】，【仮定 4】のいずれかの仮定からの逸脱が生じていることは確認できるものの，どの仮定からの逸脱が生じているかについては確認できない。また，複数の逸脱が互いにその影響を打

ち消し合っているような状況では，仮定からの逸脱を検出できない(17)。さらに，プラセボ群における，アウトカムの観測値と予測値を比較する placebo test では，【仮定 4】において，分布に不均衡が生じている未測定の予後因子は確認できるものの，分布に不均衡が生じている未測定の効果修飾因子は確認できない(17)。なお，追加で実施可能な場合，共通の実薬対照群における，アウトカムの観測値と予測値を比較する placebo test により，分布に不均衡が生じている未測定の効果修飾因子も検出できる。

相対治療効果の一般化に傾向スコアに基づく手法を用いる場合，Stuart ら(18)は，試験の対象集団（標本）とターゲット集団間の傾向スコアの平均値の差を調べる手法を提案している。傾向スコアの平均値の差が標準偏差の 0.25 倍より大きい場合，相対治療効果の一般化の大部分は外挿に基づいており，傾向スコアのモデルに過度に依存していることが示唆される。

1.2.2　較正法を用いる際に前提となる仮定

1.2.1 節では，標準化や一般化を実施する際に想定する状況，すなわち，ターゲット集団（例えば，AC 試験の集団）の部分集団である標本（例えば，AB 試験の集団）で認められた治療効果からターゲット集団で認められるはずの治療効果を推定する際に前提となる仮定について述べた。本節では，較正法を用いて，例えば，AC 試験の集団における治療 B に対する治療 C の平均治療効果を推定する際に前提となる仮定について説明する。なお，較正法では一般に AB 試験の集団と AC 試験の集団が互いに独立の状況を想定する。以下に記載のとおり，前提となる仮定に応じていくつかのアプローチがある。

まず始めは，「絶対治療効果の恒常性（constancy of absolute effects）の仮定」に基づくアプローチである。本アプローチでは，2 つの集団間で予後因子や効果修飾因子の分布に不均衡はないことを仮定する。しかしながら，試験間ではランダム化が行われている訳ではないため，この仮定は合理的でないと言える。エビデンスを統合する際には，母集団の調整を実施するか否かにかかわらず，あまりにも現実的ではない強い仮定に基づく手法は一般に受け入れられない。

第 2 は，「絶対治療効果の条件付き恒常性（conditional constancy of absolute effect）の仮定」に基づくアプローチである。較正法の文献では，この仮定は，治療特異的な条件付き恒常性（treatment-specific conditional constancy）(19)の仮定とも呼ばれる。本アプローチでは，共変量で条件付けた下では，治療 B により期待される絶対治療結果は 2 つの集団間で同じであると仮定する。全ての効果修飾因子と予後因子が利用可能であることを前提とするため，非常に強い仮定と言える(19)。この仮定に基づく間接比較では，傾向スコアに基づく重み付き解析や回帰モデルに基づく方法を用いて，AC 試験の集団における各群のアウトカムの予測値を求める。絶対治療効果の条件付き恒常性の仮定は，一般化の文献でよく用いられる無視できる標本の割り当ての仮定（上述の【仮定 4】）と同じである。

第 3 は，「相対治療効果の恒常性（constancy of relative effects）の仮定」（単に，恒常性の仮定とも呼ばれる）に基づくアプローチである。本アプローチでは，AB 試験の集団で認められた治療 A に対する治療 B の相対治療効果は，AC 試験の集団で認められるはずの治療 A に対する治療 B の相対治療効果と同一であると仮定する。しかしながら，全ての効果修飾因子（未測定の因子も含む）の分布は 2 つの集団間で同じであることを前提とするため，疑問の余地がある仮定である。この仮定は，NMA でよく用いられる（変換した尺度上での）一致性の仮定に類似している。なお，固定効果モデルに基づく解析では，

一致性の仮定の成立を前提としており，変量効果モデルに基づく解析では，一致性の仮定は期待的に成立することを前提とする。興味がある対比に対する試験数が十分に多い（スパースでないネットワークの）場合，効果修飾因子の影響は互いに相殺されるため，変量効果モデルに基づく解析において，一致性の仮定を置くことは妥当と言える。しかしながら，NICE への提出では，一般にスパースなネットワークに基づく比較を行わざる負えない場合が多い。

上述の 3 つのアプローチはいずれも比較的強い仮定に基づいている。第 4 は，「相対治療効果の条件付き恒常性（conditional constancy of relative effects）の仮定」（単に，条件付き恒常性の仮定とも呼ばれる）に基づくアプローチである。本アプローチでは，共変量を条件付けた下では，AB 試験の集団で認められた治療 A に対する治療 B の相対治療効果は, AC 試験の集団で認められるはずの治療 A に対する治療 B の相対治療効果と同一であると仮定する。本仮定の下で母集団の調整に必要となるのは効果修飾因子のみであるため, 上述の 3 つのアプローチで必要とする仮定に比べて, より現実的な仮定と言える。ただし，ランダム化により，試験内での予後因子の均衡は保持されていることを前提とする。

1.3 MAIC

MAIC はノンパラメトリックな尤度再重み付け法の一つである(8–10)。1 章で述べたとおり，本手法は傾向スコアに基づく重み付き解析を拡張した手法であり，本手法により，AC 試験の IPD を利用することなく, ロジスティック回帰モデルに基づき傾向スコアの推定が可能となる。ターゲット集団である AC 試験で治療 t（=A，B）を受けた場合に認められるはずの平均アウトカム$\mu_{t(AC)}$は，AB 試験で治療 t を受けた$N_{t(AB)}$名の対象者に対するアウトカムの重み付き平均として表される。式で表すと以下のとおりである。

$$\hat{Y}_{t(AC)} = \frac{\sum_{i=1}^{N_{t(AB)}} Y_{it(AB)} w_{it}}{\sum_{i=1}^{N_{t(AB)}} w_{it}}$$

ここで，$Y_{it(AB)}$は AB 試験で治療 t を受けた対象者 i に対するアウトカム，対象者 i に対する重みw_{it}は，各対象者が AB 試験に対して AC 試験に登録される（割り当てられる）オッズとして表される。このように，本手法は，全ての試験で IPD が利用できる場合に傾向スコア（なお，ここでの傾向スコアとは，共変量を条件付けた下で，ターゲット集団の対象者が，ある標本集団に割り当てられる条件付き確率のことを指す）の逆数で重み付けする重み付け解析［例えば，Inverse Probability Weighting（IPW）解析］と概念的には同じである。

重み自体は以下のロジスティック回帰モデルを用いて推定する。

$$\log(w_{it}) = \alpha_0 + \alpha_1^T X_{it}$$

ここで，X_{it}は治療 t を受けた対象者 i の共変量ベクトルとする。しかしながら，AC 試験における IPD を利用できないため，標準的な手法を用いて回帰パラメータを推定できない。

AC 試験における IPD が利用できないものの，仮に AC 試験における共変量の同時分布が利用可能な場合，IPD に対する十分統計量を用いることで，Nie ら(20)の尤度再重み付け法の適用が可能である。一方，AC 試験における共変量の周辺分布（例えば，平均や標準偏差）のみが利用可能な場合，Signorovitch ら(9)のモーメント法の適用が可能である。この手法では，AB 試験の対象者を重み付けした集団と AC 試験の対象者集団間で各共変量の平均（場合によっては，さらに高次のモーメントまで）が等しいとする推定方程式を解くことで$\hat{\alpha}_1$を推定する。（例えば，中心化により）$\bar{X}_{(AC)} = 0$とした場合，このアプローチは，$\sum_{t=A,B}\sum_{i=1}^{N_{t(AB)}} \exp(\alpha_1^T X_{it})$を最小化することに相当する。

$\hat{Y}_{t(AC)}$は以下のように表される。

$$\hat{Y}_{t(AC)} = \frac{\sum_{i=1}^{N_{t(AB)}} Y_{it(AB)} \exp(\hat{\alpha}_1^T X_{it})}{\sum_{i=1}^{N_{t(AB)}} \exp(\hat{\alpha}_1^T X_{it})}$$

次に，1.1 節で述べたとおり，(1-1)及び(1-2)を用いて，アンカーを有する間接比較及びアンカーを有さない間接比較をそれぞれ行う。

上述のとおり，MAIC の文献では，尺度の変換を行わずに（恒等リンク関数を用いて），アウトカムの自然尺度上での間接比較をよく例示しているものの，MAIC によりあらゆる尺度上で間接比較を行うことができる。また，通常は分散に頑健なサンドイッチ推定量を用いる。サンドイッチ推定量は，重み自体が推定されることを考慮して，データから経験的に推定する。

次に，AB 試験の対象者を重み付けした疑似集団の有効標本サイズ（Effective Sample Size：ESS）は以下の式で近似できる(9)。

$$\text{ESS} = \left(\sum_{t=A,B}\sum_{i=1}^{N_{t(AB)}} \hat{w}_{it}\right)^2 \Big/ \sum_{t=A,B}\sum_{i=1}^{N_{t(AB)}} \hat{w}_{it}^2$$

重みが固定且つ既知の場合やアウトカムと相関していない場合には，上式に基づく ESS の推定は正確であるが，そうでない場合には ESS を過少評価する可能性がある(21)。ESS が小さい場合，対象集団が十分にオーバーラップしておらず，重みのばらつきが大きいことが示唆しており，推定が不安定な可能性がある。

対象集団のオーバーラップの程度や，結果に過度に影響を及ぼす可能性のある対象者の存在を確認するには，重みの分布を直接調べることが有用である。しかしながら，傾向スコアの推定は AB 試験でのみ可能であるため，オーバーラップの程度の確認に古典的な傾向スコアに基づく手法を適用できない。モーメント法では，定義どおりに［少なくとも，AC 試験で用いた情報のレベル（例えば，平均や分散）まで］共変量の分布の均衡を保証するのみである。

1.4 STC

STC は共変量に基づく調整解析の一つである(10,11)。1 章で述べたとおり，STC は回帰モデルに基づく方法を拡張した手法であるため，まずは AB 試験における IPD に対して，以下のアウトカムモデルを当てはめる。

$$g\left(\mu_{t(AB)}(X)\right) = \beta_0 + \beta_1^T X + (\beta_B + \beta_2^T X^{EM})I(t = B)$$

ここで，β_0は切片項，β_1は予後因子の係数ベクトル，β_Bは$X = 0$の場合の治療 A に対する治療 B の相対治療効果，Xは完全な共変量ベクトル，X^{EM}は効果修飾因子ベクトル（Xのサブベクトル），β_2は効果修飾因子X^{EM}の係数ベクトル，$\mu_{t(AB)}(X)$はリンク関数$g(.)$により，選択した線形予測子の尺度に変換した下で，共変量 X（効果修飾因子X^{EM}を含む）と割り付け治療 t により推定されるアウトカムの期待値とする。なお，このモデルは，Ishak ら(10)が提案した効果修飾因子を含まないモデルのより一般的な形式のモデルに相当する。

このアウトカムモデルにおいて，各パラメータを推定した後，個々の共変量の値をそれらの平均で置き換えることで，$\hat{Y}_{A(AC)}$及び$\hat{Y}_{B(AC)}$は以下のとおりに予測できる。

$$\hat{Y}_{A(AC)} = g^{-1}\left(\hat{\beta}_0 + \hat{\beta}_1^T \bar{X}_{(AC)}\right)$$

$$\hat{Y}_{B(AC)} = g^{-1}\left(\hat{\beta}_0 + \hat{\beta}_1^T \bar{X}_{(AC)} + \hat{\beta}_B + \hat{\beta}_2^T \bar{X}_{(AC)}^{EM}\right)$$

ここで，$\bar{X}_{(AC)}$は AC 試験の集団における共変量の平均ベクトル，$\bar{X}_{(AC)}^{EM}$は AC 試験の集団における効果修飾因子の平均ベクトルとする。$\hat{Y}_{A(AC)}$及び$\hat{Y}_{B(AC)}$を用いて，例えば，自然尺度上でのアンカーを有する間接比較，アンカーを有さない間接比較はそれぞれ以下のとおりに実施可能である(10)。

$$\hat{\Delta}_{BC(AC)} = \bar{Y}_{C(AC)} - \bar{Y}_{A(AC)} - \left(\hat{Y}_{B(AC)} - \hat{Y}_{A(AC)}\right)$$

$$\hat{\Delta}_{BC\,(AC)} = \bar{Y}_{C(AC)} - \hat{Y}_{B(AC)}$$

Ishak ら(10)は，$g(.)$が恒等関数でない（$g(y) = y$でない）場合，$\hat{Y}_{A(AC)}$及び$\hat{Y}_{B(AC)}$（及び，これらの推定量に基づくアウトカムの自然尺度上での間接比較）にはバイアスを伴うことを指摘している。なぜなら，アウトカムの平均は，共変量の平均のみでなく，分布の全体に依存するからである。そこで，Ishak らは，共変量の平均で置き換えるのではなく，AC 試験における対象集団の共変量の同時分布から標本を抽出し（なお，共変量間の相関については，例えば，AB 試験の結果で補完するか，相関はないと仮定する方法がある），次に，その標本を用いて回帰モデルから予測したアウトカムの平均を取ることで，求めたい推定値を得る方法（シミュレーションに基づくアプローチ）を推奨している。しかしながら，このシミュレーションに基づくアプローチでは，AC 試験における対象集団の共変量の分布の平均から効果を推定するのではなく，AC 試験の対象集団から無作為に抽出した対象者集団に対して期待される効果を推定するため，ばらつきは大きくなり，結果として間接比較における最終的な推定値の精度を過小評価する可能性がある。

一方で，アウトカムの自然尺度上で間接比較を行うと様々な問題が生じる（1.6.2 節参照）。これらの問題を回避するため，アンカーを有する間接比較及びアンカーを有さない間接比較は，それぞれ(1-1)及び(1-2)を用いて線形予測子の尺度上で実施することが強く推奨される。モデルチェックを実施するため

の標準的なツール（AIC や DIC，残差の評価など）は，AB 試験におけるアウトカムモデルを構築するときに利用できる。しかしながら，MAIC と同様に，AC 集団におけるアウトカムを予測するには追加の仮定が必要であり，利用可能なデータが殆どない場合には，一般にモデルを評価することが困難である。

1.5 その他の手法

1.5.1 ネットワークメタ回帰を拡張した手法

AB 試験と AC 試験の両方で IPD を利用できる場合，効果修飾因子の分布の違いを調整する手法として，上述した手法以外で最も標準的な手法の一つはネットワークメタ回帰である（1 章参照）。AB 試験でしか IPD を利用できない場合，ネットワークメタ回帰を拡張した手法が提案されており，IPD と AgD をどのようにモデル化するかによって，大きく 2 つの手法に大別される。

1 つ目は Sutton ら(22)によるペアワイズのメタ回帰に基づくアプローチである。AB 試験の IPD のアウトカム及び AC 試験の AgD のアウトカムに対してそれぞれ以下の回帰モデルを同時に当てはめる。

IPD のアウトカムモデル

$$g\left(\mu_{t(AB)}(X)\right) = \beta_{0(AB)} + \beta_1^T X + (\beta_B + \beta_2^T X^{EM}) I(t = B) \quad (1\text{-}3\text{-}1)$$

AgD のアウトカムモデル

$$g\left(\mu_{t(AC)}\right) = \beta_{0(AC)} + \left(\beta_C + \beta_2^T \bar{X}_{t(AC)}^{EM}\right) I(t = C) \quad (1\text{-}3\text{-}2)$$

ただし，データが限られていることから，モデルに対して幾つかの制約を設ける必要がある。1 つ目の制約は，固定効果モデルを用いる必要がある点である。2 つ目の制約は，治療と効果修飾因子の交互作用に対する係数β_2は治療 B と治療 C の間で，さらに個人レベルと集団レベルの間で共有する必要がある点であり，この制約により，効果修飾因子は個人レベルと集団レベルの両方で同じような挙動を示すことが前提となる（係数β_2を共有するという点で，共有の効果修飾因子の仮定と呼ばれる）。この仮定は，恒等リンク関数が用いられ，未測定の効果修飾因子がない（そして適切なランダム化が行われている）場合にのみ妥当である。この仮定が成立しない場合にはエコロジカルバイアスが生じる(13,14)。最後に，$\beta_1^T X$の項は IPD のアウトカムモデルにのみ含めることとする。

2 つ目は Jackson ら(23,24)による階層的関連回帰（hierarchical related regression）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。その基本的な考え方は，AgD は集団に対して平均を取ることで得られることから，AgD のアウトカムモデルは，各対象者に対するモデルの集団に対する平均（すなわち積分）により得られるというものである。一般的な形式で表すと以下のとおりである。

IPD のアウトカムモデル

$$g\left(\mu_{t(AB)}(X)\right) = \beta_{0(AB)} + \beta_1^T X + (\beta_B + \beta_2^T X^{EM}) I(t = B) \quad (1\text{-}4\text{-}1)$$

AgD のアウトカムモデル

$$\mu_{t(AC)} = \int g^{-1}\left(\beta_{0(AC)} + \beta_1^T X + (\beta_C + \beta_2^T X^{EM}) I(t = C)\right) f_{(AC)}(X) dX \quad (1\text{-}4\text{-}2)$$

ここで，積分は AC 試験の集団における X が取りうる区間に対して実施し，$f_{(AC)}(X)$は AC 試験の集団における X の同時分布とする。なお, AC 試験の集団における完全な同時分布を利用できない場合でも，近似値を利用できる場合がある。例えば，連続変数の共変量の分布について，（論文などで）報告された平均と標準偏差から正規分布（状況に応じて，対数正規分布などの適切な分布）を仮定し，さらに共変量間の相関については，AB 試験の結果で補完するか，相関はないと仮定することが可能である。また，このモデルは，全ての試験で IPD が利用可能な場合，標準的なネットワークメタ回帰モデルに帰着し，さらに，一部の試験でしか IPD が利用できない場合，AB 試験と AC 試験の 2 試験のみ利用可能な状況だけではなく，より大きなネットワークでの利用が可能である。また，AB 試験と AC 試験の 2 試験のみ利用可能な場合，データが限られているため，(1-4-1)及び(1-4-2)において，上述した（係数β_2に関する）共有の効果修飾因子の仮定が必要となるものの，より大きなネットワークが利用可能な場合や，治療 C の効果修飾因子に関する外部情報が利用可能な場合, 共有の効果修飾因子の仮定を置く必要がない場合がある。恒等リンク関数が用いられ，且つ全ての効果修飾因子が利用可能な場合，(1-4-1)及び(1-4-2)は(1-3-1)及び(1-3-2)と同等になる。

上述のとおり，(1-4-1)及び(1-4-2)を用いた階層的アプローチは，MAIC や STC とは概念的に全く異なるアプローチである。階層的アプローチでは，個人のレベルの関連性をモデル化するため，標準的な間接比較のように，個人レベルと集団レベルの両方で内的に一貫した推測を行うことができる。一方，MAIC や STC では，各治療群の平均アウトカムの予測に IPD を用いるため，集団レベルの推測に影響を及ぼす。この点において，MAIC と STC は母集団を調整した試験レベルの間接比較に基づくアプローチ（population-adjusted study-level indirect comparison），階層的アプローチは母集団を調整した個人レベルの間接比較に基づくアプローチ（population-adjusted individual-level indirect comparison）と呼ぶことができる。TSD18 は MAIC や STC に焦点を置いているため, 階層的アプローチの詳細に関する記述はないが，その後公表された Phillippo らの論文(25)では，階層的アプローチの性能についても述べている。

1.5.2　その他の再重み付け解析

試験間で共変量の分布の均衡を保つために，IPD を有する集団の各対象者を重み付けする手法は，集団再重み付け法と呼ばれる。MAIC が最もよく用いられる集団再重み付け法の一つであるが，その他の手法として，エントロピー平衡化が挙げられる(16,26,27)。エントロピー平衡化では，傾向スコアを用いて重みを推定する代わりに，共変量の分布のモーメント（例えば，平均や標準偏差）が一致するように直接重みを推定する。上述した MAIC では，モーメント法を用いて重みを推定するため，事実上エントロピー平衡化と同じである。しかしながら，エントロピー平衡化では，重みの推定に追加の制約を必要とする。それは，エントロピー平衡化の最適な重みは，均一な重みに可能な限り近い重みである，という制約である。この追加の制約により，エントロピー平衡化では，MAIC と比較してばらつきが小さく

なることが期待される。

その他の集団再重み付き法として，群ごとに共変量の分布の均衡を保つ手法がある。すなわち，AB 試験と AC 試験の 2 試験がある場合，2 つの試験の比較対照群（治療 A）間及び興味のある治療群間（治療 B と治療 C）のそれぞれで，共変量の分布の均衡を保つ手法である。Belger ら(26,27)は，アンカーを有する場合と有さない場合のそれぞれで，この手法と他の手法との比較を行っている。

1.6 MAIC や STC の性質

1.6.1 MAIC や STC を用いる際に前提となる仮定

上述したとおり，MAIC や STC を用いる場合にも，標準化や一般化，較正法で通常置かれる仮定が必要となる。【仮定 1】，【仮定 2】，【仮定 3】については，適切に計画されたランダム化比較試験では，通常成立していると見做すことができる仮定であり，MAIC や STC に基づく統合解析においても通常前提となる仮定である。【仮定 4】については，適切な尺度上での絶対治療効果の条件付き恒常性の仮定に相当する。

標準的な間接比較や NMA では，線形予測子の尺度上で，相対治療効果の恒常性（第 3 のアプローチ，1.2.2 節参照）を仮定する。アンカーを有する間接比較を MAIC や STC を用いて実施する場合，アウトカムの自然尺度上で，相対治療効果の条件付き恒常性（第 4 のアプローチ，1.2.2 節参照）を仮定することが多い。すなわち，効果修飾因子で条件付けた下では，相対治療効果は試験間で一定であると仮定する。また，【仮定 1】，【仮定 2】，【仮定 3】により，試験内における予後因子の均衡は保持されるため，試験間における予後因子の分布の相違について仮定を置く必要はない。

一方，アンカーを有さない間接比較を MAIC や STC を用いて実施する場合，絶対治療効果の条件付き恒常性（第 2 のアプローチ，1.2.2 節）を仮定する必要がある。本仮定は，相対治療効果の条件付き恒常性よりも強い仮定であり，効果修飾因子と予後因子で条件付けた下では，絶対治療効果は一定であると仮定するため，全ての効果修飾因子と予後因子が既知であることが前提となる。これは非常に強い仮定であるため，成立していると見做すことが一般に困難である。従って，非連結のネットワークや単群試験を含むネットワークなど，アンカーを有さない間接比較の実施には注意を払う必要がある。

各手法を用いる際に前提となる仮定については表 1-1 を参照のこと。

また，MAIC や STC は基本的に 2 つの試験（一つの試験が IPD を有する試験，もう一つが AgD を有する試験，1.1 節参照）に対して適用可能な手法であり，より大きなネットワークに対して本手法を適切に適用するのは容易ではない。より大きなネットワークに対して MAIC を適用するには，例えば，IPD を有する試験の効果修飾因子の分布を，ネットワークの中から選択した，ある試験の効果修飾因子の分布に一致させてから，通常の NMA を実施する方法や，IPD を有する試験の効果修飾因子の分布を，ネットワークに含まれる全ての試験の平均的な集団の効果修飾因子の分布に一致させてから，通常の NMA を実施する方法が挙げられる(26)。しかしながら，これらの手法では，MAIC を適用する際には，IPD を有する試験と AgD を有する試験（あるいは全ての試験の平均的な集団）間で効果修飾因子の分布は異

なることを前提とする一方で，その後，NMA を適用する際には，AgD を有する試験間で効果修飾因子の分布は同じであることを前提とするため，正当化が困難な可能性がある。また，AgD を有する試験の集団（あるいは全ての試験の平均的な集団）はターゲット集団と必ずしも一致しない。

表 1-1 間接比較を実施するための各解析手法で前提となる仮定の有無

仮定	解析手法					
	標準的な間接比較や NMA	ネットワークメタ回帰 [2)]	MAIC		STC	
			アンカーを有さない場合	アンカーを有する場合	アンカーを有さない場合	アンカーを有する場合
各治療に対するアウトカムの同質性	Yes	Yes	Yes	Yes	Yes	Yes
SUTVA	Yes	Yes	Yes	Yes	Yes	Yes
強く無視できる治療の割り当て	Yes	Yes	No	Yes	No	Yes
絶対治療効果の恒常性	No	No	No	No	No	No
絶対治療効果の条件付き恒常性	No	No	Yes[4)]	No	Yes[4)]	No
相対治療効果の恒常性	Yes[1)]	No	No	No	No	No
相対治療効果の条件付き恒常性	No	Yes[1)]	No	Yes[4)]	No	Yes[4)]
共有の効果修飾因子	Not Applicable	Yes[3)]	No[5)]	No[5)]	No[5)]	No[5)]

各仮定の詳細は 1.2.1 節及び 1.2.2 節を参照。

1) 適切な線形予測子の尺度上。2) 全ての試験で IPD を利用可能な場合，全ての試験で試験レベルのデータしか利用できない場合，一部の試験でしか IPD を利用できない場合を含む。3) 適切な線形予測子の尺度上。なお，全ての試験で IPD が利用可能な場合，本仮定は必要ない。4) 一般に自然尺度上。5) 基本的に本仮定は必要ないものの，あらゆるターゲット集団における治療効果に関心がある場合，本仮定が必要な場合がある。

1.6.2　尺度の重要性と効果修飾因子との関連

標準的な間接比較や NMA は，事前に規定した尺度，すなわち，線形予測子の尺度上で実施する。例えば，割合に対してはロジット尺度，率に対しては対数尺度が一般に用いられる。効果修飾因子とは通常，比較を実施する尺度（標準的な間接比較や NMA では，一般に線形予測子の尺度）上で，治療効果に影響を及ぼす因子のことを指す。現在，MAIC や STC は一般にアウトカムの自然尺度上で適用されており，その場合，標準的な間接比較では効果修飾因子と見做された因子が，MAIC や STC の枠組みでは効果修飾因子と見做されない可能性があり，逆に，MAIC や STC では効果修飾因子と見做された因子が，標準的な間接比較の枠組みでは効果修飾因子と見做されない可能性がある。

それぞれのアウトカムに対する適切な尺度は何か？という問いに対しては議論の余地があるものの，アウトカムが割合や率の場合には，リスク差のような絶対尺度よりも，ロジット尺度や対数尺度を用いた方が，試験を通して相対治療効果が安定する（効果修飾因子の数が少なくなるか，効果修飾因子の影響が小さくなる）ことが知られている。また，間接比較における尺度選択では，異なる尺度（例えば，割合と率）を用いることにより，真逆の結論が導かれる可能性がある点が懸念点として挙げられる(28)。真逆の結論が導かれる理由としては，加法性の仮定は必ずしも全ての尺度上で成立している訳ではないにもかかわらず，何れの尺度を用いる場合も，それぞれの線形予測子の尺度上での加法性の仮定に基づいているためである(29)。従って，適切な尺度の選択は重要であり，生物学的知識及び臨床的知識を用いて選択することが推奨される(30)。さらに，あるアウトカムに対して，標準的な尺度があり，加法性の仮定の成立が一般的に受け入れられている場合，別の尺度を用いることを正当化するのは困難である。

意思決定を行う状況において，効果修飾因子であるかどうかの検討は慎重に行う必要がある。NICE の方法ガイド（NICE 2013）(31)には，効果修飾因子は事前に特定され，臨床的に妥当な因子でなければならないことや，対象とする疾患領域の徹底的なレビューまたは専門家の臨床的意見に基づき，裏付けとなる根拠を示す必要があること（NICE の方法ガイド 5.2.7 節参照）が明記されている。さらに，現状では，効果修飾因子の調整は，母集団調整法によりバイアスを制御した相対治療効果を推定するために行われているが，効果修飾因子の存在は意思決定の性質を変える可能性がある。例えば，年齢が効果修飾因子と考えられる場合，ある年齢層で有効性を有する治療が別の年齢層で有効性を有さない可能性が生じる。

以上より，母集団を調整した治療効果の推定に関連した 3 つの推奨事項が挙げられる。第 1 に，MAIC と STC の何れを用いる場合でも，通常の統合解析と同様に，母集団の調整は線形予測子の尺度上で実施すべきである。第 2 に，MAIC や STC は，基本的に相対治療効果を較正するために適用すべきである。第 3 に，母集団の調整にある因子を用いる場合，その因子を用いることにした理由を正当化する必要がある。具体的には，ある因子を効果修飾因子と見做す場合，外部情報に基づく定量的なエビデンス，専門家の意見あるいは系統的レビューによって，その適切性を裏付ける必要がある。また，その因子の分布の試験間の不均衡の程度を明確にする必要がある。本手法を適用する際には，これらの必要性を考慮することで制御可能なバイアスの程度を定量的に示す必要がある。それにより，母集団の調整により得られる治療効果と，標準的な間接比較により得られる未調整の治療効果との比較が可能となる。

TSD18 で推奨しているアプローチは, 仮に効果修飾因子が存在しなかった場合, 母集団の調整に MAIC や STC を適用したとしても，母集団の調整は行われないという特性を有する。つまり，母集団の調整により得られる治療効果は，標準的な間接比較や NMA により得られる治療効果と同じになることが期待される。この特性は以下の理由から非常に望ましい。第 1 に，現在実施されているアンカーを有する間接比較における MAIC と STC では，標準的な間接比較で通常用いられるモデルを基にしている。第 2 に，NICE への提出では，各企業がそれぞれ独自の解析を実施することが可能であるため，仮に効果修飾因子が存在しなかった場合でも，各手法を適用することで誤って何らかの調整が行われてしまうと，同じ条件下の異なる薬剤に対して，矛盾した意思決定を下す可能性が生じるからである。

1.6.3 適切なターゲット集団における治療効果の推定

MAIC や STC に基づく治療効果は対象集団に依存する。すなわち，MAIC や STC では，ターゲット集団（AgD のみを有する AC 試験の集団）に対してのみ妥当な比較を行い，バイアスのない治療効果を推定する。従って，MAIC や STC では，AC 試験の集団以外の集団における治療効果を推定することはできない。しかしながら，意思決定のためのターゲット集団には一般に UK コホート（NICE への提出の場合）や，臨床的な意思決定に関連したレジストリの集団を用いることが推奨され，その集団は一般に AC 試験の集団とは一致しない可能性が高いと考えられる。

この問題に対処するため，TSD18 では，共有の効果修飾因子の仮定と呼ばれる追加の仮定を置くことで，あらゆるターゲット集団における相対治療効果として解釈可能となる点について述べている（詳細は TSD18 2.5 節及び Appendix B 参照）。ここで述べる共有の効果修飾因子の仮定とは，興味のある実薬治療に対して置かれる仮定であり，興味のある各実薬治療の効果は，同じ効果修飾因子によってのみ，同じように影響を受けることを仮定する。例えば，治療 A（プラセボまたは標準治療）をアンカーとした AB 試験と AC 試験の 2 つの試験が利用可能な状況では，興味のある実薬治療である治療 B と C に対して共有の効果修飾因子の仮定が置かれる。この仮定により，治療 B と治療 C の比較に影響を及ぼす効果修飾因子は（それらの影響が互いに相殺されるため）ないことを意味し，従って，治療 B に対する治療 C の相対治療効果はあらゆるターゲット集団に対して解釈可能となる。共有の効果修飾因子の仮定を置くことに対する根拠は，臨床的及び生物学的知識に基づく。（例えば，生物学的特性または作用機序を共有する）同じクラスの治療に対してこの仮定を置くことは理にかなっていると言える。

1.6.4 不確実性

MAIC や STC を用いてアンカーを有する間接比較やアンカーを有さない間接比較を実施する場合, それぞれで生じうる誤差の原因を理解することが重要である。

アンカーを有する間接比較では，試験間で効果修飾因子の分布に不均衡がない限り，相対治療効果の推定で生じうる唯一の誤差は統計的なサンプリング誤差である。このサンプリング誤差は試験規模に依存する。一方，効果修飾因子の分布に不均衡がある場合，このサンプリング誤差に加えて系統的な誤差であるバイアスが生じる。MAIC や STC を用いることにより，このバイアスの影響を制御することができるものの，バイアスを完全に制御するためには，未測定の効果修飾因子がないことが前提となる。こ

れは強い仮定ではあるものの，標準的な間接比較で前提となる仮定，すなわち，測定の有無にかかわらず，分布に不均衡がある効果修飾因子はないという仮定に比べれば弱く，状況によっては理にかなった仮定と言える。

一方，アンカーを有さない間接比較では，サンプリング誤差とともに，予後因子及び効果修飾因子の分布の試験間の不均衡に伴いバイアスが生じる。MAIC や STC を適切に用いることで，このバイアスの影響を制御することができるものの，未測定の予後因子や効果修飾因子があった場合，バイアスを完全に制御することは困難である。また，アンカーを有さない間接比較においては，その制御できなかった潜在的なバイアスの程度を評価する適切な手法がなく，得られた治療効果の正確度については不明となる。Hoaglin(32)は，アンカーを有さない間接比較において，母集団を調整する手法を適用することの適切性を示していない場合，この手法によって得られた結果を評価する価値はないとまで述べている。従って，MAIC や STC を用いてアンカーを有さない間接比較を実施する場合，試験間における未測定の予後因子や効果修飾因子の分布の不均衡に伴い生じうるバイアスの程度について考察することが重要である（TSD18 Appendix C 参照）。

1.7 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（図 1-1）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，2 章で解説するシミュレーション研究を始め，母集団調整法の更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。

【推奨事項 3】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。

【推奨事項 4】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，全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 の何れの手法を用いる場合にも，全ての効果修飾因子と予後因子を調整した方が良い。

【推奨事項 5】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。

【推奨事項 6】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。

【推奨事項 7】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。

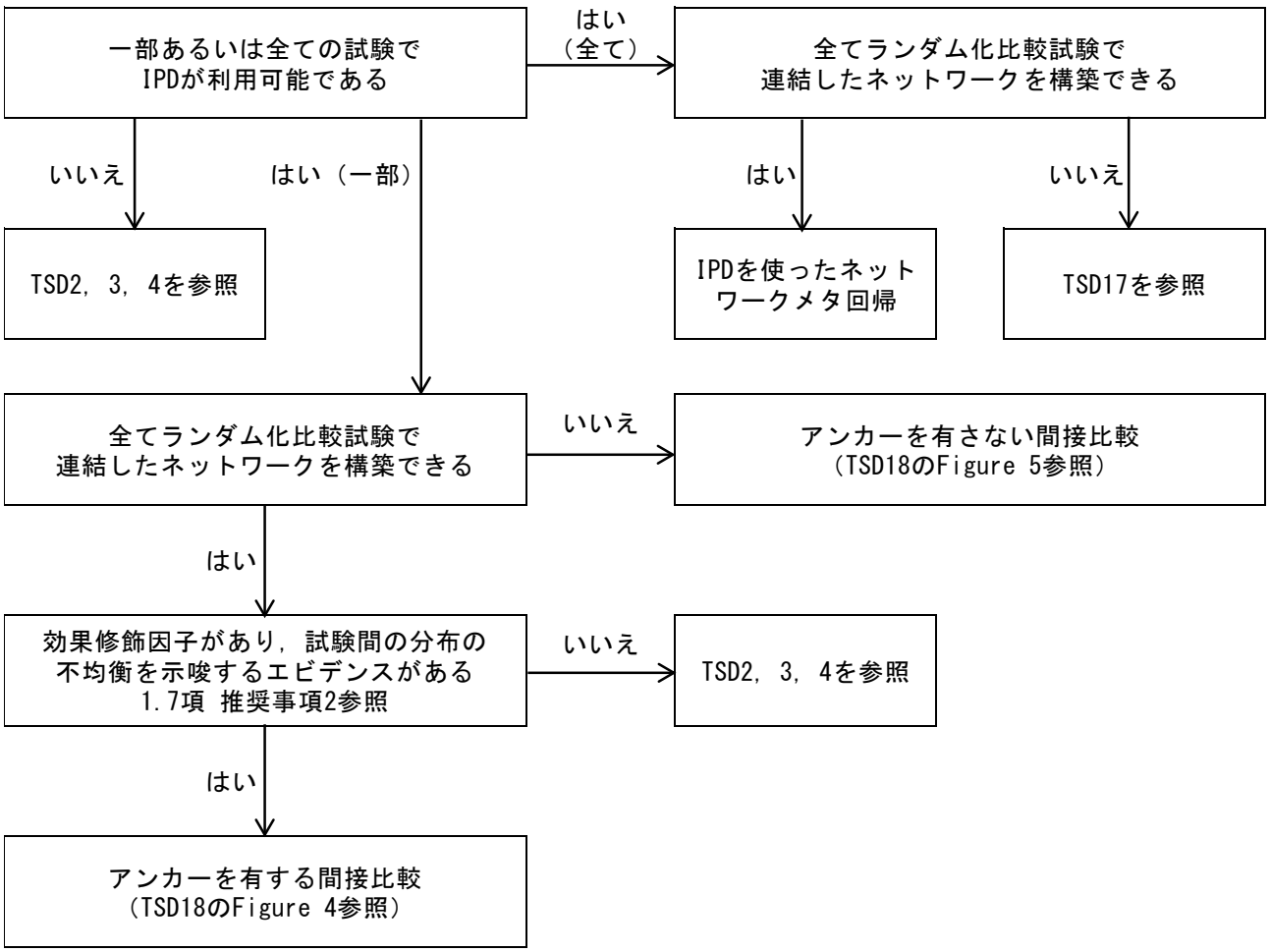


図 1-1 間接比較を実施するための手法選択のフローチャート（TSD18 の Figure 3 を一部改変）

2. シミュレーション研究の紹介

1 章での記載のとおり，MAIC 及び STC などの母集団調整法の適用には前提となる仮定が複数ある。そのため，これらの仮定が成立しない状況下では，どのような性質を持つのかが気になるところである。TSD18 では母集団調整法に関するいくつかの研究を推奨しており，そのうちのひとつとして，このような疑問に答えるためのシミュレーション研究を挙げている。手法の前提となる仮定から逸脱した状況下での母集団調整方法の性能評価や，標準的な間接比較の方法と性能を比較した評価は TSD18 発出時点(2016)では行われていなかったようであるものの，現在ではいくつかのシミュレーション研究が報告されている(26,27,33–37)。そこで本章では，母集団調整法の MAIC 及び STC に関して，アンカーを有する場合とアンカーを有さない場合の間接比較のシミュレーション研究について紹介する。なお，本章の数式の表記は紹介する論文の表記に合わせており，1 章とはやや異なっている箇所がある点をご了承いただきたい。

2.1 アンカーを有する間接比較

アンカーを有する場合の母集団調整法を用いた間接比較のシミュレーション研究は複数報告されている。特に，TSD18 の著者でもある Phillippo らは，自身の提案した階層的アプローチによるマルチレベルネットワークメタ回帰（multilevel network meta-regression：ML-NMR）（1.5.1 参照）とともに，MAIC と STC の性能を包括的に評価している(25)。その他にもいくつかの研究報告はあるものの，状況設定（シナリオ）が一部の限定されたものであるか，評価した手法が MAIC と STC のどちらかであった(26,27,33,35,36)。そこで，本節では Phillippo らによるシミュレーション研究で検討されたシナリオ及び結果の要約を紹介する。なお，著者らの提案法である ML-NMR は本報告書の範囲を超えるため，ML-NMR に関する性能評価に関する記載は割愛するものの、ML-NMR は STC と概ね同じ結果を示していた。さらに，個々のシミュレーションの結果や考察の詳細は原著論文をご参照いただきたい。なお，原著論文の付録には，本研究で用いられた R のプログラムコードが付属されている。TSD18 の付録にも R による MAIC と STC の実装例が記述されており，本報告書の付録で TSD18 の付録について紹介しているため，興味のある方はご参照いただきたい。

2.1.1 シミュレーションの設定

- 目的

2 値のアウトカムを対象に，MAIC，STC 及び標準的な間接比較の方法（Bucher 法に基づく未調整の標準的な間接比較の方法(38)）について，母集団調整法の前提となる仮定を満たすときの標準設定と仮定から逸脱するさまざまなシナリオでの性能評価が行われた。

- データ生成メカニズム

3 つの治療法（A，B，C）に対して，IPD の AB 試験，AgD の AC 試験の 2 つの試験が存在する状況

を設定する。2 値のアウトカムは，以下のロジット（対数オッズ比）モデルで生成する。ただし，AC 試験は AgD として利用するため，IPD を生成後にイベント数と共変量の要約情報のみが用いられる。

$$y_{ijk} \sim Bern\left(\theta_{ijk}\right)$$

$$logit\left(\theta_{ijk}\right) = \mu_j + q\left(\boldsymbol{x}_{ijk}^T - \boldsymbol{m}_j^T\right)\boldsymbol{\beta}_k + \gamma_k$$

$$\boldsymbol{x}_{ijk} \sim \boldsymbol{\phi}_j$$

ここで，治療kを受けている試験jの被験者 i のアウトカムy_{ijk}は，被験者のイベント発生確率θ_{ijk}を持つベルヌーイ分布に従う。アウトカムモデルのパラメータは切片μ_j，共変量$\boldsymbol{x}_{ijk}$，回帰係数$\boldsymbol{\beta}_k$，治療効果γ_kであり，標準設定での治療効果と切片は，$\gamma_A = 0$，$\gamma_B = -2$，$\gamma_C = -1.5$，$\mu_{AB} = 1$，$\mu_{AC} = 1.5$である。共変量は 2 つとし，AB 試験ではそれぞれ平均値$m_{X1(AB)} = 1$，$m_{X2(AB)} = 0.5$，標準偏差$\sigma_{X1(AB)} = 0.5$，$\sigma_{X2(AB)} = 0.1$の同時分布$\boldsymbol{\phi}_j$に従う（標準設定では相関係数 0.25 の 2 変量正規分布）。共変量$\boldsymbol{x}_{ijk}$は各試験の平均$\boldsymbol{m}_j$ によって中心化しているため，回帰係数$\boldsymbol{\beta}_k$は平均を基準とした対数オッズや対数オッズ比として解釈できる。また，関数 q(.)によって，共変量とアウトカムの関係を非線形に表現可能である（関数 q(.)は後述する）。その他の標準設定におけるパラメータの値と各シナリオに従って変更するパラメータを**表 2-1** に示した。なお，A がアンカーとなる基準治療であるため，$\boldsymbol{\beta}_A = 0$となる。シミュレーション回数は 2000 回であった。

表 2-1 検討されたシナリオ

検討項目	標準設定	標準設定からの変更
a. 症例数	500 （1 群あたり 250）	100, 1000
b. 効果修飾因子の影響	$\boldsymbol{\beta}_k = 0.1\boldsymbol{\sigma}_{X(AB)}$ 治療効果（対数オッズ比）に対して AB 試験の共変量の SD×0.1 をもつ	$\boldsymbol{\beta}_k = 0.5\boldsymbol{\sigma}_{X(AB)}$
c. 共有の効果修飾因子の状態	治療 B と C の効果修飾因子が共有: $\boldsymbol{\beta}_B = \boldsymbol{\beta}_C = 0.1\boldsymbol{\sigma}_{X(AB)}$	共有でない: $\boldsymbol{\beta}_B \neq \boldsymbol{\beta}_C$ $\boldsymbol{\beta}_B = 0.1\boldsymbol{\sigma}_{X(AB)}, \boldsymbol{\beta}_C = 0.5\boldsymbol{\sigma}_{X(AB)}$ $\boldsymbol{\beta}_B = 0.5\boldsymbol{\sigma}_{X(AB)}, \boldsymbol{\beta}_C = 0.1\boldsymbol{\sigma}_{X(AB)}$
d. 効果修飾因子の相関	$\rho_{(AB)} = \rho_{(AC)} = 0.25$	$\rho_{(AB)} = \rho_{(AC)} = 0, 0.5$
e*. 試験間のオーバーラップの程度	50%; $\kappa = 0.5$	Full overlap：AC 集団が AB集団に全て含まれる; $\kappa = 1$

検討項目	標準設定	標準設定からの変更
		No overlap : AC 集団が AB 集団の範囲外; $\kappa = 0$
f**.効果修飾因子とアウトカムの関係	線形	非線形
g. AB 試験の効果修飾因子の分布	正規分布	ガンマ分布 （標準設定と同じ平均と分散）
h*. AC 試験の効果修飾因子の分布	正規分布	ガンマ分布 （標準設定と同じ平均と分散）
i. 効果修飾因子の相関	各試験の相関係数は共通 $\rho_{(AB)} = \rho_{(AC)} = 0.25$	各試験の相関係数は異なる $\rho_{(AB)} = 0.25, \rho_{(AC)} = 0,0.5$

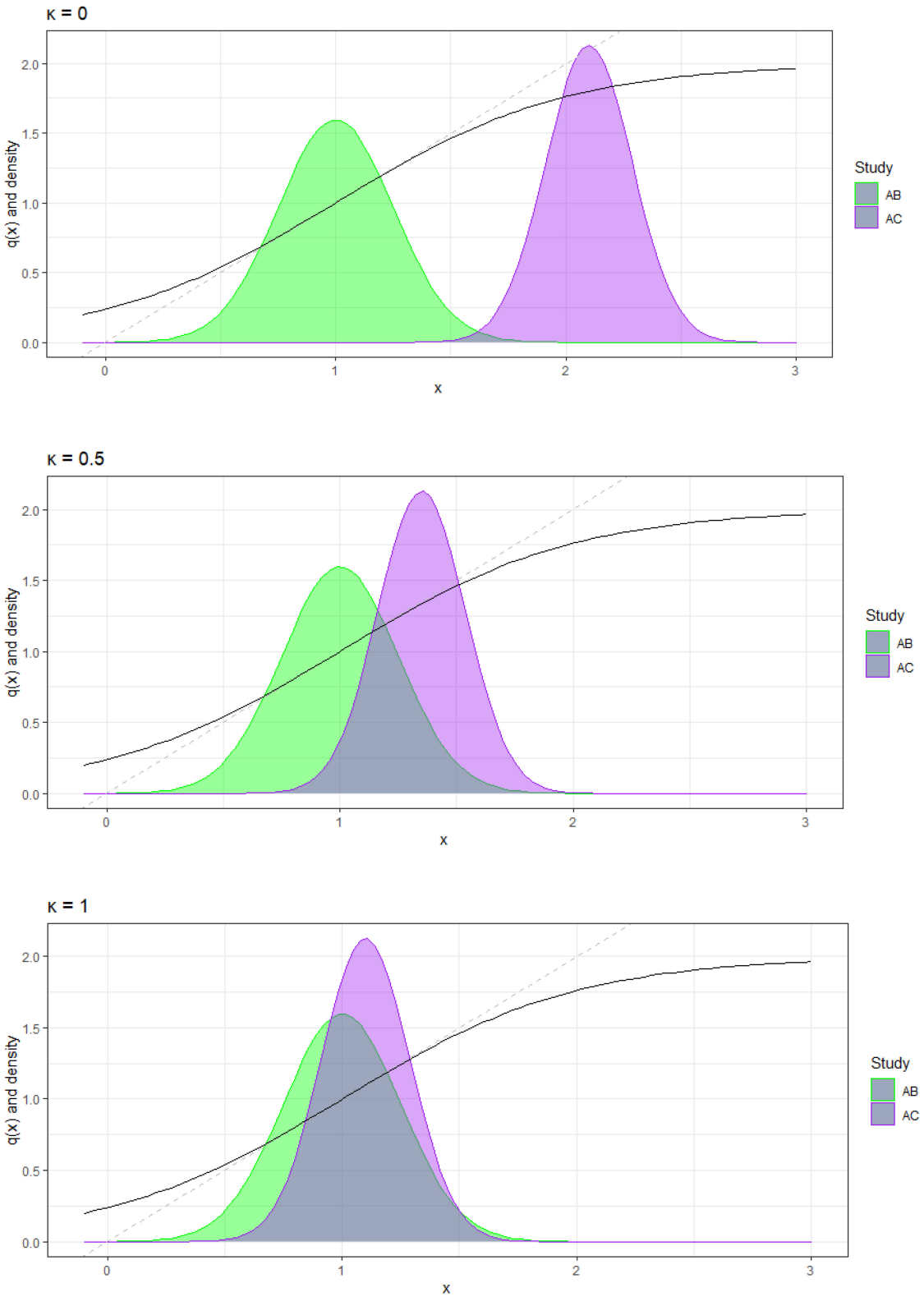
*: シナリオ e のオーバーラップの程度について，Full overlap，50% overlap，no overlap を示すパラメータをそれぞれ$\kappa = 1$，0.5，0とすると，AC 試験の共変量の平均値と SD は以下で定義される。

$$m_{X(AC)} = (1.1 + (1 - \kappa)^2) m_{X(AB)}$$
$$\sigma_{X(AC)} = 0.75\sigma_{X(AB)}$$

**: シナリオ f の共変量とアウトカムが非線形関係の場合は，以下の関数 $q(.)$を用いる。AB 試験の共変量の平均値±2SD の範囲内では $y \approx x$ の直線とみなせるが，分布の範囲外になると減衰していくシグモイド状になる関数として定義されている。

$$q(x) = 8\sigma_{X(AB)}\left(\frac{1}{1 + \exp(-(x - m_{X(AB)})/2\sigma_{X(AB)})} - \frac{1}{2}\right) + 1$$

数式では把握しにくいため，標準設定における 2 試験の共変量 X_1 の分布（つまり，正規分布）と関数$q(.)$について，オーバーラップの程度$\kappa = 0$，0.5，1の場合を**図 2-1** に示した。$\kappa = 0$のときには非線形の場合の影響が他の状況に比べて大きいことが予想できる。



実線：q(x). 点線：傾きが 1 の直線.

図 2-1 関数 $q(.)$と共変量の分布に関する 2 試験のオーバーラップの程度

- 評価方法

各方法を適用して相対治療効果を推定し，以下の指標を用いて評価する。MAIC と STC が推定できるのは，AC 試験をターゲット集団とした$d_{AB(AC)}$，$d_{BC(AC)}$と AB 試験をターゲット集団とした$d_{BC(AB)}$である。なお，MAIC では平均と分散をマッチングさせ，推定する$d_{AB(AC)}$の標準誤差にはブートストラップ法(5)を用いた。STC に用いる AB 試験の IPD に対するモデルはロジスティック回帰モデルを用いた。また母集団調整法では，共変量を 2 つとも用いる場合の他に，観測できなかった共変量がある場合を想定し，共変量を 1 つのみ用いる場合も評価した。

［評価指標］

✓ バイアス：$\frac{1}{N_{sim}}\sum_{i=1}^{N_{sim}}(\hat{d}_i - d\ )$

✓ 経験的標準誤差：$\sqrt{\frac{1}{N_{sim}-1}\sum_{i=1}^{N_{sim}}\left(\hat{d}_i - \bar{d}_i\right)^2}$

✓ モデル標準誤差：$\sqrt{\frac{1}{N_{sim}}\sum_{i=1}^{N_{sim}}\widehat{var}(\hat{d}_i)}$

✓ 被覆確率（相対治療効果の 95%信頼区間が真値を含む確率）

d：相対治療効果の真値，$\hat{d}_i$：i 回目のシミュレーションにおける d の推定値，
N_{sim}：シミュレーションの合計回数，$\bar{d}_i = \sum_{i=1}^{N_{sim}}\hat{d}_i$

2.1.2 シミュレーション結果の要約

シミュレーションの結果と著者らの考察について以下にまとめた。

STC はシミュレーション研究全体を通して，非常に似通った性能を示した。解析手法に必要な仮定が満たされていれば，標準的な間接比較で生じるバイアスを大きく減少させ，標準誤差も小さくなる結果であった。一方で，MAIC は全てのシナリオでパフォーマンスが悪く，サンプルサイズが最小の場合(n=100)には，標準的な間接比較と比較してバイアスが増大していた。MAIC の問題点は，外挿できないということに起因しているため，試験間のオーバーラップの標準設定が 50%である本シミュレーションでは，MAIC はバイアスが残ったままという結果になった。MAIC を用いてバイアスを完全に除去できるのは，AgD 試験の集団が完全に IPD 試験の集団内に含まれている場合のみである。しかしながら，2 試験の母集団が完全に重なっている場合には，効果修飾因子の不均衡の程度も小さいと思われるので，調整しなくともバイアスは比較的小さい。従って MAIC が有用な場面は，共変量の分布が AgD の集団の共変量の分布より広い試験の IPD を所有している立場から実施したときである。共変量の分布がより限定されている試験の IPD を所有している立場から実施された MAIC では，標準的な間接比較よりもバ

イアスのある推定値が得られてしまうだろう。MAIC を検討したこれまでのシミュレーション研究では，これらのオーバーラップに関する問題は言及されていなかった。過去の研究では，MAIC がうまく機能するようなオーバーラップの設定に基づいたシミュレーションのみを行っているか，連続量の共変量ではなく，2 値の共変量に焦点を当てて，IPD 試験の共変量の割合が 0 または 1 に近い場合にどの手法も機能しないことを観測したためだろうと著者は推察している。MAIC の実際の適用場面においては，サンプルサイズが小さい場合と，2 試験の共変量の分布があまりオーバーラップしていない場合に懸念があるだろうが，このような状況では MAIC の性能が悪いことが示された。共変量の分布のオーバーラップ・不均衡に起因する問題を考えると，MAIC が STC を一貫して凌駕するような現実的なシナリオはありえないと考察されている。

今回のシミュレーションでは，MAIC によるマッチングに共変量間の相関を含めることは検討されていない。これらの情報は AgD からは一般的に入手できないためである。しかし，共変量の相関の情報が得られている場合は，共変量の相関を AgD に合わせることも可能である。一方で，相関関係の違いが MAIC や STC の性能に与える影響は軽微であった。

STC は IPD の範囲を超えて外挿することができ，各集団の分布がオーバーラップしていない場合でも MAIC とは異なり推定値を得ることができる。しかし，効果修飾因子とアウトカムの関係が仮定と異なる場合には推定値はバイアスを持つ。例えば，オーバーラップしていないとき，効果修飾因子とアウトカムの関係が IPD の範囲外で非線形であるにもかかわらず，線形関係のモデルを用いた場合は，STC はバイアスのある推定値となった。なお，標準的な方法と比べると標準誤差は大きいものの，バイアスは小さい結果が得られていた。また，共有の効果修飾因子の仮定が成り立たない場合，AB 試験をターゲット集団とした$d_{BC(AB)}$のバイアスは，標準的な方法よりは若干小さいものの，どの母集団調整法でも大きいバイアスであった。

効果修飾因子が調整モデルに含まれていないとき，つまり相対治療効果の条件付き恒常性の仮定が崩れている場合，検討された全ての母集団調整法はバイアスが生じやすい。標準的な方法よりは影響は小さいものの，被覆確率が名目よりも小さくなっていた。このことからも潜在的な効果修飾因子がないか十分に検討する必要がある。全ての効果修飾因子が調整モデルに含まれている場合は，STC および ML-NMR は母集団調整された間接比較に基づく推定値を得るためのロバストな手法といえる。

2.2 アンカーを有さない間接比較

アンカーを有さない場合の母集団調整法を用いた間接比較のシミュレーション研究を調査した限り，STC を評価したものはなかった。一方で，MAIC の評価は数例報告されており(34,37)，その中でも Hatswell らが多くのシナリオで性能評価を行っている。本節では，Hatswell ら(34)よるシミュレーション研究について紹介する。詳細については原著論文をご参照いただきたい。

2.2.1 シミュレーションの設定

- 目的

著者らの調査によると, MAIC が適用された試験のほとんどがオンコロジー領域であった。そのため，生存時間データを対象に，単群の介入試験とヒストリカルコントールを用いて間接比較する際の性能評価が行われた。本研究では，介入試験の対象集団（IPD が利用可能な集団 A）の被験者は，過去の対照治療を受けた被験者（AgD が利用可能な集団 B：ヒストリカルコントロール）よりも予後が良い共変量の分布と想定し，集団 A に介入した治療方法に有利なバイアスが生じる状況を設定した。MAIC を用いて 1 次のモーメント（集団 B の平均）または 1 次及び 2 次のモーメント（集団 B の平均及び SD）でマッチングさせた場合にどのように性能が変わるかを検討するとともに，未調整の単純比較による方法，及び集団 B の IPD も利用した IPW（原著論文での表記は propensity score weighting：PSW）との比較を行う。

- データ生成メカニズム

標準設定：患者特性として独立な正規分布に従う共変量を 6 個持つと設定し，ワイブル分布に従う生存時間データを集団 A と集団 B でそれぞれ 1000 例発生させる。その後，集団 B の要約統計量を算出し，MAIC と及び単純比較で用いるデータとする。6 個の共変量のうち 4 個は観測可能な共変量としてマッチングに用いられ，2 個は未観測の共変量として設定された。

検討されたシナリオ：生存関数が指数分布の場合，標準設定と比べて，ハザード比が大きい/小さい場合，共変量の影響が大きい/小さい場合，観測可能な共変量の分布の違い（共変量同士の相関，正規分布以外：二項分布/対数正規分布/切断正規分布，集団間のオーバーラップの程度），サンプルサイズが 30/300/3000 の場合。

- 評価方法

各シナリオでそれぞれの方法を適用し，Cox 比例ハザードモデルでハザード比を推定する。評価指標として，平均パーセント誤差（mean percentage error），平均絶対パーセント誤差（mean absolute percentage error），平均標準誤差（mean standard error）及び被覆確率が用いられた。さらに，シミュレーションのそれぞれにおいて，未調整の単純比較よりもハザード比のバイアスが増加した頻度が評価指標に用いられた。

2.2.2 シミュレーション結果の要約

標準設定のシナリオでは，MAIC は単純比較よりもバイアスを改善した。1 次のモーメント（平均）をマッチングさせた MAIC では，重要な因子がマッチングに含まれていない場合にはバイアスが生じたものの，ほとんどのシナリオで IPW と同等の性能であった。また，1 次及び 2 次のモーメント（平均及び分散）をマッチングさせた MAIC では，多くのシナリオで 1 次のモーメントと同様の性能を示すものの，明確に優れている性能を示すことはなかった。それどころか，集団 A の共変量が正規分布ではない

という誤特定の状況では，1 次及び 2 次のモーメントをマッチングさせた MAIC は，単純比較の推定値よりも大きなバイアスが生じており，1 次のモーメントをマッチングさせた MAIC のほうが良い結果となった。また，集団 A と B のどちらかのサンプルサイズが 30 例と少ない場合には，MAIC のバイアスの中央値は良好に見えるが，精度は非常に悪かった（ただし，IPW でも同様である）。MAIC の被覆確率は一見高いようにも見えたが，ESS が小さくなり，結果として得られる 95%信頼区間が広くなっていただけであった。

まとめとして，著者らは次のように述べている。MAIC を適切に使用するには，十分なサンプルサイズが必要であり，試験間の患者背景などの分布も十分にオーバーラップしていることが望まれる。MAIC は欠点がないとはいえないが，今回のほとんどのシナリオでは IPW と同等の性能であった。ただし，交絡を含む状況などの今回のシナリオよりも複雑な場合には，利用可能な追加のデータを IPW に用いたり，データをトリミングしたりすることで，IPW の推定が良くなる可能性が考えられる。実務においては，2 つの試験の IPD が得られず，IPW を使用できない状況も想定される。このような場合でも，使用される状況を慎重に考慮すれば，MAIC によって，相対治療効果の正確な推定値を提供する可能性がある。しかしながら解析担当者は，このアプローチに内在する仮定を慎重に検討し，適合性を判断することを強く推奨する。

本研究での標準設定は 2 つの集団とも 1000 例という大規模なものであり，この場合には平均絶対パーセント誤差が単純比較の 11.8%と比べて 2.6%と良く改善され，単純比較と比べてバイアスが増加した頻度は 2%と小さい結果が示されている。しかしながら，2 つの集団とも 300 例のときには，平均絶対パーセント誤差の改善度は，単純比較の 12%に対して 8.7%と改善の程度が低下し，バイアスが増加した頻度は 30%に増えていた。アンカーを有さない場合，実際にはサンプルサイズが 1000 例もない試験の場合が多いことが想定されるため，シミュレーションで得られた各シナリオの評価は，サンプルサイズの影響によってもう少し大きい誤差があると考えたほうがよいかもしれない。なお，サンプルサイズが増えるにつれて精度は向上しており，IPD である集団 A のサンプルサイズの影響のほうが重要である傾向がみられた。調査した限りでは，アンカーを有さない場合の STC に対するシミュレーション研究の報告例はなかったものの，重要な共変量が存在するにもかかわらずアウトカムモデルに含まれなかった場合には，MAIC と同様に STC でもバイアスは生じると考えられる。

3. 母集団調整法の適用事例

本章では，母集団調整法の適用について 2 種類のレビューを行う。3.1 節では，現在最も一般的に使用されている母集団調整法である MAIC と STC の適用に関する文献をレビューする。3.2 節では，MAIC と STC が文献で初めて提案された 2010 年(9,11)以降に発表された全ての NICE の TA での母集団調整法の使用についてレビューする(39)。

3.1 MAIC と STC の文献での適用例

本節では，MAIC と STC が実際にどのように使用されているか，また，それらの基礎となる方法論や仮定がどの程度理解されているかを検討するために，MAIC と STC の適用論文をレビューする。適用論文は，以下の方法により Phillippo(5)により精査され確認された。

- 2016 年 7 月 7 日に Scopus と PubMed で「matching-adjusted indirect comparison」と「simulated treatment comparison」のタイトル，抄録，キーワードの中から検索
- 方法論（引用）の論文を引用した論文を確認(9–11)
- スコーピングレビュー[1](40)で確認された論文

3.1.1 MAIC の文献での適用例

MAIC(9)と STC(11)に関する最初の論文が発表されてから，これらの手法，特に MAIC の使用は比較的早くから増加した。**表 3-1** に文献検索の結果 2016 年 7 月 7 日までに確認された MAIC の適用例を 10 報示す。**表 3-1** の治療群のうち，「企業」は IPD を有し，「比較対照」では AgD を有する。また，マッチング変数の数の「with SD」はマッチング変数のうち，二次モーメントまで利用しマッチングした変数の数を表す。

[1] 入手可能な研究文献の総数や範囲についての目安を把握するための初期評価であり，研究エビデンス（現在進行中の研究を含む）の内容と規模を特定することを目指しているレビュー

表 3-1 文献での MAIC 適用例

論文 [発表年]	治療群	IPD サンプルサイズ	AgD サンプル サイズ	マッチング 変数の数	効果修飾因子の状態のエビデンスが提示されている変数の数	不均衡のエビデンスが提示されている変数の数	アンカーを有する／有しない間接比較	間接比較におけるアウトカムの尺度
Signorovitch et al .(9) [2010]	企業： Adalimumab (B) vs. プラセボ (A) 2 つの AB 試験をプール 比較対照： Etancercept (C) vs. プラセボ (A)	オリジナル：1359 選択基準内：1025 ESS：591	330	10 (5 with SD)	2 臨床的理由に基づく	4 統計的に有意	アンカーを有する	反応率，PASI 変化率
Chang et al. (41) [2011]	企業： Bevacizumab + cisplatin (B) 比較対照： Pemetrexed + cisplatin (C) 2 つの単群試験	オリジナル：2172 選択基準内：72 ESS：46	67	2 (0 with SD)	0 潜在的な予後因子と記載	2 数値的に異なる	アンカーを有しない	PFS 中央値
Signorovitch et al. (42) [2011]	企業： Nilotinib (B) vs. Imatinib (A) 比較対照： Dasatinib (C) vs, Imatinib (A)	オリジナル：A 282，B 283 選択基準内：A 280，B 273 ESS：報告無し	A： 260 C：259	10 (0 with SD)	0	3 数値的に異なる	アンカーを有しない	1 年時の PFS，OS，MMR の割合
Signorovitch et al. (43) [2011]	企業： Vildagliptin (B) vs.プラセボ (A) Vildagliptin (B) vs. Voglibose (D) 比較対照： Sitagliptin (C) vs. プラセボ (A) Sitagliptin (C) vs. Voglibose (D) 2 つの AC 試験を用量ごとにプール	オリジナル：AB 148, BD 380 選択基準内：AB 148, BD 363 ESS：報告無し	AC：145 CD：319	6 (5 with SD)	0 過去のメタアナリシスで異質性が大きかったとの記載あり	3 統計的に有意	アンカーを有する プラセボと Voglibose を共通の比較対照として MAIC をそれぞれ実施し結果をプールした	HbA1c 平均値
Kirson et al. (44) [2013]	企業： Adalimumab (B) vs. プラセボ (A) 比較対照： Etancercept (C) vs. プラセボ (A) Infliximab (D) vs. プラセボ (A)	オリジナル：313 選択基準内：AC 296, AD 234 ESS：報告無し	AC：205 AD：200	For AC: 12 (6 with SD) For AD: 17 (11 with SD)	0	AC: 2 AD: 4 統計的に有意	アンカーを有する	反応率，変化率
Signorovitch et al. (45) [2013]	企業： Everolimus (B) vs. プラセボ (A) 比較対照： Sunitinib (C) vs. プラセボ (A)	オリジナル：410 選択基準内：394 ESS：報告無し	171	9 (0 with SD)	0	3 統計的に有意	PFS：アンカーを有する OS：アンカーを有しない	対数ハザード比
Sikirica et al. (46) [2013]	企業： Guanfacine (B) vs. プラセボ (A) AB をプールした集団 比較対照： Atomoxetine (C) vs. プラセボ (A)	オリジナル：631 選択基準内：A 136，B 82 ESS：:報告無し	A：83 C：84	4 (4 with SD)	0	1 統計的に有意	アンカーを有しない 重みは，プラセボ群が正確に一致するように制約	平均 ADHD スコア
Sherman et al. (47) [2015]	企業： Everolimus (B) 比較対照： Axitinib (C) 共通の比較対照は無し，他の群は無	オリジナル：277 選択基準内：43 ESS：報告無し	194	3	0 潜在クラスモデルを用い PFS に影響を与える変数を抽出	3 数値的に異なる	アンカーを有しない	PFS 中央値

論文 [発表年]	治療群	IPD サンプルサイズ	AgD サンプル サイズ	マッチング 変数の数	効果修飾因子の状態のエビデンスが提示されている変数の数	不均衡のエビデンスが提示されている変数の数	アンカーを有する／有しない間接比較	間接比較におけるアウトカムの尺度
	視された							
Van Sanden et al. (48) [2016]	企業： Simeprevir + peginterferon alfa 2a + ribavirin (B) 比較対照： Peginterferon alfa 2a + ribavirin (C1–C5) 単群試験，複数の C 集団	オリジナル：107 選択基準内:C1 35，C2 35，C3 57 ，C4 35，C5 19 ESS:C1 29，C2 15，C3 14，C4 26，C5 17	C1：30 C2：18 C3：95 C4：40 C5：109	5–6	0 経験豊富な 2 人の肝臓専門医に治療反応に関連する変数を相談	数個 数値的に異なる	アンカーを有しない	持続的なウイルス学的奏効を達成した割合
Swallow et al. (49) [2016]	企業： Daclatasvir + sofosbuvir (B) 比較対照： Sofosbuvir + ribavirin (C) 2 つの C 集団をプール， 全てオープン単群試験	オリジナル：153 選択基準内：91 ESS：報告無し	455	14 (3 with SD)	0	4 統計的に有意	アンカーを有しない	持続的なウイルス学的奏効を達成した割合

3.1.2　アンカーを有する間接比較とアンカーを有さない間接比較

10 報のうち 6 報においては，共通の比較対照を有するランダム化比較試験が含まれていた。6 報中 4 報では，アンカーを有する間接比較が実施された。また，6 報中 3 報では，アンカーを有さない間接比較が行われていた（1 報では異なるアウトカムについてアンカーを有する間接比較とアンカーを有さない間接比較の両方が行われていた）。アンカーを有さない間接比較が行われた 3 報中 2 報は治療スイッチングを許容した試験であり，プラセボ群で疾患の進行が確認された後に実薬群に切り替えられた患者が含まれていた。そのため全生存期間（OS）が正しく評価できず，アンカーを有さない間接比較が用いられた。OS ではなく無増悪生存期間（PFS）に変更することで，この問題は回避可能であるが，経済モデルでは通常 OS の推定値が必要である。Signorovitch ら(45)の分析では，PFS に対するアンカーを有する間接比較と OS に対するアンカーを有さない間接比較が行われている。OS の解析において治療スイッチングを考慮する方法としては，rank-preserving structural failure time model(50)，反復パラメータ推定(52)，inverse probability of censoring weighting(51)などがあるが，いずれの解析においてもこれらのアプローチを利用していない［治療スイッチングの概要と NICE の TA での使用に関するガイダンスは Latimer and Abrams（TSD16(52)），日本製薬工業協会の報告書(53)によって示されている］。

Sikirica ら(46)の分析では，2 つの試験間でプラセボ群が共通であるにもかかわらず，アンカーを有さない間接比較を行っていた。マッチング手順において，試験間でプラセボ群のアウトカムが正確にバランスがとられるように重み付けを行っていることで正当化している。この方法はまだシミュレーション研究などでも評価されたものではないため，アンカーを有する間接比較と比較した場合の特性や性能は不確かである。特に，プラセボ群のアウトカムのバランスをとることが，観察されていない予後因子によるバイアスを除去するためのランダム化と同等であるとは考えにくい。

残りの 4 報では，単群試験や共通の比較対照がない状況下で MAIC を適用していた。このような場合はアンカーを有さない間接比較を行うことが唯一の選択肢となる。アンカーを有さない間接比較が実施される他の全ての場合と同様に，全ての予後因子と全ての効果修飾因子が明らかとなり，正確に特定されているという非常に強い仮定が必要である。アンカーを有さない MAIC の適用例は，観察されていない予後因子と効果修飾因子による残留バイアス（除ききれていないバイアス）の可能性を認識しているが，残留バイアスの潜在的な大きさの分析はなされてない。従って，アンカーを有さない MAIC の推定値の誤差の程度については不明であるが，結果として得られる推定値の正確性が完全に未知であることには言及していない。さらに，単群試験を分析に含めることは，これらの研究デザインによって生じる追加の仮定やバイアスの影響を受けてしまうことになる(54)。

3.1.3　複数の研究の利用可能性

公表された分析の半数では，複数の IPD または同じ治療群の複数の AgD の取り扱いにおいて問題が生じていた。IPD を持つ複数のデータが利用可能な場合，データは単純にプールされ，1 つの大きなデータとして扱われた。この場合，構成された試験間の違いを考慮しておらず，IPD のメタアナリシスでは，バイアスを発生させ検出力を低下させることが確認されている(55)。試験間の違いを考慮した MAIC の

手法がない場合のより良い選択肢は，各 IPD に基づいて同一の MAIC を実施し，標準的な（ネットワーク）メタアナリシス手法を用いて相対治療効果の推定値をプールすることである(56–60)。

複数の AgD が存在した 3 報中 2 報ではプールされ，残り 1 報では別々に分析されていた。AgD をプールする場合には，尺度のコンフリクトを避けるために，常に線形予測子の尺度での相対治療効果で行われるべきである。そのような分析方法として次の 2 つの方法がある。

(i) 各 AgD に同一の MAIC を実行し，$\hat{d}_{BC}$相対治療効果推定値をプールする。

(ii) AgD と $\hat{d}_{AC}$相対治療効果推定値をプールし，プールされたデータに単一の MAIC を実行する。

いずれの場合も，相対治療効果推定値のプールは，標準的な方法(56–60)を用いて線形予測子の尺度で行われるべきであり，結果として得られるターゲット集団は AgD の（適切に加重された）組み合わせとなる。この得られた集団が，意思決定のために用いる真のターゲット集団と一致するかどうかは分からない。AgD ごとに個別の分析を行った場合，結果として得られる推計値は，それぞれ異なるターゲット集団（すなわち，各 AgD）に対して有効であり，ターゲット集団が効果修飾因子のバランスのとれた分布を持っていない限り，比較可能なものではないことに留意すべきである。

3.1.4 より大きな治療ネットワーク

公表された分析のうち 2 報では，3 つ以上の治療法を含む分析が行われている。Signorovitch ら(43)は 4 つの治療法を四角いネットワーク（**図 3-1** (a)）に配置しており，基本的には関心のある治療法 B と C の間に 2 つの共通の比較対照（プラセボと別の有効な治療法）を与えている。Kirson ら(45)の別の例では，スターネットワーク（**図 3-1** (b)）で 4 つの治療法を配置していたが，この場合は B との間接比較を行うために 2 つの対照治療として C と D を含んでいた。

Signorovitch ら(43)では，AB 試験と BD 試験の IPD と，AC 試験と CD 試験の AgD を利用した。従って，2 つの MAIC 分析が可能であった。1 つは治療法 A によるもの，もう 1 つは治療法 D によるものである。結果として得られた 2 つの間接比較推定値は，AC 試験と CD 試験の異なる対象集団に対して有効であり，それらはプールされた。この場合の MAIC のターゲット集団は，AC と CD の集団の加重組み合わせであり，意思決定のために用いる真のターゲット集団と一致する可能性は低い。

Kirson ら(45)では，間接比較を構成するために AC 試験と AD 試験の AgD を持つ 2 つの対照治療 C と D があるという同様のシナリオに直面した。この場合も 2 つの MAIC が実行され，今回は $d_{BC(AC)}$と $d_{BD(AD)}$の推定値が得られた。これらの相対治療効果の推定値は，2 つのターゲット集団が効果修飾因子のバランスのとれた分布を持っていない限り，異なるターゲット集団（AC と AD それぞれ）に対してそれぞれで有効であるため，比較可能ではない。現在の MAIC では，AC と AD の試験集団が効果修飾因子の点で異なる場合，この場合の 4 つの治療法全てに対する首尾一貫した比較を達成する方法はない。

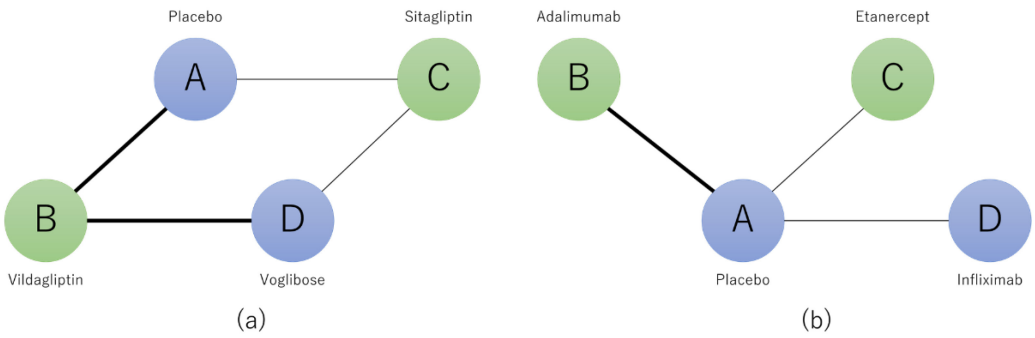


図 3-1 ネットワークの例

* (a) Signorovitch ら(43)は，異なる共通対照を介して 2 つの MAIC を実行し，(b) Kirson ら(44)は，2 つの異なる対照治療について 2 つの MAIC を実行する。太い線は IPD が利用可能であることを示し，細い線は AgD のみが利用可能であることを示す。

3.1.5　有効サンプルサイズと重み分布

公表された MAIC の分析のうち，有効サンプルサイズまたは重み分布について言及しているのはわずか 4 報であった。3 報は有効サンプルサイズ，残りの 1 報には重み分布の要約が含まれていた。これらのうち少なくとも 1 つを報告することは，IPD と AgD の間の重なり具合を理解し診断するための基本となる。有効サンプルサイズが著しく減少した場合，あるいは重みが大きく変動する場合に推定値は不安定になり，推定値は一部の個人データに大きく依存することになる。有効サンプルサイズを報告している 3 報の論文では，元のサンプルサイズから平均 80%の減少が見られた（範囲：57%～98%）。有効サンプルサイズの絶対値は 14～591 の範囲であった。

3.1.6　マッチング変数の選択

公表された MAIC の分析で使用されたマッチング変数の数は 2～17 個の範囲であった。ほとんどの分析では，共変量の標準偏差，平均または集団間の他の要約統計量とのバランスが取れていたが，重み付けモデルに交互作用項または高次項を含んでいたのは 1 つだけであった(46)。従って，公表された MAIC の分析の大部分では，同時分布の代わりに周辺共変量分布を使用しているため，より強い仮定に従うことになる。特に，集団間の共変量相関のバランスに関して，あるいは間接比較の尺度で誘導される潜在的なアウトカムモデルの交互作用項の欠如に関して，追加の仮定がなされなければならない。

臨床的専門知識や事前の経験的エビデンスを用いて，重み付きモデルに含まれる変数が効果修飾因子であることを正当化しようとしたアンカーを有する分析は 1 報もなかった。NICE の方法ガイド（NICE 2013）(31)では，効果修飾因子の状態は解析の前に正当化されるべきであると明示されている。アンカーを有さない比較では，効果修飾因子と同様に全ての予後因子が含まれるべきである。含まれる変数が予後または効果修飾因子であると正当化されたのは 3 報だけであった。

一般的に，公表済みのアンカーを有する MAIC では，重み付き調整前後の比較推定値が報告されており，その違いに言及している。しかし，分析後の相対治療効果の違いに関する考察は，アンカーを有す

る MAIC が標準的な間接比較よりも好ましいことを正当化するために用いられるべきではない。NICE TA では，本来全ての分析が明確に事前規定されるべきである。よって，このような議論はその場しのぎの推論にすぎない（NICE 2013）(31)。効果修飾因子の不均衡が間接比較に与える影響の大きさの評価は，どの分析でも事前にも行われていなかった。

共通のプラセボ群が存在するケースでは，MAIC の妥当性を正当化する試みとして placebo test が実施されていた。しかしながら，このようなテストは，観察されているかどうかにかかわらず，予後因子の不均衡のみを検出することができ，効果修飾因子の不均衡を完全に検出することはできない。この状況での placebo test が付加価値をもたらすかどうかは議論の余地がある。アンカーを有する間接比較は，設計上集団間の予後因子の違いを考慮しており，不均衡な予後因子は間接比較のバイアスにはつながらないが，placebo test が失敗する原因となりうる。placebo test は，その低い検出力のため，アンカーを有さない間接比較を正当化するために使用されるべきではない。

3.1.7 尺度の選択

間接比較の尺度の選択は，効果の加法性，予後因子と効果修飾因子の定義，分布特性に関し仮定が含まれているため重要である。ほとんど全ての公表された MAIC では，アウトカムの尺度のままで間接比較が実施された。このため多くの場合では，対数オッズ比ではなく確率差など，メタアナリシスでは一般的に使用されない尺度で間接比較が行われた。メタアナリシスと同様に，適切な尺度は生物学的及び臨床的知識に照らし合わせてケースバイケースで検討されるべきであり，基準の尺度は既存の文献を参考に決定されるべきである。

3.1.8 STC の文献での適用例

今回の文献検索の結果では，STC を用いた事例は 1 報のみであった。Nixon ら(61)は，再発寛解型多発性硬化症の経口治療の分析で STC を用いた。ネットワーク図を**図 3-2** に示す。AB 集団は，IPD を持つ 2 つの試験において fingolimod（B）またはプラセボ（A）のいずれかに無作為に割り付けられた 1556 人の患者で構成されている。プールされた IPD を用いた MAIC とは異なり，Nixon らはアウトカムモデルに試験レベルのベースラインリスク項（各 AB 試験ごとの独立した切片）を含めることで IPD のクラスタリングを正しく説明した。他の 3 試験は AgD のみを有し，そのうち 2 つの試験では合計 2301 人の患者を対象に dimethyl fumarate（C）とプラセボを比較し，もう 1 つの試験では 1088 人の患者を対象に teriflunomide（D）とプラセボを比較した。2 つの AC 試験のリスク比と共変量分布は，逆分散重み付け解析（本質的には 2 つの試験の固定効果メタアナリシス）を用いて単純にプールされた。AC 試験と AD 試験の間の共変量とアウトカムの定義の違いにより，Nixon らは 2 つの STC モデルを作成し，一方は AC 集団への予測に AC の定義を用い，もう一方は AD 集団への予測に AD の定義を用いた。

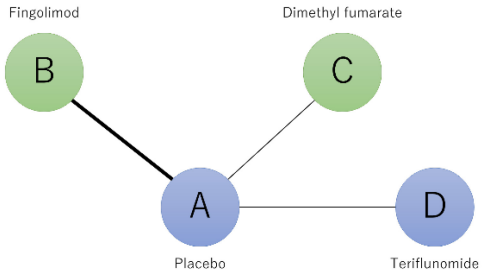


図 3-2 Nixon et al. (2014)による STC 解析のネットワーク図

* 太い線は IPD が利用可能であることを示し，細い線は AgD のみが利用可能であることを示す。

MAIC と STC に関する全ての公表された文献の中で，Nixon ら(61)のみが，全変数の効果修飾因子の状態を正当化しようとした。専門家の臨床的意見と過去のサブグループ解析結果の両方がエビデンスとして使用された。3 集団間の共変量の不均衡に関する分析は行われなかったが，最終モデルを選択するために AIC ベースの変数減少法による選択アルゴリズムを使用したことから，残りの共変量がアウトカムの有意な予測因子であることが示唆された。結果モデル自体は線形確率モデルであり，共変量に対する反応の確率を回帰させるために恒等リンク関数を使用している。先に述べたように，これはあまり一般的ではないモデルの選択であり，そのようなモデルでは 0 から 1 の範囲外となってしまう予測確率をもたらす可能性がある。同様に，この場合のモデル・尺度の選択は，（対数）相対リスク尺度で構成されるアンカーを有する間接比較での問題にもつながる。アウトカム回帰では，予後因子（及び効果修飾因子）は，線形確率尺度に関して定義されるが，アンカーを有する間接比較で対数相対リスクを使用することは，予後因子がうまく相殺されないことを意味する。

3.2 NICE 技術評価（TA）での適用例

母集団調整法の使用は，NICE TA において普及してきているが，限られた IPD しか利用できないのが一般的である。規制当局や償還機関に提出する企業は，自社の試験では利用可能な IPD を保有しているが，競合他社の試験は公表された AgD しか利用できない可能性が高い。本節では，母集団調整法の使用状況を示すことを目的として，NICE(62)が発表した TA のレビューを行う。単に母集団調整法の使用状況を調査するだけでなく，実務上，特に意思決定の際の母集団調整法の妥当性を評価するために，主たる仮定がどの程度満たされて使用されているのかに注目した。これらの方法がアプレイザル委員会ではどのように受け止められているか，またそれらが意思決定にどのような影響を与えているかについても示す。

3.2.1 レビューの概要

2010 年 1 月 1 日から 2018 年 4 月 20 日までに公表された全ての NICE の TA の母集団調整法の使用有無について Phillippo ら(39)によりレビューされた。対象となった全ての研究に対して IPD へのアクセス

が可能な TA を除外し，IPD が部分的にしか利用できない TA に絞った。1 つ以上の母集団調整法を使用した TA の企業の提出資料から以下の情報を抽出した。

- 母集団調整法名
- アンカーの有無
- アウトカムのタイプ
- 臨床領域
- 調整された共変量の数
- 共変量の選択方法
- MAIC を用いた TA の場合，重み付け後の有効サンプルサイズ
- より大きなネットワーク構造の存在の有無，存在する場合はその取扱い方法

3.2.2 レビューの結果

MAIC と STC に関する最初の論文が発表された 2010 年以降(9,11), 2018 年 4 月 20 日までに, 合計 268 件の TA が NICE により公表された。これらのうち 21 件で母集団調整法が用いられていたが，そのうち 3 件では含まれる全ての研究で IPD が利用可能であったため，部分的な IPD のみを有する残りの 18 件の評価に絞った。**図 3-3** に選定のプロセスを示す。対象となった TA は**表 3-2** にまとめた。

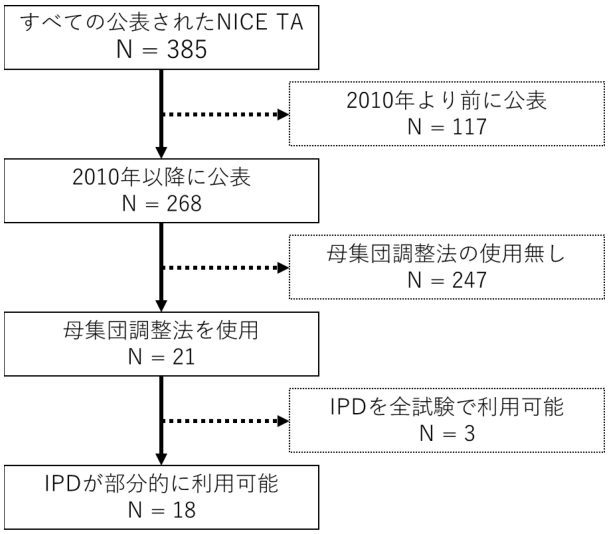


図 3-3 TA の選定プロセスのフローチャート

表 3-2 NICE TA における母集団調整の適用例

共変量列の記号は，隣接する列で詳述されたアプローチで選択された共変量の部分集団を示す

TA [公表年]	公表 時期	母集団調整法	アンカーを有する／有しない間接比較	臨床領域	アウトカムの種類	共変量	どのように共変量が選ばれたか	MAIC 有効サンプルサイズ (%)	利用可能な場合，より大きなネットワークはどのように取り扱ったか
TA510:Daratumumab monotherapy for treating relapsed and refractory multiple myeloma(63) [2018]	2018 年 3 月	MAIC	アンカーを有しない	がん領域	生存時間	不応状態*†，ECOG*†，前治療歴*，クレアチニンクリアランス*，診断からの経過時間，骨髄腫サブタイプ，人種，骨病変，ASCT の既往，年齢	文献レビューと専門家の意見を用いて，共変量を重要度でランク付けし，次に変数増加法で選択した。 MAIC は 2 つのターゲット試験で実施され，最初の試験では 4 つの共変量(*)，2 つ目の試験では 2 つの共変量(†)で調整した。	84 (57.8), 80 (54.1)	複数の母集団を調整した間接比較が実施された後，対象集団の結果のみ提示
TA500: Ceritinib for untreated ALK-positive non-small-cell lung cancer (64) [2018]	2018 年 1 月	MAIC	アンカーを有しない	がん領域	生存時間	年齢，性別，人種，喫煙状況，腺がん，ECOG，転移性疾患，脳転移	両試験で利用可能な全てのベースライン特性	171 (90.4), 174 (92.1)	複数の母集団を調整した間接比較が実施された後，対象集団の結果のみ提示
TA492:Atezolizumab for untreated locally advanced or metastatic urothelial cancer when cisplatin is unsuitable (65) [2017]	2017 年 12 月	STC, MAIC	アンカーを有しない	がん領域	生存時間	年齢*，性別*，ECOG*，肝転移*，前治療歴数	STC は，交差検証された予測性能を最大化するために共変量(*)を選択した。 MAIC は利用可能な全ての共変量を使用した。	-	ネットワークに接続するために各試験で実薬群を予測し，NMA を使用して分析した
TA462:Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma (66) [2017]	2017 年 7 月	MAIC	アンカーを有しない	がん領域	生存時間，反応率	年齢，性別，病期，B 症状，ヘモグロビン，リンパ球，白血球数，アルブミン，結節外部位，ECOG，腫瘍径，前治療歴数	-	81 (42)	-
TA478:Brentuximab vedotin for treating relapsed or refractory systemic anaplastic large cell lymphoma (67) [2017]	2017 年 10 月	MAIC	アンカーを有しない	がん領域	生存時間	年齢，性別，乳酸脱水素酵素上昇，病期，ECOG，一次治療への反応	両試験で利用可能な全てのベースライン特性	4.8 (8.3)	-
TA457:Carfilzomib for previously treated multiple myeloma (68) [2017]	2017 年 7 月	MAIC	アンカーを有しない	がん領域	生存時間	年齢，ISS 病期，診断からの経過時間，クレアチニンクリアランス，前治療歴，前治療歴，前治療歴 SCT，前治療歴ボルテゾミブ，前治療歴 IMiD，前治療不応性	英国の臨床専門家によって予後因子として同定された共変量が両研究で報告された	335.5 (52)	-

TA [公表年]	公表 時期	母集団調整法	アンカーを有する／有しない間接比較	臨床領域	アウトカムの種類	共変量	どのように共変量が選ばれたか	MAIC 有効サンプルサイズ (%)	利用可能な場合，より大きなネットワークはどのように取り扱ったか
TA449:Everolimus and sunitinib for treating unresectable or metastatic neuroendocrine tumours in people with progressive disease (69) [2017]	2017 年 6 月	MAIC	アンカーを有する	がん領域	生存時間	-	-	-	-
TA451: Ponatinib for treating chronic myeloid leukaemia and acute lymphoblastic leukaemia (70) [2017]	2017 年 6 月	MAIC	アンカーを有しない	がん領域	反応率，持続期間	年齢，性別，T315I 変異，人種，罹患期間，ECOG	両試験で利用可能な全てのベースライン特性	69 (25.8)	-
TA432:Everolimus for advanced renal cell carcinoma after previous treatment (71) [2017]	2017 年 2 月	MAIC	アンカーを有しない	がん領域	生存時間	-	-	-	-
TA429: Ibrutinib for previously treated chronic lymphocytic leukaemia and untreated chronic lymphocytic leukaemia with 17p deletion or TP53 mutation (72) [2017]	2017 年 1 月	MAIC	アンカーを有しない	がん領域	生存時間	17p 欠失状態，前治療数，プリン体再充填状態，年齢，binet/RAI，IGVH 状態，β2-マイクログロブリン，del 11q，クレアチニンクリアランス，血小板，性別，ヘモグロビン，リンパ球	両試験で利用可能な臨床的に関連性のあるベースライン特性は全て，臨床専門家によってレビューされた。	30 (15.4)	複数の母集団を調整した間接比較が実施された後，対象集団の結果のみ提示
TA427:Pomalidomide for multiple myeloma previously treated with lenalidomide and bortezomib (73) [2017]	2017 年 1 月	MAIC	アンカーを有しない	がん領域	生存時間	年齢，ECOG，前治療数，サリドマイドの前治療歴	-	-	追加の独立した比較に使用される他の単群 IPD データ
TA410:Talimogene laherparepvec for treating unresectable metastatic melanoma (74) [2016]	2016 年 9 月	予測モデル	アンカーを有しない	がん領域	生存時間	性別，ECOG，内臓状態，脳転移，LDH	公表されている修正予測モデル（Korn モデル）	-	-
TA383: TNF-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis (75) [2016]	2016 年 2 月	MAIC, STC	アンカーを有しない	リウマチ領域	ベースラインからの変化量	性別，人種，年齢，DMARD の併用，NSAID の併用，HLA-B27，BASDAI，BASFI，ASDAS，CRP	両試験で利用可能な全てのベースライン特性	-	複数の母集団を調整した間接比較が実施された後，対象集団の結果のみ提示
TA380:Panobinostat for treating multiple myeloma after at least 2 previous treatments (76) [2016]	2016 年 1 月	MAIC	アンカーを有しない	がん領域	生存時間	年齢，性別，診断からの経過時間，ECOG，前治療歴の数，サリドマイド前，ボルテゾミブ前，幹細胞移植前，β2-マイクログロブリン	両試験で利用可能な全てのベースライン特性	137 (35.4), 23 (5.9)	利用可能な試験のネットワーク，ターゲットの単一の比較に MAIC を使用

TA [公表年]	公表 時期	母集団調整法	アンカーを有する／有しない間接比較	臨床領域	アウトカムの種類	共変量	どのように共変量が選ばれたか	MAIC 有効サンプルサイズ (%)	利用可能な場合，より大きなネットワークはどのように取り扱ったか
TA364:Daclatasvir for treating chronic hepatitis C (77) [2015]	2015 年 11 月	MAIC	アンカーを有しない	肝臓病領域	割合	年齢，BMI，人種，性別，HCV 遺伝子型，血漿 HCV RNA，線維化ステージング，IL28B 遺伝子型，プラテッツ，ALT，前治療歴	両試験で全てのベースライン特性が利用可能 。一方の MAIC では，1 つの共変量（HCV RNA）を調整するのに十分なサンプルがなかった。	-	不明瞭
TA331: Simeprevir in combination with peginterferon alfa and ribavirin for treating genotypes 1 and 4 chronic hepatitis C (78) [2015]	2015 年 2 月	MAIC	アンカーを有しない	肝臓病領域	割合	線維症スコア，ウイルス量，BMI，年齢，性別	-	15 (14)	複数の母集団を調整した間接比較が実施された後，対象集団の結果のみ提示
TA333: Axitinib for treating advanced renal cell carcinoma after failure of prior systemic treatment (79) [2015]	2015 年 2 月	STC	アンカーを有しない	がん領域	生存時間	性別，年齢*，腎摘出術の有無，放射線治療歴，サイトカイン治療歴，MSKCC* †，明細胞癌，ECOG，サブチニブ投与時間	回帰モデルにおける PFS（*）と OS（†）の有意なアウトカム予測因子（p＜0.1）。	-	-
TA311:Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation (80) [2014]	2014 年 4 月	MAIC	アンカーを有しない	がん領域	反応率	ISS 病期，β2 ミクログロブリン，細胞遺伝学的異常 t4，年齢，性別，軽鎖骨髄腫，IG-A，IG-D，IG-G	-	-	利用可能な試験のネットワーク，ターゲットの単一の比較に MAIC を使用

最初に母集団調整が使用された TA は，2014 年の TA311(80)であった。それ以降，TA における母集団調整法の使用は増加しており，母集団調整法を利用した TA 数とその割合のいずれも増加している（**図 3-4**）。2017 年に母集団調整法を使用した TA は 9 件で，同年の全 TA の 14.5%を占めている。

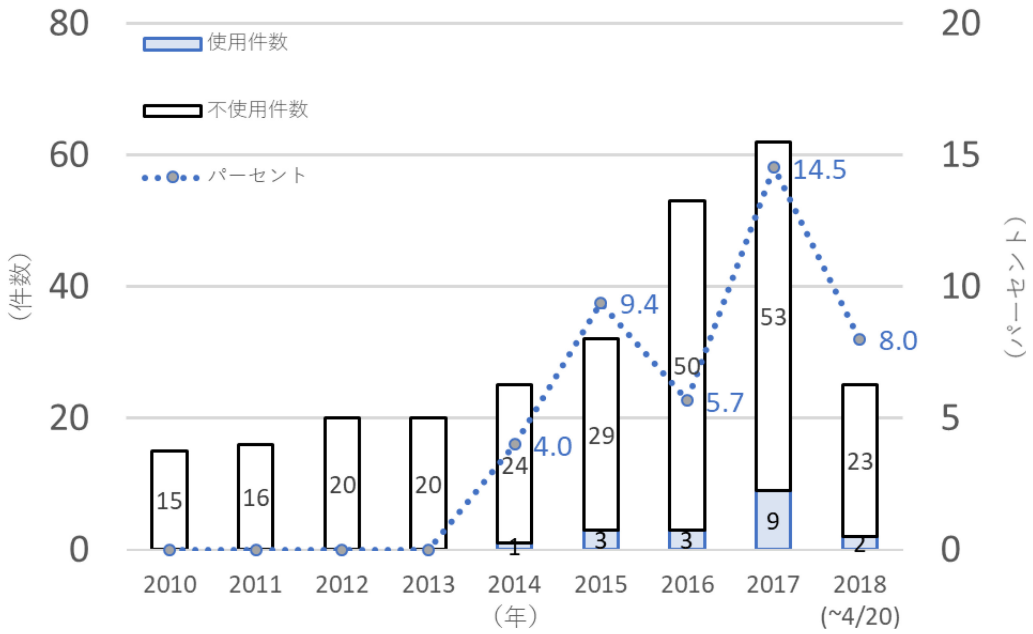


図 3-4 母集団調整法が使用された TA の件数とパーセント

3.2.3 臨床領域別の使用状況

2010 年以降，公表された TA のほぼ半数ががん領域であった（268 件中 127 件，47.4%）。そのうち 15 件（11.8%）が母集団調整法を使用しており，これまでの TA における母集団調整法の使用の 80%以上を占めている。他の臨床領域で母集団調整法が使用されたのは 3 件のみであった。肝臓病領域の 12 件中 2 件（16.7%）（いずれも C 型肝炎）とリウマチ領域の 28 件中 1 件（3.6%）であった。がん領域の TA における母集団調整法の使用は，2010 年以降使用している TA の件数と割合いずれも増加している。2014 年は 1 件（9.1%）であったが，2017 年には 9 件（25.7%）の TA で母集団調整法が使用されている（**図 3-5**）。全体の中で最も大きな割合を占めるがん領域の TA での母集団調整法の利用増加が，**図 3-4** における全体での利用増加の要因となっている。

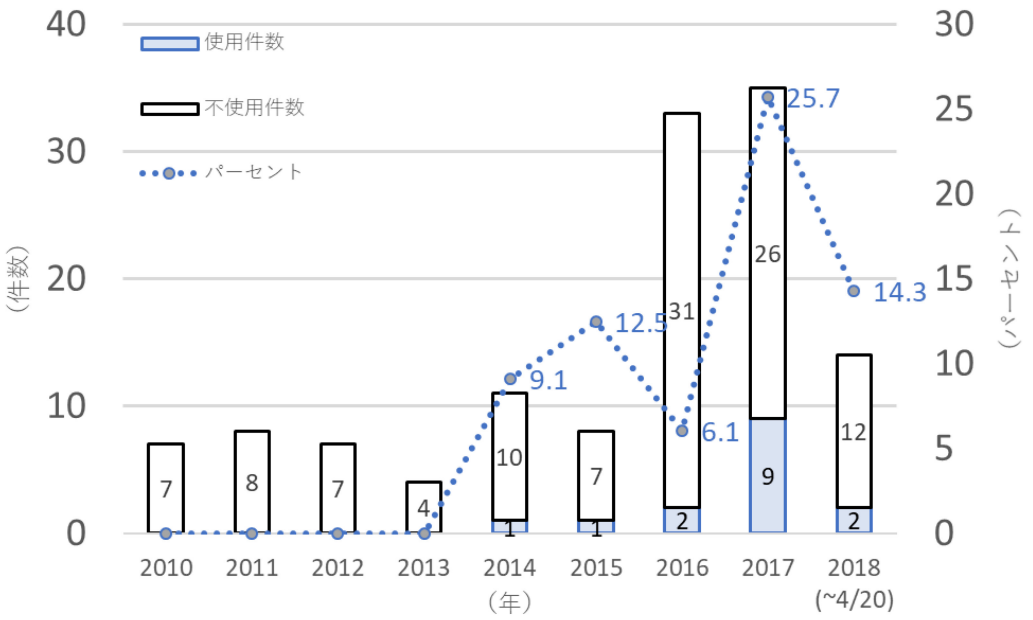


図 3-5 がん領域で母集団調整法が使用された TA の件数とパーセント

3.2.4 アウトカムの種類

母集団調整法ががん領域の TA で大部分が用いられているため当然ではあるが，母集団調整法で使用されるアウトカムの種類としては，生存アウトカム〔無増悪生存（PFS），全生存（OS）など〕が最も多く利用された。18 件の TA のうち 13 件（72.2%）には母集団調整生存率が含まれていた。5 件の TA では奏効率などの比率アウトカムが用いられ，持続期間とベースラインからの変化量が 1 件ずつで用いられた。TA462(66)，TA451(70)では，2 種類以上のアウトカム（生存率と奏効率，奏効率と持続時間）に対して母集団調整法が用いられていた。

3.2.5 母集団調整法

何らかの母集団調整法を使用している TA の大多数では MAIC を使用していた（18 件中 16 件, 88.9%）。STC はあまり使用されておらず，わずか 3 件のみであった（16.7%）。TA383(75)，TA492(65)では MAIC と STC の両方が使用され，結果の比較が行われたが，いずれの場合も同様の結果であったと報告されている。TA410(72)では MAIC も STC も使用されていない。本 TA では以前の評価(81)のために開発された予測モデルを用いて，AgD の試験の生存曲線を IPD の試験の母集団に合わせて調整した。MAIC を実施した 16 件のうち，有効サンプルサイズが報告されているのはわずか 9 件（56.3%）であった。有効サンプルサイズの中央値は 80.0（範囲：4.0～335.5，四分位範囲：15.4～52.0）であり，有効サンプルサイズの中央値は元のサンプルサイズの 74.2%（範囲：7.9%～94.1%，四分位範囲：48.0%～84.6%）であった。有効サンプルサイズの大きな減少は，多くの場合で IPD の試験と AgD の試験間のオーバーラップが不十分であることを示している。多くの TA が小さい有効サンプルサイズを報告しており，結果として得られた比較は IPD 試験の少数の個体に依存してしまい不安定である可能性がある。

3.2.6　アンカーを有する間接比較とアンカーを有さない間接比較

アンカーを有する間接比較を行った TA は 18 件中 TA383，TA449 の 2 件（11.1%）にとどまった。残りの 16 件の TA（88.9%）は共通の比較対照を持たず，アンカーを有さない間接比較を行っており，正当化が非常に困難な強い仮定に依存しているため，未知の残留バイアスの影響を受けている可能性が高い。残留バイアスを定量化しようとした TA はなかったが，これは非常に困難である（TSD18 参照）。そのため，アプレイザル委員会やエビデンスレビューグループは，アンカーを有さない間接比較での推定値を非常に注意深く扱っている。

3.2.7　調整された共変量

アンカーを有さない間接比較を使用した TA では，調整された共変量数の中央値は 6 個で，1～13 個の範囲であった。アンカーを有する比較を報告した 2 件の TA のうち，共変量の選択に関する情報を提示したのは TA383(75)の 1 件のみであり，10 個の共変量が調整された。

がん領域の TA でよく用いられる共変量は，年齢，Eastern Cooperative Oncology Group（ECOG）のパフォーマンスステータス，性別，前治療の回数及び／または種類であった。多くの TA ではバイオマーカーレベルや疾患サブタイプなどの他の臨床因子も調整されていた。

肝臓病領域の TA364(77)，TA331(78)では，年齢，BMI（body mass index），性別，線維化ステージ，ウイルス量が調整された。TA364(77)では，遺伝子型と対照治療法が異なる 2 種類の MAIC の分析で人種，遺伝子型，及びいくつかのバイオマーカーレベルをさらに調整したが，3 番目の MAIC の分析では，ウイルス量を調整するのに十分なサンプルサイズが得られなかった。

リウマチ領域の唯一の TA である TA383(75)では，年齢，性別，人種，併用治療，2 つのバイオマーカー，3 つの機能／活動性スコアを含む 10 の共変量が調整された。

アンカーを有さない間接比較を使用した TA の中で共変量の選択を正当化する最も一般的な理由は，研究で報告された全てのベースライン特性を単純に調整することであった。アンカーを有する間接比較では，効果修飾因子である共変量のみ調整が必要だが，アンカーを有する間接比較を使用した TA383(75)でも同様であった。アンカーを有する間接比較を用いたもう 1 件（TA449(69)）では，変数選択に関する情報は報告されていない。不必要な調整がバイアスを導くことはない（TSD18 2.3.1.1 節参照）が，特に MAIC の場合には不確実性を高める可能性がある。2 つの評価（TA429(72)，TA457(68)）では，専門家の臨床的意見を用いて共変量の選択が正当化された。MAIC を用いた TA510(63)では専門家に共変量を重要度別にランク付けしてもらい，重要度の高い順に共変量を 1 つずつモデルに追加し，最終的なモデルの選択は有効サンプルサイズを考慮して決定された。特にアンカーを有さない MAIC では，潜在的な予後因子の数が多いため，有効サンプルサイズと調整変数の数はトレードオフの関係となる。しかし，全ての予後因子と効果修飾因子が調整に含まれない限り，推定値は偏ったままである。さらに，有効サンプルサイズの減少が最も大きい共変量は，母集団間で最も不均衡な共変量であり，予後因子または効果修飾の強さが類似している共変量の中でも優先して調整することが重要である。STC を使用した 2 つの

TA では，共変量を選択するために統計的手法を使用されていた。TA333(79)は，回帰モデルにおいて「有意な」共変量を選択したが，これもまた残留バイアスが発生する可能性が高い。特にサンプルサイズが少ない場合にはその可能性はさらに高まる(82)。TA492(65)では，交差検証された予測性能を最大化するような共変量が選択された。これは，STC が AgD 全体の正確な予測に依存していることを考えると，「有意性」に基づく選択よりも適切であるが，それでも交差検証の限界は存在する（TSD18）。

3.2.8　より大きな治療ネットワーク

MAIC と STC は関心のある複数の比較対照及び／または複数の AgD を持つより大きなネットワーク構造に拡張することは想定されていない。しかし，実際にはこれらのシナリオが頻繁に発生しており，18 件の TA のうち 10 件（55.6%）が，より大きな治療／試験のネットワークを含んでいた。

これらのうち 5 件（71.4％；TA331(77)，TA383(75)，TA429(72)，TA500(64)，TA510(63)）では，複数の母集団を調整した間接比較が実施された後，対象集団の結果のみ残されている。これらの推定値はそれぞれ異なる対象集団に対して有効であるため，（アンカーを有する間接比較では効果修飾因子の点で，アンカーを有さない間接比較では予後因子の点でも）全ての対象集団が実際に同一であるという追加の仮定が満たされない限り，わかりやすく解釈することはできない。

TA492(65)では，STC（及び MAIC を感度分析として使用）を用いて，無接続ネットワーク内の各単群試験の有効な治療群を予測し，NMA を用いて新たに接続されたネットワークを分析した。これにより理解しやすい相対治療効果の推定値のセットが得られた。しかし，アンカーを有さない間接比較に必要な非常に強力な仮定もさることながら，この分析では NMA に含まれる単群試験間で効果修飾因子に不均衡がないことも仮定しなければならない。もう一つの重大な懸念は，予測された有効な治療群が繰り返し使用されていることであり，これらは全て同じデータセットに基づいているため独立していない。

TA311(80)，TA380(76)は，主に関心のある 2 つの治療法を含むより広い治療法と試験のネットワークを持っていたが，完全には接続されていなかった。これらのネットワークは，ネットワークを接続するために，2 つの治療法（TA311(80)）とマッチドペア分析（TA380(76)）で同等性の仮定を用いて母集団調整を行わない NMA を使用して分析された。感度分析として母集団調整済みの比較を実施するために，アンカーを有さない MAIC が別途使用された。

TA427(73)では，追加の単群 IPD があり，それを用いて追加の単群比較が行われた（この場合，生存時間アウトカムに対しコックス回帰を用いた）。NICE の TA における IPD を用いた単群比較に関するガイダンスは，Faria ら（TSD17(83)），日本製薬工業協会の報告書(6)によって発表されている。

最後に，TA364(77)では，複数の比較対照が存在し，中には複数の AgD の試験が利用可能なものもあったため，分析方法が不明瞭であった。しかし，アンカーを有さない MAIC が使用されたことを考えると，推定値が単独の推定値として残されているか，ネットワークとして合成されているかにこの分析は依存するため，上述と同様の問題に陥りやすい。

4. ディスカッション

4 章では，意思決定者により良いエビデンスを提供し，分析者がより良い手法の選択ができるようになることを目指し，現状存在するいくつかの重要な改善点を示す。これらは Phillippo ら(5,39)により論点が整理されたものであるが，MAIC や STC の適用は慎重に検討すべきであるという本タスクフォースの考えと一致している。

3 章での 2 種類のレビューにおいて，母集団調整法が公表文献や NICE TA において，徐々に普及してきていることを示した。他の償還機関への提出では，分析の受け取り方や解釈が異なる場合があると考えられる。いずれのレビューも，MAIC と STC が発表された後の限られた期間に限定されている。TSD18 が 2016 年後半に公表されてからは，母集団調整法を使用したその後の全ての TA で TSD18 が引用されており，より適切な運用がなされていくことが期待されている。一方，アプレイザル委員会やエビデンスレビューグループ（ERG）は，母集団調整法に関する結論を通知し，正当化するために TSD18 を使用している。

NICE TA では，意思決定は臨床効果のみに基づくものではなく，費用効果分析では費用も考慮される。従って，母集団調整法が評価に与える影響は費用にも及ぶ。NICE 内の議論(39)を通じて，アプレイザル委員会が MAIC や STC で作成されたエビデンスの質に懸念を抱くことが多いことが明らかとなってきた。分析に使用されたデータの多くは脆弱（例えばフォローアップデータが不完全な場合や小規模な単群試験など）であり，調整された共変量（どの共変量がどのように選択されたか，観察されていない共変量がバイアスをもたらしたかどうか，それはどの程度か）についても不確実性が高い状況であった。その結果，アプレイザル委員会は多くの場合，質が低く不確実なエビデンスによるリスクを相殺するために，肯定的な推奨を行う前に，より大きな費用対効果を示すことを求めていた。母集団調整法による分析の結果が追加の検証分析と一緒に提示された際に，その方法の不確実性が認められその程度が検討され示される（例えば感度分析の使用）と，アプレイザル委員会は意思決定に母集団調整法による分析結果を参照する傾向があった。

公表済みの文献と NICE TA のいずれにおいても，多くの分析ではアンカーを有さない場合が多く，非常に強い仮定に依存している。アンカーを有さない分析の増加の要因は，米国食品医薬品局（Food and Drug Administration：FDA）や欧州医薬品庁（European Medicines Agency：EMA）などの規制当局による迅速承認または条件付き承認の増加に伴う単群試験の増加が大部分を占めていることから，今後も拡大する可能性が高い(84)。しかし，臨床的有効性を証明するための（許認可を得るための）エビデンス要件は，費用対効果を証明するための（償還を得るための）エビデンス要件と異なる場合もある。NICE のアプレイザル委員会やエビデンスレビューグループは，このエビデンスギャップを埋めるためにアンカーを有さない母集団調整法を使用することを警戒しており，その結果は慎重に解釈すべきであり，未知のバイアスが含まれている可能性があるとの意見を度々述べている。今後，規制当局と償還機関間の対話を増やすことは，このエビデンス要件のギャップを埋めるのに役立つかもしれないと考える。

現在用いられている母集団調整法は，アンカーを有する間接比較を行う際には，測定されていない効果修飾因子が存在しないことを前提としている。アンカーを有さない間接比較では，さらに測定されて

いない予後因子も存在しないことが前提とされている。この後者の仮定は特に強く正当化が難しい。測定されていない予後因子と効果修飾因子による残留バイアスの定量化は今後の研究課題である。

文献及び NICE TA におけるいくつかの適用例では，複数の比較対照及び／または AgD が存在した（3.1.4 節，3.1.8 節，3.2.8 節参照）。現行の MAIC 及び STC の方法論では，より大きなネットワーク構造を扱うことができないため，各事例では複数の分析が実施され，その後，独立した比較として残されるか，あるいは NMA を用いてそれ自体が合成される。後者の場合は更なる仮定が必要となる。さらに，現在の MAIC と STC は，追加の仮定なしに AgD（一般には競合他社の試験データ）に対してのみ有効な推定値を生成するため，意思決定対象集団と一致しない可能性が高い（TSD18）。TA451(70)で実施された MAIC では，意思決定対象集団に近いと考えられる IPD の試験の結果を，意思決定対象集団とはかけ離れた集団である AgD の方に調整していることが指摘されているが，このような対応がなされているということは，これまでの NICE TA や公表された文献ではほとんど見落とされてきた。明らかに効果修飾因子が存在する場合に，単に「不偏」な推定値を作成するだけでは十分でない。作成された推定値は意思決定のための集団に固有のものでなければならない。このことは，より大きな治療ネットワークにも自然に拡張でき，与えられたターゲットの意思決定対象集団の推定値を作成できる新しい手法の開発に動機づけを与えている。さらに，全ての試験が 1 つ以上の効果修飾因子に関して意思決定対象集団の部分集団である場合には，調整は外挿に頼らなければならず，現在の MAIC と STC ではこれらの効果修飾因子が離散的である場合には，調整は不可能である。

公表された分析や NICE TA の大部分では, 母集団調整された間接比較のために MAIC が使用された。有効サンプルサイズは小さくなる傾向があり，多くは元のサンプルサイズに比べて大幅に小さくなってしまっている。これは IPD と AgD の間にオーバーラップがない可能性があることを示している。オーバーラップの欠如は，MAIC のような再重み付け法では特に懸念される。それは，IPD で観察された値の範囲外の共変量を考慮して外挿することができないからである。このことについては，集団間のオーバーラップがあまりないシナリオでの MAIC（及び他の母集団調整法）のロバスト性をシミュレーション研究で検討している（2.1.1 節）。シミュレーション研究によると，オーバーラップが 50%程度のときには，STC にはなかったバイアスが MAIC では残ってしまい，状況によっては未調整の標準的な間接比較よりもバイアスを生じさせてしまう可能性もある。

含まれる全ての試験で IPD が利用可能であった 3 件の TA がレビューから除外された(84–86)。これらの評価は全て, がん領域における生存アウトカムでのアンカーを有さない間接比較であり, 傾向スコア, 共変量マッチング，回帰分析の中から一部の方法が使用されていた。全ての試験で IPD が利用可能であることがゴールドスタンダードであり，可能であればそれが望ましい。IPD により統計的な検出力が上がり，より厳格ではない仮定に依存する分析が可能となり，仮定の検証が容易となるからである。全ての試験で IPD を用いた分析については，NICE DSU から方法論的ガイダンスが提供されている（TSD17(83)）

TA を含む意思決定の場面において，期待する効果を得るためには，母集団調整法による分析は，現状の実務においていくつかの改善が必要である（TSD18）。まずは，意思決定者に関連するターゲット集団を定義し，この集団に対する推定値を作成しなければならないことである。現在の母集団調整法では，更なる仮定がない限り，AgD で表された集団に対してのみ有効な推定値が作成され，それは手元の意思

決定結果を適切に表していない可能性がある。このことは，これまでの TA と公表された文献の適用例ではほとんど見落とされてきた（ただし 3 章で検索された文献や TA は，TSD18 の発表前のものがほとんどであることに注意が必要である)。アンカーを有する間接比較では，先行研究や臨床の専門知識に基づく経験的エビデンスに基づいて，事前に効果の修正を行うための明確な正当化が必要である。これまでのところ，アンカーを有する母集団調整法の適用例の中で，効果の修飾を正当化しようとする試みを行ったものは 1 報のみであった(61)（3.1.8 節参照)。アンカーを有さない母集団調整法による分析を報告した文献，TA やその他の適用例の中では，そのような正当化の根拠を示したものはなかった。アンカーを有さない間接比較では，予後因子と効果修飾因子の両方を調整することで絶対効果の信頼性の高い予測を必要とし，ランダム化されていないために，観察されない予後因子と効果修飾因子の影響を非常に受けやすい。現在一般的に行われているような，利用可能な全ての共変量を単に調整するだけでは実は十分でない。アンカーを有さない間接比較が影響力を持つためには，共変量は予測性能を念頭に置いて選択されるべきであり，残留バイアスの潜在的な範囲の推定値が必要である。そうでなければ，推定値のバイアスの大きさは不明であり，無調整の比較よりもバイアスが大きくなってしまう可能性もある。これを実現するのは大変困難ではあるが，このような不安な点がない状況でなければ，意思決定者はアンカーを有さない分析の結果を当然慎重に評価することになるであろう。上記の問題の多くは，分析に含まれる全ての試験で IPD が利用可能であることで，少なくとも部分的には緩和される。

おわりに

本報告書では，NICE DSU が公開した TSD18 をもとに，間接比較を行う際に効果修飾因子の分布の違いを調整して治療効果を推定する母集団調整法について，MAIC 及び STC を中心に解説した。また，これらの方法を評価したシミュレーション研究とともに，文献及び NICE TA での適用事例について調査した結果をまとめた。効果修飾因子を無視した単純な間接比較や NMA は妥当な推定結果は得られない可能性があるものの，本報告書で取り上げた手法は，2 つの試験を用いて間接比較を行う際に，一方の試験の IPD さえ利用できれば，妥当な推定結果を得られる可能性があるという点で魅力的な手法である。さらに，アンカーを有さない場合にも適用できるという特徴があることから，適用事例が近年増加している。しかしながら，これらの手法には非常に強い仮定が必要であり，状況によっては未調整の標準的な間接比較よりもバイアスを生じさせてしまう可能性もある。従って，これらの手法を適用する際や結果を解釈する際には，十分に検討すべきであろう。全ての試験で IPD が利用可能であることが望ましいものの，実際には，自社が実施した試験の IPD しか利用できない状況のほうが多いと思われる。その際には本報告書がお役に立てば幸いである。

Appendix：Rによる実装例

TSD18 の付録 D には，アンカーを有する間接比較の場合の MAIC と STC を R で実装するコードとその解説が記載されている。ここではその R コードに日本語で最小限の解説・補足を加えた。詳細な解説については TSD18 をご参照いただきたい。なお，R コードは web 上で公開されている。
http://nicedsu.org.uk/technical-support-documents/population-adjusted-indirect-comparisons-maic-and-stc/

TSD18のMAICの実装例では，標準誤差にサンドイッチ推定量を用いており，専用のsandwichパッケージが使用されている。その他の使用パッケージは，シミュレーションデータの発生，データハンドリング，グラフ作成といった用途であり，母集団調整法専用のパッケージではないことから，同様のロジックを用いればSASなどの他の統計解析ソフトでも実装可能と思われる。TSD18発出以降，MAICに関してはパッケージがいくつか開発されており，以下のリンク先（GitHub，CRAN）で公開されている。

MAIC：https://github.com/Roche/Global-HTA-Evidence-Open/tree/master/Rpackages/MAIC

maic：https://cran.r-project.org/web/packages/maic/index.html https://github.com/heorltd/maic

乱数生成方法の変更

R version 3.6.0以降では乱数生成の標準方法が変更されているため，TSD18の例と同じ結果を得たい場合は，以下を実行する必要がある。

```
RNGkind(sample.kind = "Rounding") # TSD18 の乱数生成方法と同じにする。

# RNGversion(getRversion()) #使用中のR のバージョンの標準方法に戻す
```

必要なRパッケージの準備

```
### Initial setup
if(!require(dplyr)) {install.packages("dplyr"); library(dplyr)}

if(!require(tidyr)) {install.packages("tidyr"); library(tidyr)}

if(!require(wakefield)) {install.packages("wakefield"); library(wakefield)}

if(!require(ggplot2)) {install.packages("ggplot2"); library(ggplot2)}

if(!require(sandwich)) {install.packages("sandwich"); library(sandwich)}
```

解析に用いるデータの準備

2つの2群比較試験，IPDが利用可能なAB試験とAgDのみ利用可能なAC試験について治療Aをアンカーとして間接比較を行う。効果修飾因子として年齢，予後因子として性別を設定する。

アウトカムは2値として，治療t（=A，B，C）を受けている患者iのイベント発生確率をp_{it}とする以下の

アウトカムモデルによって生成する。各パラメータはオッズ比として解釈できる。

$$\text{logit}(p_{it}) = 0.85 + 0.12 * male_{it} + 0.05 * (age_{it} - 40) + \left(\beta_t + -0.08 * (age_{it} - 40)\right) I(t \neq A)$$

AB試験 サンプルサイズ：$N(AB)=500$， 年齢：45-75歳， 女性の割合：64%
AC試験 サンプルサイズ：$N(AC)=300$， 年齢：45-55歳， 女性の割合：80%

上記のパラメータを以下で定義する。

```
set.seed(61374988)

# Study characteristics
N_AB <- 500
N_AC <- 300
agerange_AB <- 45:75
agerange_AC <- 45:55
femalepc_AB <- 0.64
femalepc_AC <- 0.8

# Outcome model
b_0 <- 0.85
b_gender <- 0.12
b_age <- 0.05
b_age_trt <- -0.08
b_trt_B <- -2.1
b_trt_C <- -2.5
```

AB試験のデータを作成する。

```
### Generate AB trial
AB.IPD <-
  rbind(

  # Generate A arm
  r_data_frame(n = N_AB/2,     # Number of individuals in arm A
               id,             # Unique ID
               age = age(x = agerange_AB),   # Generate ages
               gender = gender(prob = c(1 - femalepc_AB, femalepc_AB)), # Generate gender
s
               trt = "A"       # Assign treatment A
               ),
```

```
  # Generate B arm
  r_data_frame(n = N_AB/2,     # Number of individuals in arm B
               id,             # Unique ID
               age = age(x = agerange_AB),   # Generate ages
               gender = gender(prob = c(1 - femalepc_AB, femalepc_AB)), # Generate gender
s
               trt = "B"       # Assign treatment B
               )
  ) %>%

  # Generate outcomes using logistic model
  mutate(
    yprob = 1 / (1 + exp(-(
      b_0 + b_gender * (gender == "Male") + b_age * (age - 40) +
        if_else(trt == "B", b_trt_B + b_age_trt * (age - 40), 0)
    ))),
    y = rbinom(N_AB, 1, yprob)
  ) %>%
  select(-yprob)   # Drop the yprob column
```

AB試験のデータを確認する。

```
# Tabulate
AB.IPD %>% group_by(trt) %>%
  summarise(n(), mean(age), sd(age), `n(male)`=sum(gender=="Male"),
            `%(male)`=mean(gender=="Male"), sum(y), mean(y))

## # A tibble: 2 x 8
##   trt   `n()` `mean(age)` `sd(age)` `n(male)` `%(male)` `sum(y)` `mean(y)`
##   <chr> <int>       <dbl>     <dbl>     <int>     <dbl>    <int>     <dbl>
## 1 A       250        60.0      9.15        87     0.348      217     0.868
## 2 B       250        60.1      9.21        92     0.368       39     0.156
```

同様にAC試験のデータを作成する。

```
### Generate AC trial
AC.IPD <-
  rbind(

  # Generate A arm
```

```
  r_data_frame(n = N_AC/2,     # Number of individuals in arm A
               id,             # Unique ID
               age = age(x = agerange_AC),   # Generate ages
               gender = gender(prob = c(1 - femalepc_AC, femalepc_AC)), # Generate gender
s
               trt = "A"       # Assign treatment A
               ),

  # Generate C arm
  r_data_frame(n = N_AC/2,     # Number of individuals in arm C
               id,             # Unique ID
               age = age(x = agerange_AC),   # Generate ages
               gender = gender(prob = c(1 - femalepc_AC, femalepc_AC)), # Generate gender
s
               trt = "C"       # Assign treatment C
               )
  ) %>%

  # Generate outcomes using logistic model
  mutate(
    yprob = 1 / (1 + exp(-(
      b_0 + b_gender * (gender == "Male") + b_age * (age - 40) +
        if_else(trt == "C", b_trt_C + b_age_trt * (age - 40), 0)
    ))),
    y = rbinom(N_AC, 1, yprob)
  ) %>%
  select(-yprob)   # Drop the yprob column
```

同様にAC試験のデータを確認する。

```
# Tabulate
AC.IPD %>% group_by(trt) %>%
  summarise(n(), mean(age), sd(age), `n(male)`=sum(gender=="Male"),
            `%(male)`=mean(gender=="Male"), sum(y), mean(y))

## # A tibble: 2 x 8
##   trt   `n()` `mean(age)` `sd(age)` `n(male)` `%(male)` `sum(y)` `mean(y)`
##   <chr> <int>       <dbl>     <dbl>     <int>     <dbl>    <int>     <dbl>
## 1 A       150        49.5      2.96        33      0.22      115     0.767
## 2 C       150        50.1      3.19        27      0.18       17     0.113
```

最後にAC試験のAgDを作成する。

```
# Create aggregate data
AC.AgD <-
  cbind(
    # Trial level stats: mean and sd of age, number and proportion of males
    summarise(AC.IPD, age.mean = mean(age), age.sd = sd(age),
              N.male = sum(gender=="Male"), prop.male = mean(gender=="Male")),

    # Summary outcomes for A arm
    filter(AC.IPD, trt == "A") %>%
      summarise(y.A.sum = sum(y), y.A.bar = mean(y), N.A = n()),

    # Summary outcomes for C arm
    filter(AC.IPD, trt == "C") %>%
      summarise(y.C.sum = sum(y), y.C.bar = mean(y), N.C = n())
  )

AC.AgD

##   age.mean   age.sd N.male prop.male y.A.sum   y.A.bar N.A y.C.sum   y.C.bar
## 1 49.80667 3.082363     60       0.2     115 0.7666667 150      17 0.1133333
##   N.C
## 1 150
```

以上でデータの準備が終わり，実際にMAICとSTCを行うためのコードは以降となる。

MAICによる間接比較

全ての効果修飾因子を持つロジスティックモデルによって傾向スコアを推定する。個々の重みは $\log(w_{it}) = \alpha_0 + \alpha_1^T \mathbf{X}_{it}^{EM}$ に相当し，AB試験とAC試験の間の効果修飾因子の分布を一致させるモーメント法を使って推定できる。これは$\bar{X}_{(AC)}^{EM} = 0$のときに，$\sum_{i,t} \exp(\alpha_1^T \mathbf{X}_{it}^{EM})$を最小化することに相当し，この最小化する目的関数と最小化アルゴリズムによって利用する勾配関数（その微分）を以下で定義する。

```
### MAIC
# Objective function
objfn <- function(a1, X){
  sum(exp(X %*% a1))
}

# Gradient function
```

```
gradfn <- function(a1, X){
  colSums(sweep(X, 1, exp(X %*% a1), "*"))
}
```

次に，$\bar{X}_{(AC)}^{EM} = 0$を満たすために，両方の試験で，X^{EM}から$\bar{X}_{(AC)}^{EM}$を差し引くことにより，効果修飾因子を中心化する。ここでは，年齢が効果修飾因子であり，平均と標準偏差の両方でバランスを取ることを考える。AC試験ではage.meanとage.sdとして観測されている。そのため，重み付きモデルでのAB試験の個々のデータに対する中心化としてageとage^2を含める。平均を中心化する際には，var(**X**)＝E(**X**^{2}) − E(**X**)^{2}より，AC試験のage.mean^2＋age.sd^2を用いて，AB試験のage^2を中心化する。

```
# Centred EMs
X.EM.0 <- sweep(with(AB.IPD, cbind(age, age^2)), 2,
                with(AC.AgD, c(age.mean, age.mean^2 + age.sd^2)), '-')
```

α_1を推定するために，optim()を用いて準ニュートン法に属するBFGS法によって先ほど定義したobjfn()を最小化する。

```
# Estimate weights
print(opt1 <- optim(par = c(0,0), fn = objfn, gr = gradfn, X = X.EM.0, method = "BFGS"))

## $par
## [1]  3.15553663 -0.03224845
##
## $value
## [1] 202.1651
##
## $counts
## function gradient
##       67       14
##
## $convergence
## [1] 0
##
## $message
## NULL

a1 <- opt1$par
```

$convergence=0から収束したことがわかり，その推定値$\hat{\alpha}_1$は$parにある。

次に，各個人での重みの推定量$\widehat{w_{ij}} = \exp\left(\bar{X}_{(AC)}^{EM}\hat{\alpha}_1\right)$を算出する。なお，$\alpha_0$は定数で相殺されるため，推定する必要はない。さらに重みを再スケール化し平均を1にする。

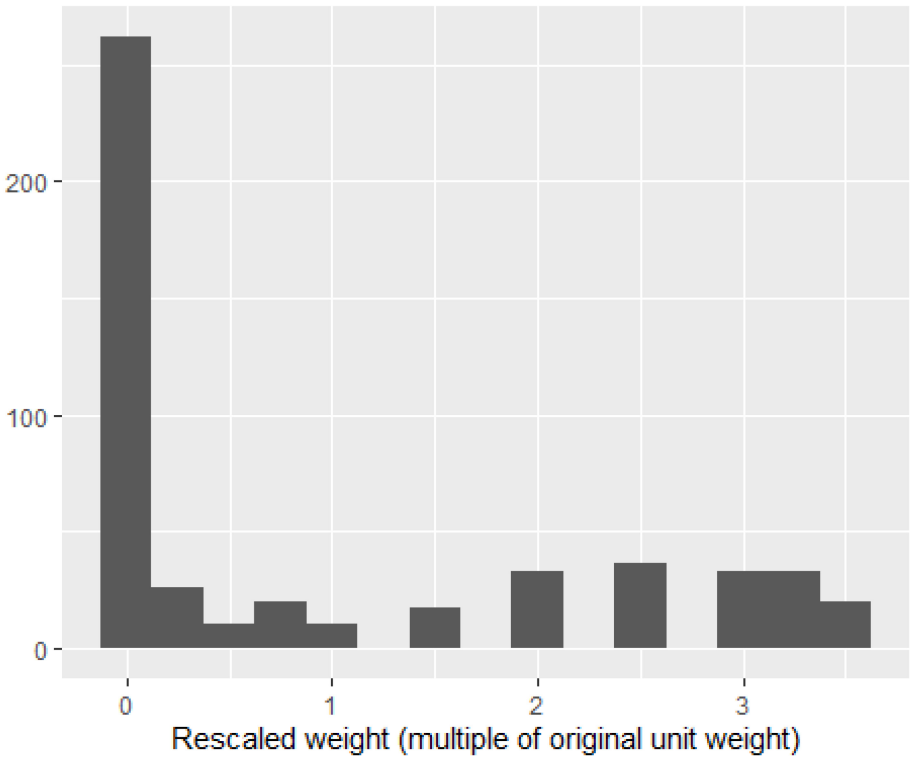
```
wt <- exp(X.EM.0 %*% a1)
wt.rs <- (wt / sum(wt)) * N_AB   # rescaled weights
```

再スケール化された重みの分布を確認する。

```
# Summary of weights, histogram
summary(wt.rs)

##       V1
##  Min.   :0.000000
##  1st Qu.:0.000023
##  Median :0.065993
##  Mean   :1.000000
##  3rd Qu.:2.096045
##  Max.   :3.444506

qplot(wt.rs, geom="histogram",
     xlab = "Rescaled weight (multiple of original unit weight)",
     binwidth=0.25)
```



AB 試験の被験者の再スケール化された重みの分布をみると，非常に歪んでおり，多くの被験者が 0 に近い重みが与えられていることがわかる。これは AB 試験での年齢範囲（45-75 歳）が AC 試験（45-55 歳）より広いからである。このときの ESS は次のように計算され，AB 試験の例数である 250/群から

185.6451 に減少していることがわかるが，この数値は十分に大きいと TSD18 では述べられている。

```
# Effective sample size
sum(wt)^2/sum(wt^2)

## [1] 185.6451
```

重み付けした後のAB集団とAC集団のバランスを確認する。

```
# Check balance
AB.IPD %>%
  mutate(wt) %>%
  summarise(age.mean = weighted.mean(age, wt),
            age.sd = sqrt(sum(wt / sum(wt) * (age - age.mean)^2))
            )

## # A tibble: 1 x 2
##   age.mean age.sd
##      <dbl>  <dbl>
## 1     49.8   3.08

AC.AgD[, c("age.mean", "age.sd")]

##   age.mean   age.sd
## 1 49.80667 3.082363

## Create the weighting estimator using a simple linear model, use sandwich
## estimator for standard error.
```

年齢の平均とSDのバランスが取れていることがわかる。

AC集団における相対治療効果の推定値$\hat{d}_{AB(AC)}$は，AB試験におけるアウトカムの重み付き平均を取ることで得られる。なお，サンドイッチ推定量を求める関数sandwich::vcovHC()には，一般化線形モデルのglm()（もしくは一般線形モデルlm()）の結果が必要であるため，以下ではglm()による線形モデルを当てはめて，治療Bの効果も算出する。

```
# Binomial GLM
fit1 <-
  AB.IPD %>% mutate(y0 = 1 - y, wt = wt) %>%
  glm(cbind(y,y0) ~ trt, data = ., family = binomial, weights = wt)

# Sandwich estimator of variance matrix
V.sw <- vcovHC(fit1)
```

```
# The log OR of B vs. A is just the trtB parameter estimate,
# since effect modifiers were centred
print(d.AB.MAIC <- coef(fit1)["trtB"])

##     trtB
## -3.215136

print(var.d.AB.MAIC <- V.sw["trtB","trtB"])

## [1] 0.1628077
```

最後に，$d_{BC(AC)} = d_{AC\,(AC)} - d_{AB(AC)}$より，間接比較の推定値 $\hat{d}_{BC(AC)}$ を求める。

```
# Estimated log OR of C vs. A from the AC trial
d.AC <- with(AC.AgD, log(y.C.sum * (N.A - y.A.sum) / (y.A.sum * (N.C - y.C.sum))))
var.d.AC <- with(AC.AgD, 1/y.A.sum + 1/(N.A - y.A.sum) + 1/y.C.sum + 1/(N.C - y.C.sum))

# Indirect comparison of C vs. B in AC trial
print(d.BC.MAIC <- d.AC - d.AB.MAIC)

##       trtB
## -0.03158391

print(var.d.BC.MAIC <- var.d.AC + var.d.AB.MAIC)

## [1] 0.2664171
```

MAICによる治療Bに対する治療Cの対数オッズ比の推定値は-0.032と求まる。SEは print(sqrt(var.d.BC.MAIC))を実行すると0.516と求まる。

STCによる間接比較

アウトカム回帰モデルをAB集団に当てはめ，これを利用してAC集団の結果を予測する。アウトカムモデルでバイアスを防ぐためには，不均衡のある全ての効果修飾因子を含むことが必要だが，さらにその他の共変量を追加することで精度を向上できる。今回の場合，モデルに効果修飾因子である年齢を含める必要があるが，効果修飾因子ではない性別を追加することで，精度を向上できる可能性があるためそれも試みた。

まず，アウトカムモデルを当てはめる。AC集団の平均値で年齢を中心化させている点に注意していただきたい。

```
### STC

AB.IPD$y0 <- 1 - AB.IPD$y  # Add in dummy non-event column
```

```r
# Fit binomial GLM
STC.GLM <- glm(cbind(y,y0) ~ trt*I(age - AC.AgD$age.mean),
               data = AB.IPD, family = binomial)
summary(STC.GLM)

##
## Call:
## glm(formula = cbind(y, y0) ~ trt * I(age - AC.AgD$age.mean),
##     family = binomial, data = AB.IPD)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2617  -0.5932   0.4016   0.5426   2.0739
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   1.54856    0.25205   6.144 8.05e-10 ***
## trtB                         -2.99962    0.35089  -8.549  < 2e-16 ***
## I(age - AC.AgD$age.mean)      0.03685    0.02103   1.752   0.0797 .
## trtB:I(age - AC.AgD$age.mean) -0.06167    0.02851  -2.163   0.0305 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 692.86  on 499  degrees of freedom
## Residual deviance: 406.73  on 496  degrees of freedom
## AIC: 414.73
##
## Number of Fisher Scoring iterations: 4
```

この例では，間接比較のSEを小さくするために，性別をモデルに追加することを試みたものの，以下の結果のとおり，改善しなかったため，性別は採用しなかった。なお，推定された係数の全てが，真のロジスティックモデルに一致するわけではないことに注意が必要である。その理由は，ここでの基準年齢がAC集団の平均年齢であるのに対して，真のモデルの基準年齢は40歳であるという違いがあるためである。

```r
# Try adding prognostic variables to improve model fit
add1(STC.GLM, ~.+gender, test="Chisq")
```

```
## Single term additions
##
## Model:
## cbind(y, y0) ~ trt * I(age - AC.AgD$age.mean)
##        Df Deviance    AIC     LRT Pr(>Chi)
## <none>      406.73 414.73
## gender  1   406.33 416.33 0.40192   0.5261
```

作成したモデルを用いて，AC集団における治療Aに対する治療Bの対数オッズ比$d_{AB(AC)}$を推定する。

```
# Estimated log OR of B vs. A in the AC trial
print(d.AB.STC <- coef(STC.GLM)["trtB"])

##    trtB
## -2.99962

print(var.d.AB.STC <- vcov(STC.GLM)["trtB","trtB"])

## [1] 0.1231246

# Indirect comparison of C vs. B in AC trial
print(d.BC.STC <- d.AC - d.AB.STC)

##      trtB
## -0.2470995

print(var.d.BC.STC <- var.d.AC + var.d.AB.STC)

## [1] 0.226734
```

STCによる治療Bに対する治療Cの対数オッズ比の推定値は-0.247が得られた。

結果のまとめ

AC集団における真の対数オッズ比の算出と未調整の標準的な間接比較を行い，MAICとSTCの推定結果をフォレストプロットで示す。

```
### Summary
# True B vs. A effect in the AC population
d.AB.TRUE <- b_trt_B + b_age_trt * (AC.AgD$age.mean - 40)

# Naive approach
AB.IPD %>% group_by(trt) %>%
summarise(y.sum = sum(y)) %>%
spread(trt, y.sum) %>%
  with({
```

```
  d.AB.AB <<- log(B * (N_AB/2 - A) / (A * (N_AB/2 - B)))
  var.d.AB.AB <<- 1/B + 1/(N_AB/2 - A) + 1/A + 1/(N_AB/2 - B)
})
```

同様に，AC集団における治療Bに対する治療Cの治療効果の真値の算出と標準的な間接比較を行う。

```
# True C vs. B effect in AC trial
d.BC.TRUE <- b_trt_C - b_trt_B

# Naive approach
d.BC.NAIVE <- d.AC - d.AB.AB
var.d.BC.NAIVE <- var.d.AC + var.d.AB.AB
```

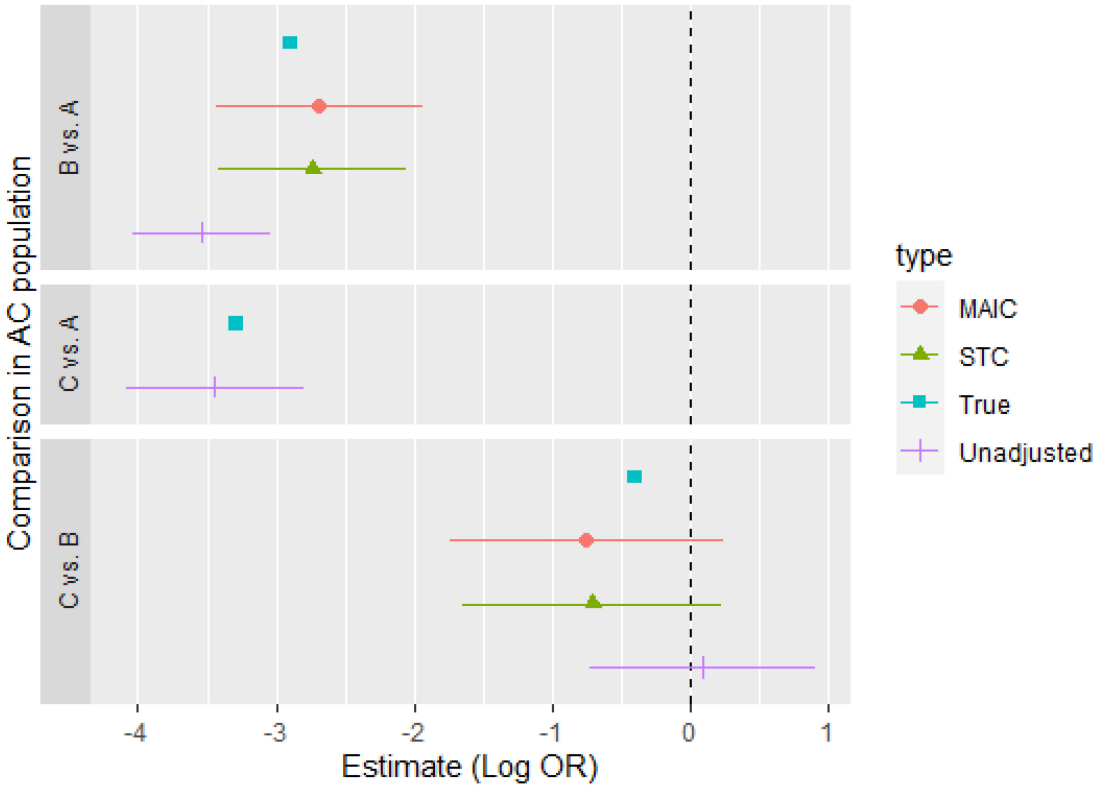
最後にフォレストプロットを作成する。なお，ここでのwarningメッセージはパッケージのバージョンによる注意書きであり，結果には影響しないため，必要に応じて変更して実行いただきたい。

```
# Forest plot of results
plotdat <- data_frame(
  id = 1:10,
  Comparison = factor(c(rep(1,4), 2, 2, rep(3,4)),
                      labels = c("B vs. A", "C vs. A", "C vs. B")),
  Estimate = c(d.AB.TRUE, d.AB.MAIC, d.AB.STC, d.AB.AB,
               b_trt_C + b_age_trt * (AC.AgD$age.mean - 40), d.AC,
               d.BC.TRUE, d.BC.MAIC, d.BC.STC, d.BC.NAIVE),
  var = c(NA, var.d.AB.MAIC, var.d.AB.STC, var.d.AB.AB,
          NA, var.d.AC,
          NA, var.d.BC.MAIC, var.d.BC.STC, var.d.BC.NAIVE),
  lo = Estimate + qnorm(0.025) * sqrt(var),
  hi = Estimate + qnorm(0.975) * sqrt(var),
  type = c("True", "MAIC", "STC", "Unadjusted",
           "True","Unadjusted",
           "True", "MAIC", "STC", "Unadjusted")
)

## Warning: `data_frame()` was deprecated in tibble 1.1.0.
## Please use `tibble()` instead.

ggplot(aes(x = Estimate, y = id, col = type, shape = type), data = plotdat) +
  geom_vline(xintercept = 0, lty = 2) +
  geom_point(size = 2) +
  geom_segment(aes(y = id, yend = id, x = lo, xend = hi), na.rm = TRUE) +
```

```
  xlab("Estimate (Log OR)") +
  facet_grid(Comparison~., switch = "y", scales = "free_y", space = "free_y") +
  scale_y_reverse(name = "Comparison in AC population", breaks = NULL, expand = c(0, 0.
6))
```



フォレストプロットから，MAIC と STC による推定値は標準的な間接比較よりもバイアスが小さく，信頼区間が真値を含んでいることがわかる。また，未調整の標準的な間接比較（Unadjusted）にはバイアスがあることが視覚的にもわかる。仮に未調整の推定値しか得られなかった場合は，誤って治療 B を支持してしまうことになる。今回のシナリオでは，STC が真値に近く，信頼区間も MAIC より狭く見えるが，この結果だけではどちらの方法が望ましいかまではいえないため，シミュレーション研究による性能評価などが TSD18 で推奨されている。

参考文献

1. 医薬品，医療機器及び再生医療等製品の費用対効果評価に関する取扱いについて [Internet]. 2019 [cited 2021 Apr 24]. Available from: https://c2h.niph.go.jp/tools/system/notification_cost-effectiveness.pdf
2. National Institute for Health and Care Excellence. [cited 2021 Apr 24]; Available from: https://www.nice.org.uk/
3. NICE Decision Support Unit. [Internet]. [cited 2021 Apr 24]. Available from: http://nicedsu.org.uk/
4. 日本製薬工業協会. 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 [cited 2021 Dec 1]; Available from: https://www.jpma.or.jp/information/evaluation/results/allotment/nice_tsd_evidence_synthesis.html
5. 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.
6. 日本製薬工業協会. 非ランダム化比較試験データを用いて治療効果を推定するための統計的手法. 2020 [cited 2021 Dec 1]; Available from: https://www.jpma.or.jp/information/evaluation/results/allotment/non-randomization.html
7. 日本製薬工業協会. ICH-E9(R1)の理解に役立つ因果推論シンポジウム. 2021 [cited 2021 Dec 1]; Available from: https://www.jpma.or.jp/information/evaluation/symposium/2021_03_19.html
8. Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, et al. Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. Value in Health. 2012;15(6).
9. Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, et al. Comparative effectiveness without head-to-head trials: A method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. PharmacoEconomics. 2010;28(10).
10. Ishak KJ, Proskorovsky I, Benedict A. Simulation and Matching-Based Approaches for Indirect Comparison of Treatments. PharmacoEconomics. 2015;33(6).
11. Caro JJ, Ishak KJ. No head-to-head trial? Simulate the missing arms. PharmacoEconomics. 2010;28(10).
12. Jansen JP. Network meta-analysis of individual and aggregate level data. Research Synthesis Methods. 2012;3(2).
13. Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. Statistics in Medicine. 2012;31(28).
14. Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. Statistics in Medicine. 2013;32(6).
15. Thom HH, Capkun G, Cerulli A, Nixon RM, Howard LS. Network meta-analysis combining individual patient and aggregate data from a mixture of study designs with an application to pulmonary arterial hypertension. BMC Medical Research Methodology. 2015;15(1).
16. Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis. 2012;20(1).
17. Hartman E, Grieve R, Ramsahai R, Sekhon JS. From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2015;178(3).
18. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of

results from randomized trials. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2011;174(2).

19. Zhang Z, Nie L, Soon G, Hu Z. New methods for treatment effect calibration, with applications to non-inferiority trials. Biometrics. 2016;72(1).
20. Nie L, Zhang Z, Rubin D, Chu J. Likelihood reweighting methods to reduce potential bias in noninferiority trials which rely on historical data to make inference. The Annals of Applied Statistics. 2013;7(3).
21. Vartivarian S LRJ. Does weighting for nonresponse increase the variance of survey means? Statistics Canada, Survey Methodology. 2005;31(2).
22. Sutton AJ, Kendrick D, Coupland CAC. Meta-analysis of individual- and aggregate-level data. Statistics in Medicine. 2008;27(5).
23. Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data. Statistics in Medicine. 2006;25(12).
24. Jackson C, Best and N, 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). 2007;171(1).
25. Phillippo DM, Dias S, Ades AE, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. Statistics in Medicine. 2020;39(30).
26. Belger M, Brnabic A, Kadziola Z, Petto H, Faries D. Inclusion Of Multiple Studies In Matching Adjusted Indirect Comparisons (Maic). Value in Health. 2015;18(3).
27. Belger M, Brnabic A, Kadziola Z, Petto H, Faries D. Alternative Weighting Approaches For Matching Adjusted Indirect Comparisons (Maic). Value in Health. 2015;18(3).
28. Norton EC, Miller MM, Wang JJ, Coyne K, Kleinman LC. Rank Reversal in Indirect Comparisons. Value in Health. 2012;15(8).
29. van Valkenhoef G, Ades AE. Evidence Synthesis Assumes Additivity on the Scale of Measurement: Response to "Rank Reversal in Indirect Comparisons" by Norton et al. Value in Health. 2013;16(2).
30. Caldwell DM, Welton NJ, Dias S, Ades A. Selecting the best scale for measuring treatment effect in a network meta-analysis: a case study in childhood nocturnal enuresis. Research Synthesis Methods. 2012;3(2).
31. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence. 2013.
32. Hoaglin DC. An indirect comparison of everolimus versus sorafenib in metastatic renal cell carcinoma-a flawed analysis and a problematic response. Expert Opinion on Pharmacotherapy. 2013;14(12).
33. Hatswell A, Freemantle N, Baio G. PRM239 - DOES MATCHING ADJUSTED INDIRECT COMPARISON (MAIC) WORK? RESULTS FROM A SIMULATION STUDY. Value in Health. 2018;21.
34. Hatswell AJ, Freemantle N, Baio G. The Effects of Model Misspecification in Unanchored Matching-Adjusted Indirect Comparison: Results of a Simulation Study. Value in Health. 2020;23(6).
35. Leahy J. The Impact of Performing a Network Meta-Analysis with Imperfect Evidence. PhD thesis Trinity College Dublin [Internet]. 2019 [cited 2021 Feb 19]; Available from: http://hdl.handle.net/2262/86070
36. Signorovitch J, Ayyagari R, Cheng D, Wu EQ. Matching-adjusted indirect comparisons: A simulation study of statistical performance. Value in Health. 2013;16(3).

37. Jiang Y, Ni W. Performance of unanchored matching-adjusted indirect comparison (MAIC) for the evidence synthesis of single-arm trials with time-to-event outcomes. BMC Medical Research Methodology. 2020;20(1).
38. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of Clinical Epidemiology. 1997;50(6).
39. Phillippo DM, Dias S, Elsada 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).
40. Veroniki AA, Straus SE, Soobiah C, Elliott MJ, Tricco AC. A scoping review of indirect comparison methods and applications using individual patient data. BMC Medical Research Methodology. 2016;16(1).
41. Chang GC, Ahn MJ, Wright E, Kim HT, Kim JH, Kang JH, et al. Comparative effectiveness of bevacizumab plus cisplatin-based chemotherapy versus pemetrexed plus cisplatin treatment in East Asian non-squamous non-small cell lung cancer patients applying real-life outcomes. Asia-Pacific Journal of Clinical Oncology. 2011;7(SUPPL.2).
42. Signorovitch JE, Wu EQ, Betts KA, Parikh K, Kantor E, Guo A, et al. Comparative efficacy of nilotinib and dasatinib in newly diagnosed chronic myeloid leukemia: A matching-adjusted indirect comparison of randomized trials. Current Medical Research and Opinion. 2011;27(6).
43. Signorovitch JE, Wu EQ, Swallow E, Kantor E, Fan L, Gruenberger JB. Comparative efficacy of vildagliptin and sitagliptin in japanese patients with type 2 diabetes mellitus: A matching-adjusted indirect comparison of randomized trials. Clinical Drug Investigation. 2011;31(9).
44. Kirson NY, Rao S, Birnbaum HG, Kantor E, Wei RS, Cifaldi M. Matching-adjusted indirect comparison of adalimumab vs etanercept and infliximab for the treatment of psoriatic arthritis. Journal of Medical Economics. 2013;16(4).
45. Signorovitch J, Swallow E, Kantor E, Wang X, Klimovsky J, Haas T, et al. Everolimus and sunitinib for advanced pancreatic neuroendocrine tumors: A matching-adjusted indirect comparison. Experimental Hematology and Oncology. 2013;2(1).
46. Sikirica V, Findling RL, Signorovitch J, Erder MH, Dammerman R, Hodgkins P, et al. Comparative efficacy of guanfacine extended release versus atomoxetine for the treatment of attention-deficit/hyperactivity disorder in children and adolescents: Applying matching-adjusted indirect comparison methodology. CNS Drugs. 2013;27(11).
47. Sherman S, Amzal B, Calvo E, Wang X, Park J, Liu Z, et al. An Indirect Comparison of Everolimus Versus Axitinib in US Patients with Advanced Renal Cell Carcinoma in Whom Prior Sunitinib Therapy Failed. Clinical Therapeutics. 2015;37(11).
48. van Sanden S, Pisini M, Duchesne I, Mehnert A, Belsey J. Indirect comparison of the antiviral efficacy of peginterferon alpha 2a plus ribavirin used with or without simeprevir in genotype 4 hepatitis C virus infection, where common comparator study arms are lacking: A special application of the matching adjusted indirect comparison methodology. Current Medical Research and Opinion. 2016;32(1).
49. Swallow E, Song J, Yuan Y, Kalsekar A, Kelley C, Peeples M, et al. Daclatasvir and Sofosbuvir Versus Sofosbuvir and Ribavirin in Patients with Chronic Hepatitis C Coinfected with HIV: A Matching-adjusted Indirect Comparison. Clinical Therapeutics. 2016;38(2).

50. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics - Theory and Methods. 1991;20(8).
51. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000;56(3).
52. Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting Survival Time Estimates in the Presence of Treatment Switching. 2014.
53. 日本製薬工業協会. 費用対効果評価における生存時間解析の諸問題. 2020 [cited 2021 Dec 1]; Available from: https://www.jpma.or.jp/information/evaluation/results/allotment/c_e_a_problems.html
54. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technology Assessment. 2003;7(27).
55. Abo-Zaid G, Guo B, Deeks JJ, Debray TPA, Steyerberg EW, Moons KGM, et al. Individual participant data meta-analyses should not ignore clustering. Journal of Clinical Epidemiology. 2013;66(8).
56. Ades AE. A chain of evidence with mixed comparisons: Models for multi-parameter synthesis and consistency of evidence. Statistics in Medicine. 2003;22(19).
57. Dias S, Sutton AJ, Ades a E, Welton NJ. NICE DSU Technical Support Document 2: A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomised Controlled Trials. 2014.
58. Hasselblad V. Meta-analysis of multitreatment studies. Medical Decision Making. 1998;18(1).
59. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. Statistics in Medicine. 1996;15(24).
60. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine. 2004;23(20).
61. Nixon R, Bergvall N, Tomic D, Sfikas N, Cutter G, Giovannoni G. No Evidence of Disease Activity: Indirect Comparisons of Oral Therapies for the Treatment of Relapsing–Remitting Multiple Sclerosis. Advances in Therapy. 2014;31(11).
62. NICE Technology appraisals guidance and advice list [Internet]. [cited 2021 Dec 1]. Available from: https://www.nice.org.uk/guidance/published?ngt=Technology%20appraisal%20guidance&ndt=Guidance
63. Technology appraisal guidance [TA510] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta510
64. Technology appraisal guidance [TA500] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta500
65. Technology appraisal guidance [TA492] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta492
66. Technology appraisal guidance [TA462] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta462
67. Technology appraisal guidance [TA478] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta478
68. Technology appraisal guidance [TA457] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta457
69. Technology appraisal guidance [TA449] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta449

70. Technology appraisal guidance [TA451] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta451
71. Technology appraisal guidance [TA432] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta432
72. Technology appraisal guidance [TA429] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta429
73. Technology appraisal guidance [TA427] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta427
74. Technology appraisal guidance [TA410] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta410
75. Technology appraisal guidance [TA383] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta383
76. Technology appraisal guidance [TA380] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta380
77. Technology appraisal guidance [TA364] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta364
78. Technology appraisal guidance [TA331] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta331
79. Technology appraisal guidance [TA333] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta333
80. Technology appraisal guidance [TA311] [Internet]. [cited 2021 Jan 28]. Available from: https://www.nice.org.uk/guidance/ta311
81. Technology appraisal guidance [TA319] [Internet]. [cited 2021 Feb 3]. Available from: https://www.nice.org.uk/guidance/ta319
82. Steyerberg EW, Eijkemans MJC, Habbema JDF. Stepwise selection in small data sets: A simulation study of bias in logistic regression analysis. Journal of Clinical Epidemiology. 1999;52(10).
83. Faria R, Alava MH, Manca A, Wailoo AJ. NICE DSU TECHNICAL SUPPORT DOCUMENT 17: THE USE OF OBSERVATIONAL DATA TO INFORM ESTIMATES OF TREATMENT EFFECTIVENESS IN TECHNOLOGY APPRAISAL: METHODS FOR COMPARATIVE INDIVIDUAL PATIENT DATA. 2015.
84. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: Analysis of EMA and FDA approvals 1999-2014. BMJ Open. 2016;6(6).

2021 年度データサイエンス部会 KT3

作成担当者
渥美 淳　　東レ株式会社
河田 祐一　　中外製薬株式会社
吉田 瑞樹　　ファイザーR&D 合同会社*

レビュー担当者
東 美恵　　エーザイ株式会社
荒西 利彦　　日本イーライリリー株式会社
奥山 ことば　　MSD 株式会社*
直井 一郎　　大日本住友製薬株式会社
中島 章博　　帝人ファーマ株式会社*
野島 俊秋　　興和株式会社
町田 光陽　　塩野義製薬株式会社

*TF リーダー

監事
酒井 弘憲　　エーザイ株式会社