

<解説>

費用対効果評価ガイドラインの考え方
—比較対照技術と分析対象集団について—

白岩健

国立保健医療科学院保健医療経済評価研究センター

Selection of a comparator and population for cost-effectiveness evaluation

SHIROIWA Takeru

Center for Outcomes Research and Economic Evaluation for Health (C2H), National Institute of Public Health

抄録

我が国における費用対効果評価は2019年4月より制度化された。医薬品・医療機器・再生医療等製品の費用対効果を検討し、その結果を公定価格の調整に用いている。製造販売業者による分析あるいは公的分析を実施するにあたっては、諸外国におけると同様に、分析手法についてまとめられた参照すべきガイドラインを作成している。14章から構成されるガイドラインのうち特に、「3 分析対象集団」「4 比較対照技術」の両章については、費用対効果評価を実施する際の分析枠組みに直結するものであるため、結果に与える影響も大きく、その考え方などについて様々な議論がなされることも多い。そこで、本稿においては分析ガイドラインのうち「3 分析対象集団」「4 比較対照技術」の両章に焦点を当てて、その考え方を改めて解説した。主な見解としては、我が国における費用対効果評価の制度応用状況を鑑みるに、比較対照技術を「最も置き換えるもの」と考えるのは必ずしも適切ではない。また、分析対象集団についても、個々の評価については様々な議論があろうが、少なくともサブグループ解析の解釈は、その限界、例えば α エラーや β エラーがインフレーションすること等について配慮を行っており、結果をP値等により一律に判断するようなことはしていない。臨床的観点等を十分に踏まえた総合的評価に基づいているものである。

キーワード：費用対効果評価，費用効果分析，医療の経済評価，医療技術評価，比較対照技術

Abstract

In Japan, cost-effectiveness evaluation was officially introduced to the pricing system in April 2019. The results of cost-effectiveness are used to adjust the official prices of drugs, medical devices, and regenerative medical products. The official guideline for cost-effectiveness evaluation was published, which describes the methodological issues. The guideline must be referred to when manufacturers submit cost-effectiveness data and researchers review them. The guideline consists of 14 chapters, which, especially, Chapters 3 (target population) and 4 (comparator(s)), attract users' attention because they directly relate to "frameworks of analysis" and their influence on the results of cost-effectiveness. In this paper, we focus on both chapters and explain the background of the description. In our view, the idea that a comparator should be selected from the most replaceable ones is not appropriate if we consider the current Japanese cost-effective evaluation system. In addition, when the results of subgroup analysis are interpreted, we always consider the

連絡先：白岩健
〒351-0197 埼玉県和光市南2-3-6
2-3-6 Minami, Wako, Saitama 351-0197, Japan.
Tel: 048-458-6285
Fax: 048-458-6715
E-mail: shiroiwa.ta@niph.go.jp
〔令和5年3月22日受理〕

limitation of the analysis, for example, the inflation of α error and β error, and not discussed based on the P value only. The results, including the clinical views, are deliberately appraised by the committee.

keywords: Cost-effectiveness evaluation, Cost-effectiveness analysis, Health economic evaluation, Health technology assessment, Comparator

(accepted for publication, March 22, 2023)

I. 緒言

我が国における費用対効果評価は2019年4月より制度化された。医薬品・医療機器・再生医療等製品の費用対効果を検討し、その結果を公定価格の調整に用いている。この費用対効果評価制度においては、対象となった品目を有する製造販売業者が費用対効果の分析結果を厚生労働省に提出するのが原則である。提出された分析結果については、中立的な立場から科学的なレビューを行い、その妥当性を検討する。もし、分析に必ずしも妥当でない部分があると判断されれば、再分析を実施し、科学的な観点からより妥当な分析結果を費用対効果評価専門組織に提出することとなる。このプロセスを「公的分析」と呼んでおり、公的分析は国立保健医療科学院が中心となって実施することとなっている。

製造販売業者による分析あるいは公的分析を実施するにあたっては、諸外国におけると同様に、分析手法についてまとめられた参照すべきガイドラインを作成している。〔『中央社会保険医療協議会における費用対効果評価の分析ガイドライン第3版』2022年1月19日中医協総会了承、以下ガイドライン[1]〕このガイドラインは、学術的な医療経済評価には様々な手法や考え方が存在することを前提とした中で、我が国で実装された政策応用に適する形で可能な限り統一した分析手法を提示するものである。本ガイドラインはその内容に応じて、「1 ガイドラインの目的」「2 分析の立場」「3 分析対象集団」「4 比較対照技術」「5 追加的有用性」「6 分析手法」「7 分析期間」「8 効果指標の選択」「9 データソース（費用を除く）」「10 費用の算出」「11 公的介護費用・生産性損失の取り扱い」「12 割引」「13 モデル分析」「14 不確実性の取り扱い」の14章に分かれて記述されている。なお、本ガイドラインは費用対効果評価分野の専門家により構成される厚生労働科学研究班（政策科学総合研究事業（政策科学推進研究事業）「医薬品・医療機器等の費用対効果評価における公的分析と公的意思決定方法に関する研究」班（研究代表者：福田 敬））により作成されたものであり、筆者もそのメンバーの一人として作成に関与している。

さて、このようなガイドラインの内容については、アカデミアや関係団体等を含む様々なステークホルダーから意見、時には批判を受けるのは当然である。このようなフィードバックを受けながら当該ガイドラインも初版から第3版へと改訂作業を続けてきた。うち特に、「3 分析対象集団」「4 比較対照技術」の両章については、

費用対効果評価を実施する際の分析枠組みに直結するものであるため、結果に与える影響も大きく、その考え方などについて様々な議論がなされることも多い。しかし、それらには分析ガイドラインの考え方がうまく伝わっていないことから、意図せぬ誤解を生じてしまっているケースも見受けられるように感じられる。そこで、本稿においては分析ガイドラインのうち「3 分析対象集団」「4 比較対照技術」の両章に焦点を当てて、その考え方を改めて解説することとした。

II. 比較対照技術について

評価を行う際の比較対照技術は、評価対象技術が分析対象集団への治療として導入された時点で、臨床的に幅広く使用されており、評価対象技術によって代替されると想定されるもののうち、治療効果がより高いものを一つ選定することが原則的な考え方である[1]。

1. 諸外国における比較対照技術の考え方

比較対照技術との相対的關係性においてその技術の医療経済性を評価するのが費用対効果評価の原則であり、比較対照技術として何を選択するかは重要な問題である。臨床研究等でもそうであるように、費用対効果評価においても比較対照技術の取り方に結果が大きく依存する。そのため、比較対照技術の設定方法については、すでに医療技術評価あるいは医療経済評価を政策応用している国々においても、様々な検討がなされてきた。（表1）

比較対照技術における一つの考え方は、「置き換わるもの」である[2-9]。表1にある「標準治療」等の表現も「置き換わるもの」として解釈できる。つまりある医療技術が新規導入されれば、以前の不要になった医療技術（標準治療など）は医療システムの中で使用されなくなっていく。つまり、新規技術の導入により、古い技術が置き換えられる。この置き換えりに注目して費用対効果評価における比較対照技術を設定する方法である。この考え方は、特に税方式で医療を実施している国々、つまり厳しい予算制約の下で費用対効果を検討している国々においてはなじみやすい。新たな医療技術を導入するためには、古い技術に使用していた既存予算を削減して、新たな技術の予算を確保する必要がある。このような医療技術のinとoutの関係が明確な国々においては受け入れやすい発想だろう。

一方で、数は少ないものの比較対照技術を費用対効果フロンティア（cost-effectiveness frontier）（効率性フロ

表1 諸外国のガイドラインにおける比較対照技術の記載

(a)「置き換わるもの」系		
国名	比較対照技術の設定方法	主な英語表記
イギリス	標準治療	established practice
カナダ	標準治療	current care
オーストラリア	最も置き換わるもの	the most replaced
スウェーデン	最も用いられているもの	e.g. the most used
オランダ	標準治療	standard of care or usual care
ノルウェー	置き換わるもの	the completely or partially replaced
ニュージーランド	最も置き換わるもの	most prescribers or clinicians would replace
韓国	シェアの最も高いもの	drug with the highest market share

(b)「費用対効果フロンティア」系		
国名	比較対照技術の設定方法	主な英語表記
フランス	関連するすべてのもの	all clinically relevant interventions
ベルギー	費用対効果フロンティア	efficiency frontier
ドイツ	費用対効果フロンティア	

ンティア (efficiency frontier) とも呼ぶ) に基づいて検討している国々もある[10-12]. 費用対効果フロンティアは、図1のように複数の医療技術が存在する際に、費用効果平面上にそれらをプロットし、効率的となる医療技術を凸関数となるように(増分効果が大きくなるにつれてICERが大きくなるように)結んでいくものである。ただし、費用・効果がともに他の技術より劣るもの(ドミナント、劣位)あるいは2つの医療技術を混合して使用するとその医療技術と同等の増分費用・増分効果が得られるもの(拡張劣位)についてはフロンティアから除外する。

閾値が与えられている場合、ICERが閾値を上回る医療技術については費用対効果がよくないものとして、直線では結ばない。通常のこのフロンティアの先端部に位置する医療技術は費用対効果が最もよい、左記の表現が誤解を招くようであれば、与えられた閾値のもとで原点

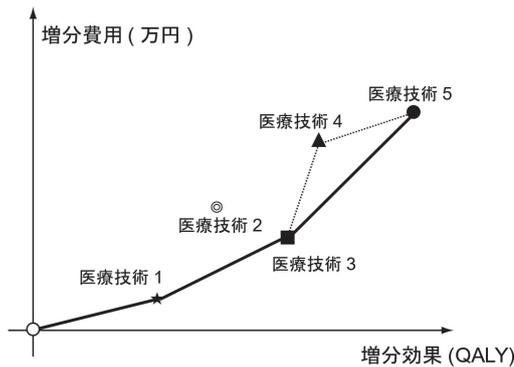
からみたときのINB (Incremental net benefit: 増分純便益) (式1) が最大になる。費用対効果フロンティアに基づき検討する際には多くの場合、新規の医療技術をこの先端部の技術と比較することになる。

$$\text{式 1: } =INB_0 = \gamma \times IE_0 - IC_0 = \gamma \sum_i IE_i - \sum_i IC_i = \sum_i INB_i$$

γ : 閾値 (JPY/QALY), INB_0 : 原点からみたときのINB (JPY), IE_0 : 原点からみたときの増分効果 (QALY), IC_0 : 原点からみたときの増分費用 (JPY), INB_i : 医療技術iのINB (JPY), IE_i : 医療技術iの増分効果 (QALY), IC_i : 医療技術iの増分費用 (JPY)

2. 「置換されるもの」と比較するとはどのようなことを意味するのか?

さて、我が国の費用対効果評価制度において、比較対照技術をどのように設定すべきであろうか。分析ガイド



	効果 (QALY)	費用 (万円)	増分効果 (QALY)	増分費用 (万円)	ICER (万円/QALY)
医療技術 1	1	50			
医療技術 2	1.5	200			劣位
医療技術 3	2	150	1	100	100
医療技術 4	2.25	300			拡張劣位
医療技術 5	3	350	1	200	200

図1 費用対効果フロンティア

ラインの立場からの主張は「『置換されるもの』という概念は我が国の制度において不適である」ということである。それはおよそ以下のような理由による。

そもそも新規技術を「置換されるもの」と比較するとはどのようなことを意味するのであろうか。「置換されるもの」と比較する場合、ここでなされる評価は「新規医療技術を導入するか、導入しないか」という二つの意思決定について検討するものである。すなわち意思決定の結果として出現する「新規技術が導入された世界」と「導入されない世界」を比較することに他ならない。

つまり、何らかの意思決定のもと、現状のまま新規技術が導入されなかった世界（「世界1」）と新たに導入された世界（「世界2」）を考える。この2つの世界は何が違うか。まずは、当然だが当該新規技術が導入されている。そして、新規技術の導入により、今まで使用されていた医療技術の一部あるいは全部に置き換えが起こっている。例えば、ある抗がん剤からより有効性の高い抗がん剤に使用がシフトする。

このとき、患者はより高い治療効果を楽しむのみならず、新規抗がん剤は1ヶ月に1回の投与で済み、既存の抗がん剤は2週間に1回の投与が必要であれば、患者にとっては通院が半分の頻度ですむため、交通費や仕事を休む時間も半分になるかもしれない。病院にとっては、空いた診療枠に別の患者を入れられる。新たな診療枠に入れた患者にとっては、新たに抗がん剤治療が開始され、一定の治療効果が得られる一方で、医療費はより多くかかるだろう。そして、新規抗がん剤を開発したメーカーにとっては、売り上げが大きく増加する。増加した売り上げを研究開発に回すことにより次の新薬が開発される確率が高まるかもしれない。

…といった無限の連鎖を経て、「世界1」は「世界2」へと変化する。このときの変化を可能な限り捉えてそれらを含めた「増分費用（「世界2」-「世界1」）」と「増分効果（「世界2」-「世界1」）」から医療経済性を検討しようというのが「社会の立場」からの費用効果分析である。あるいはアウトカムを金銭化すればそれを「費用便益分析」と呼ぶ、少し誤解されているように感じるが本来の意味での「社会の立場」というのは、生産性費用や介護費用をどうこうしようという狭小な問題ではない。

しかし、これはこれで風呂敷が広すぎるかもしれない。そのような社会全体での変化を捉えることは多くの場合実際には困難であるし、一方で評価の目的が医療費レベルでの効率性を考えるのであればその世界でのみ考えれば十分、あるいはむしろその方が望ましいかもしれない。もう一歩進めて考えれば、医療費の世界に限定した上でも、波及効果を含めてその影響のすべてを考えることは難しい。そこで、世界全体ではなく平均的個人のレベルでの費用や効果の変化（例えば前述の待機患者などは考えずに）のみに着目して、その効率性を検討するにとどめておく考え方もありうる。新規技術が使用されなかつ

たときの「平均的個人1」と使用されたときの「平均的個人2」の比較である。これが、学術的には「医療費支払者の立場」など様々な呼び方がなされるもの、ガイドライン上は「公的医療の立場」と言われ、変化した世界の中の平均的個人レベルに着目して期待費用と期待効果（「個人2」-「個人1」）を考えるものである。

なお、上記の議論を参照すれば、いずれの場合でも「『最も置換されるもの』とは、置換を一つの医療技術に限定して影響力の大きいもののみ検討を行おうというあくまで近似的な分析であり、本来であれば置換される医療技術すべてと比較を実施し、その置換割合で重みつき平均した仮想的な合成技術と比較すべきものと考えている。通常、置換される技術は一つではなくたくさんある。

3. 「置換されるもの」という概念は我が国の制度において不適である

しかし、我が国における費用対効果評価制度はそのような状況とは異なる。我が国における費用対効果評価制度は、新規の医療技術を導入するか（保険償還するか）を意思決定するものではなく、あくまで価格調整に用いるものである。新規技術の導入はすでになされた後で、その付け値が適切であるかを評価する。すなわち新規技術により置換されない世界（「世界1」）というものは存在しない。

そのような状況の中で、無理に「置換されるもの」という考え方にに基づき、2つの意思決定の評価を実施すれば、置換されるものは「付け値のままの医療技術」である。意思決定の結果、付け値の価格そのままで価格調整がなされなかった世界（「世界3」）と価格調整された世界（「世界4」）のどちらかが出来る。このとき価格調整を行うことによって、医療技術レベルでの使用パターンの変化は起こらない（とする）。保険償還から外れることもない。あくまで価格が異なる2つの世界が存在するだけである。

であれば、「公的医療の立場」から評価する以上、価格調整された「世界4」の方が医療費は少なくともすむわけであり、費用対効果に優れるのは自明である。あるいは価格調整の結果、価格が上昇すればその逆である。「増分費用（「世界4」-「世界3」）」は異なるが、常に「増分効果（「世界4」-「世界3」）」は同一である。このような評価には意味がないことは言うまでもない。

4. 費用対効果フロンティアの導入

では、このような「置換されるもの」という概念、2つの意思決定を比較するというやり方を放棄する場合、どうするのか。繰り返しになるが、原点からみるときINBが（あるいは累積のINBが）最も大きくなるのは、費用対効果フロンティアの上に存在するもののうち、その先端に位置する技術である。（なお、もちろんフロンティア先端の技術も多くは「最も置換されたもの」「標準治療」であろう。）

このとき、比較対照としてフロンティア上の（あるいはそれ以外の）どの医療技術と比較すべきか、フロンティアの先端に位置する技術は、原点から見たとき医療システムにとって最も大きな便益を与える医療技術である。このようなフロンティアの先端に位置する技術と比較して、さらに便益を増加させる医療技術を費用対効果がよいものとするのは自然であろう。逆に言えば、費用対効果がよい新規技術は医療システム全体の便益をさらに増加させる。置換すべきかどうかという単純な2択の意思決定が無効な状況で、フロンティア先端の技術との相対的な関係性によって、INBがさらに増加するかどうかを検討するわけである。

このことは、フランスやドイツ（ドイツは実施例がないにせよ）といった費用対効果評価により新規技術を使用するか否かを決定する、といった活用方法ではなく、あくまで価格調整に使用することを志向する国々において、フロンティアに基づいた議論がなされていることと整合しているものと考えている。

なお、このような考え方はドイツで導入されたが、未だ使用されていない効率性フロンティア法に基づくものではないか（もう少し言えば、使用されない方法を用いることは不適当ではないか）、という疑問を浮かべられる方もいるかもしれない。この効率性フロンティア法との類似に関する指摘は、半分は正しく、半分は誤っている。効率性フロンティア法は、効率性フロンティアに基づき比較対照技術と閾値の設定を行うものである。フロンティアの先端の技術を比較対照として分析を実施することは比較対照技術の点では同一であるが、閾値については制度上で設定された外生的な値を用いることから、閾値の設定方法は異なっている。そもそも費用対効果フロンティア自体は、効率性フロンティア法とは無縁に以前より存在している考え方であるし、この考え方自体の妥当性はドイツが効率性フロンティア法を採用するかどうかという議論とは全く関係ない。

ただし、現状では、分析ガイドラインにおいて、「費用対効果フロンティアの先端」と明示せずに「効果の最も高いもの」としている。それは、費用対効果フロンティアの先端の技術を比較対照とするのであれば、費用対効果の劣るもの、つまりフロンティアに乗らないものについては、定義からして自動的に比較対照の候補から除外される。それはそのような技術についてはフロンティア先端の技術と比較してINBが減少することからして当然である。

しかし、比較対照技術として「費用対効果の悪いものを除外するか」は議論のあるところなのかもしれない。政治的な理由での反対もある。筆者個人としては、費用対効果を考えるときに、「費用対効果の悪いものと比較して費用対効果がよい」ということは意味がないと考える。例えば、「ICERが1億円/QALYの技術と比較して、ドミナントでした」と言われても解釈が困難である。ましてや政策的に値上げをすべきものとは思わない。そ

ういった医療技術は比較対照から除外すべきである。

一方で、費用対効果の悪い技術を比較対照技術の候補から除外しない、すなわち閾値を無限大とすれば、フロンティアの先端は「最も効果の高い医療技術」と一致する。このことから、現在の分析ガイドラインの表現になっているものと理解されたい。しかし、費用対効果評価の分析ガイドラインで費用対効果を全く考慮せず「最も効果の高い医療技術」を比較対照にせよとは、畢竟、論理矛盾を来しているように感じるのは筆者だけではないはずだ。この点につき、今後の検討課題であることは間違いない。

III. 分析対象集団について

複数の適応がある場合、あるいは同一疾患内においても治療成績や使用方法・用法用量、比較対照技術が異なる主要な集団がある場合は、各集団についてそれぞれ分析を実施することを原則とする[1]。

1. 分析対象集団とサブグループ解析

費用対効果評価においては、増分費用効果比（incremental cost-effectiveness ratio: ICER）の値をもって価格調整を実施することとなっている。このICERについて、同一疾患内でも異なる集団が存在する場合、その集団ごとに評価を実施するのがガイドライン上の原則である。例えば、あるターゲット遺伝子によって疾患が集団Aと集団Bに区分され、評価対照技術の有効性について、集団Aの方が高い場合を考えれば、両集団のICERが異なる可能性がある。よって、集団を2つに分けて評価することが適切である。これは、集団間で費用対効果の程度に異質性があり、集団Aでは費用対効果がよく、集団Bが悪いとき、2つの集団をあわせた結果、例えば全体集団として費用対効果が悪くなった、としても解釈が難しい。それでは集団Aで費用対効果がよかったことを意思決定に反映できなくなるし、その逆もまたしかりである。費用対効果の程度が異なるのに、2つの集団を混ぜた結果について、一律に「費用対効果がよい」あるいは「悪い」という結論を導くのは一般的には問題があろう。

なお、このように集団を分割することによって、いかようにも恣意的に費用対効果の値を操作できるのではないかと、といった懸念を仄聞することもある。しかし、あくまで対象集団の線引きは外生的な要因（例えば先ほどの遺伝情報など臨床的なもの）によるわけであり、医療経済学の立場から操作しうるものではない。医療技術にはそれが有効な人も、有効でない人もいるわけであるが、そういうばらつきに基づいて集団を分割して分析することは通常しない。就中、医療技術の有効性や比較対照技術等が異なるわけではないのに、性別や年齢、所得など人々の属性によって異なる集団に分けることは、全く適切ではないだろう。実際にそのような評価は行われていない。

費用対効果評価は、リサーチクエスション（分析枠組み）が先立つものであり、データの利用可能性によってそのような分析の枠組みが影響を受けるのは好ましくない。しかし、分析対象集団を分割すると多くの場合、臨床試験の主要なエンドポイントには対応しなくなり、サブグループ解析の結果を参照せざるを得なくなる。サブグループ解析の結果を参照することは当然のことながら、主要エンドポイントの結果等と比べれば信頼性に課題がある。しかし、好ましくないからといって政策的に関心のある分析枠組みに回答が得られなくなるのも問題である。 α や β といった問題は臨床データの解釈にとって重要かもしれないが、それはあくまで検討すべき要素の一つだからである。

サブグループ解析の結果については、様々な課題が統計学的に指摘されており、

- (1)多くのサブグループについて検定を繰り返すことによる α エラーの増大
- (2)集団を分割することによりサンプル数が小さくなる結果としての β エラーの増大
- (3)事前に規定のないサブグループ解析を実施することによる恣意的な結果の導出

などがあげられている。費用対効果評価においても、サブグループ解析の結果を活用する際にこのような点については、もちろん慎重に配慮しなければならない。

(1)と(2)の点については、サブグループ解析の結果を解釈する際には、単純な有意差検定に基づいてP値による判断を実施しないことは、常々強調してきた。データ提出側の製造販売業者としては、特に β が増大することにより検出力が落ちることへの懸念が表明されることが多いが、公的分析や費用対効果評価専門組織においては、巷間言われるように単純にP値に基づいた議論がなされているわけではない。有意水準を大きくすれば、 β エラーはいくらでも低下するわけである。単純なP値に基づいて結果を解釈しない以上、それは必ずしも検出力の問題ではない。実際に、信頼区間の端が1を少し超えたから、といったことをあげつらう解釈はしていない。その代わりに、推定値の大きな不確実性は前提とした上で、臨床的な観点からの見解やその他の様々な立場からの考えを総合して、費用対効果評価専門組織で総合的判断（アブレイザル）を行う。このような「多元的民主主義」「手続的民主主義」（それに値する成熟した組織であるかについては常に再検討の必要があるだろう）によってサポートされる形でデータの解釈を行っていくことが費用対効果評価（医療技術評価）の特徴であることは、ぜひとも強調しておきたい。そして、そのことが究極的には不確実な状況下における意思決定の α エラーや β エラーのインフレーションを防ぐことになるのではないかと考える。一方、(3)の問題については、費用対効果評価の分析枠組みは費用対効果評価専門組織によって最終的に決定されるものであり、データを解析する側の都合で実施するものではない。費用対効果に適したサブグループ解

析については事前にプロトコル等に規定されていないことも多いが、そういった点を考慮しても、サブグループ解析につきまとう恣意性への懸念は通常よりは小さいのではないかと。

2. [C2H1901] テリルジーエリプタをめぐる評価

このようなサブグループ解析に関する懸念は、我が国で初めて実施された評価である「[C2H1901] テリルジーエリプタ」の分析枠組みや評価結果から生じている可能性がある。テリルジーはCOPD（慢性閉塞性肺疾患）患者を対象に開発されたICS（吸入ステロイド薬）/LABA（長時間作用性 β_2 刺激薬）/LAMA（長時間作用性抗コリン薬）3剤の合剤であり、COPDにおける3剤併用療法を2吸入ではなく、初めて1吸入で実施可能にしたものである。

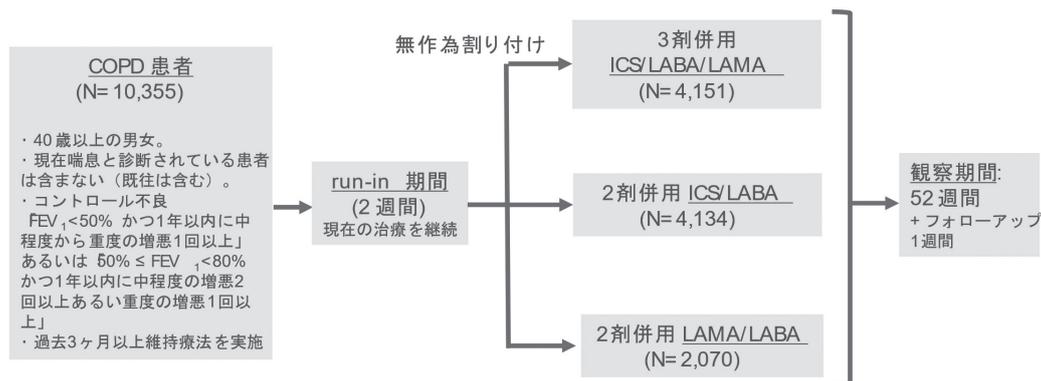
このテリルジーの評価においては、16の評価対象集団を設定し、費用対効果評価を実施したが、多くのサブグループ解析を実施することによる α エラーの増加や検出力の低下等に関連する懸念が出され、種々の批判を受けることとなった。もちろんこのような懸念に対しては一定の理解はするものの、しかし、このテリルジーの評価はかなり特殊な例であり、この評価を過度に一般化して、サブグループ解析の問題点として指摘することは慎重であるべきであると考えている。それは、テリルジーのピボタル試験が費用対効果評価、あるいは医療技術評価の観点からはいくつかの問題を有していることに由来する問題だからである。そこでまずは、その背景となるピボタル試験のIMPACT試験[13]について検討していきたい。

IMPACT試験は、国際共同第3相無作為化二重盲検比較試験であり、約10,000例のCOPD患者を(a)3剤併用群（テリルジー：ICS/LABA/LAMA）、(b)2剤併用群（ICS/LABA）、(c)2剤併用群（LAMA/LABA）に割り付けている。割り付け比率は2:2:1である。（図2）52週後の観察時点で、3剤併用群はいずれの2剤併用群に対しても主要評価項目である増悪の年間発現率において、有意差があった。（表2）

しかし、このIMPACT試験が掲載されたと同号のNew England Journal of Medicine誌のeditorialでは、すでに本試験について以下のような懸念が示されている[14]。

『IMPACT試験の結果を解釈するのは困難をとまなう。なぜなら、約40%の患者が組み入れ時に3剤併用療法を受けており、70%以上がICSを使用し、患者の中には喘息歴のある人も含まれるためである。そして、LAMA/LABA群に割り付けられた患者は、その多くが治療をステップダウンされており、ランダム化時点でICSが突然に打ち切られた。このことはCOPDの増悪を引き起こしうるものである。』

『さらなるエビデンスが利用できるようになるまでは、GOLD 2017ガイドラインに従い、LAMA/LABAによる気管支拡張剤を最大限使用した後で、頻回の増悪をとまなうGOLDのD群の症状を示す患者に限定して、3剤への増



主要評価項目 52 週間での中程度から重度の増悪発現率 (回・年)
ICS/LABA/LAMA とICS/LABA の比較、ICS/LABA/LAMA とLAMA/LABA の比較 (co-primary endpoint)

図 2 IMPACT試験のデザイン

表 2 IMPACT試験の主要評価項目 (医薬品医療機器総合機構の審査報告書より)

投与群	テリルジー群 (ICS/LABA/LAMA)	ICS/LABA	LAMA/LABA
被験者数	N=4151	N=4134	N=2070
総観察期間 (人/年)	3714.9	3457.9	1698.3
COPD 増悪発現回数	3428	3636	1949
年間発現率 (回/人・年)	0.92	1.05	1.15
年間発現率 ^{a)} (回/人・年) [95%信頼区間]	0.91 [0.87 0.95]	1.07 [1.02 1.12]	1.21 [1.14 1.29]
ICS/LABA 群又は LAMA/LABA 群に対する比 ^{a)} [95%信頼区間]		0.85 [0.80 0.90] P<0.001	0.75 [0.70 0.81] P<0.001

a) COPD 増悪発現件数を目的変数、対数変換した観察期間をオフセット、投与群、性別、中等度又は重度の COPD 増悪歴 (1 回以下、2 回以上)、喫煙歴 (スクリーニング時)、地域、気管支拡張薬吸入後の FEV1 の予測値に対する割合 (スクリーニング時) を説明変数とした負の二項回帰モデル
https://www.pmda.go.jp/drugs/2019/P20190322005/340278000_23100AMX00294_A100_1.pdf

加を検討すべきである。』

医師が必要性を認めてICSを投与していた患者に対して、ランダムにICSを打ち切れば、何らかの反応が起こりうるものであろう。特にステロイドのような性質の薬剤であればなおさらである。一方で医師がLAMA単剤あるいはLAMA/LABAで十分であると判断していた患者に、ランダムにICSを加えることは果たして意味があるのかという疑問もわく。実際に、医薬品医療機器総合機構の審査報告書では、

『前治療としてLAMA単剤を使用していた患者では UMEC/VI (注: LAMA/LABA) と本剤の有効性が同程度である可能性が示唆され、本剤投与時には肺炎等の有害事象の発現リスクに留意する必要があること等を踏まえ [以下省略]

との指摘を受けている。このことから、諸外国における医療技術評価機関においても、例えばドイツのIQWiGでは、

『[注: GOLDガイドラインの2剤で治療効果不十分の場合は3剤にステップアップすべきという]推奨が、介入群のICS/LABA/LAMA群では実施されていた。しかし、以前に増悪しているにもかかわらず、LAMA/LABA

群では、ICSが試験の開始とともに突然中止された。比較対照群におけるICSを抜いた治療は、治療のステップダウンとなっており、この時点では不適切である。突然のICSの中止は増悪を引き起こしうる。IMPACT試験でLAMA/LABA群に割り付けられた患者が、適切な治療を受けているかどうかは疑わしい。(中略)よって、IMPACT試験は[注: ICS/LABAとの追加的有用性を検討している]有用性評価には使用できない。』

と評価されていた[15]。このことから、IMPACT試験を用いて、我が国における費用対効果評価制度の中で追加的有用性を評価するためには、全体集団の結果を用いるのではなく、実施されていた前治療別に検討を行う必要がある。そうでなければIMPACT試験を解釈できないと公的分析では判断した。また、ICSの増悪予防効果は、喘息用病態の有無に依存する可能性がIMPACT試験のサブ解析から示されており、そのため好酸球数100/uL等 (Goldガイドライン2019でも好酸球数が100/uL以上で3剤併用へのステップアップを考慮するとされていた) をカットオフ値として、喘息様病態の有無で集団を分けて分析する必要があると考えられた。

その結果として、16のサブグループ (前治療4パター

ン、好酸球数2パターン、比較対照技術2パターン)に分割して費用対効果を検討する分析枠組みが設定されたわけである。この16のサブグループに分割するにあたっては、もちろん事前に検出力などの計算は不可能であったが、いたずらに挑んだものではなく、サンプル数等からおおよそ点推定値については解釈可能な数値が得られることを想定して実施したものである。そもそもIMPACT試験は被験者数が10,000例を超える大規模な臨床試験であり、主要評価項目のP値がかなり小さくなっていることからわかるようにoverpower気味の臨床試験であることも考慮した判断であった。

得られたサブグループ解析の結果と追加的有用性の評価は以下のようなものである。(表3) 集団IIについての解釈を例に取ろう。サンプル数が少ないため、点推定値の不確実性は大きく、統計学的有意差は認めないものの、IMPACT試験を用いて評価をしているC-Lのサブ集団のうち、本集団IIにおいてのみ増悪のリスク比の点推定値は1を超過しており、唯一3剤併用療法群の増悪発生頻度が高い傾向が示されていた。

また、臨床的判断からステロイド剤を含まないLAMAあるいはLAMA/LABAが処方されており、かつ好酸球数も低いような本分析対象集団においては、ステロイド剤を追加することの有用性は臨床的に必ずしも明らかではないとされた。加えてIMPACT試験の事後解析によると、好酸球数が低い患者集団では治療法間で増悪の発生頻度に差がないことも示されている[16]。

そのような観点から、本集団においてのみ増悪のリスク比が高くなり、追加的有用性に疑義が生じることは臨床的見地とも齟齬がないとの臨床家からの見解を得たこ

ととあわせて、好酸球数低値の集団IIについては、追加的有用性なしと判断している。このようにサブグループ解析の結果を決して、単純にP値や点推定値のみから議論しているわけではないことに留意してほしい。

上記の見解、特に臨床的な解釈には様々な批判等もあろうことはもちろん承知している。しかし、このことの賛否とは別にサブグループ解析の結果を単にP値のみで判断して、追加的有用性を否定している事例はその他の集団も含めて存在しない。公的分析においても、費用対効果評価専門組織においても臨床的な見解あるいはその他の臨床データ等を参照しながら、結果を慎重に解釈しているところであり、それは不確実性の大きなデータに基づく意思決定という難問に対して、我々なりに真摯に対応してきた結果でもある。

なお、本件についての見解は、製造販売業者とは最後まで一致しなかったところであるが、一方で報告書についてはキーとなる未公表のサブグループ解析の結果について、多くの社が一般公開されていないデータを非公表とする中で、黒塗り等をせずに公開に同意していただいた。この点については、製造販売業者に深く謝意を表すところである。

IV. 結論

本稿では、分析ガイドラインの中で議論になることの多い、比較対照技術と分析対象集団の課題について、その考え方を中心に解説した。我が国における費用対効果評価の制度応用状況を鑑みるに、比較対照技術を「置き換えるもの」と考えるのは必ずしも適切ではない。また、

表3 IMPACT試験のサブグループ解析の結果

	前治療	好酸球数	比較対照	N (テリルジー群)	N (比較対照群)	ハザード比	追加的有用性
C	MITT (2製剤吸入における3剤併用)	100/ μ L未満	ICS/LABA	398	356	0.83 [0.69, 0.99]	分析不能
D		100/ μ L以上		1181	1204	0.86 [0.77, 0.95]	分析不能
E		100/ μ L未満	LAMA/LABA	398	207	0.78 [0.63, 0.97]	示されていない
F		100/ μ L以上		1181	619	0.67 [0.59, 0.76]	示されていない
G	ICS/LABA	100/ μ L未満	ICS/LABA	307	303	0.78 [0.61, 1.00]	あり
H		100/ μ L以上		910	874	0.82 [0.71, 0.95]	あり
I	LAMA/LABA もしくはLAMA	100/ μ L未満	LAMA/LABA	170	69	1.37 [0.83, 2.24]	示されていない
J		100/ μ L以上		478	261	0.81 [0.65, 1.01]	あり
K	LAMA	100/ μ L未満	ICS/LABA	93	96	0.67 [0.41, 1.09]	あり
L		100/ μ L以上		195	250	0.87 [0.64, 1.19]	あり

(集団A, BについてはIMPACT試験の結果を用いた評価していないことから、本表からは除外した)

個々の評価については様々な議論があろうが、少なくともサブグループ解析の解釈は、その限界、例えば α エラーや β エラーがインフレーションすること等について十分な配慮を行っており、巷間言われるような結果をP値等により一律に判断するようなことはしていない。臨床的観点等を十分に踏まえた総合的評価に基づいている。分析ガイドラインをめぐって、費用対効果評価の結果は分析手法により大きな影響を受けるため、種々の議論はもちろん歓迎するところであるが、それらがより生産的なものとなることを望んでいる。

『世界を解釈するもの、それは私たちの欲求である、私たちの衝動とこのものの賛否である。いずれの衝動も一種の支配欲であり、いずれもがその遠近法をもっており、このおのれの遠近法を規範としてその他すべての衝動に強制したがつているのである。』(ニーチェ「権力への意思」)

利益相反 (Conflict of Interest: COI) に関する情報開示

利益相反なし

引用文献

- 保健医療経済評価研究センター. 中央社会保険医療協議会における費用対効果評価の分析ガイドライン (第3版). 2022. Center for Outcomes Research and Economic Evaluation for Health. [Guideline for preparing cost-effectiveness evaluation to the Central Social Insurance Medical Council: Version 3.0.] 2022. https://c2h.niph.go.jp/tools/guideline/guideline_ja.pdf (accessed 2023-02-22)
- The National Institute for Health and Care Excellence. NICE health technology evaluations: the manual (PMG36). 2022. <https://www.nice.org.uk/process/pmg36/> (accessed 2023-02-22)
- Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada. 2017; 4th edition. <https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition> (accessed 2023-02-22)
- The Department of Health. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee. 2016; 5th edition. <https://pbac.pbs.gov.au/> (accessed 2023-02-22)
- The Pharmaceutical Benefits Board General guidelines for economic evaluations from the Pharmaceutical Benefits Board. 2003. <https://www.tlv.se/download/18.2e53241415e842ce95514e9/1510316396792/Guidelines-for-economic-evaluations-LFNAR-2003-2.pdf> (accessed 2023-02-22)
- National Health Care Institute. Guideline for economic evaluations in healthcare. 2016.
- The Norwegian Medicines Agency. Guidelines for the submission of documentation for single technology assessment (STA) of pharmaceuticals. 2018. <https://legemiddelverket.no/Documents/English/Public%20funding%20and%20pricing/Documentation%20for%20STA/Guidelines%20151018.pdf> (accessed 2023-02-22)
- The Pharmaceutical Management Agency. Prescription for Pharmacoeconomic Analysis. 2015; ver 2.2. <https://pharmac.govt.nz/medicine-funding-and-supply/the-funding-process/policies-manuals-and-processes/economic-analysis/> (accessed 2023-02-22)
- Bae EY, Hong J, Bae S, Hahn S, An H, Hwang EJ, et al. Korean Guidelines for Pharmacoeconomic Evaluations: Updates in the Third Version. *Appl Health Econ Health Policy*. 2022; 20(4):467-477.
- Haute Autorité de Santé. Choices in methods for economic evaluation. 2020. https://www.has-sante.fr/upload/docs/application/pdf/2020-11/methodological_guidance_2020_-choices_in_methods_for_economic_evaluation.pdf (accessed 2023-02-22)
- Belgian Health Care Knowledge Centre. Belgian guidelines for economic evaluations and budget impact analyses. 2012; 2nd edition. https://kce.fgov.be/sites/default/files/2021-11/KCE_183_economic_evaluations_second_edition_Report_update.pdf (accessed 2023-02-22)
- Institute for Quality and Efficiency in Health Care. General Methods. 2022; ver 6.2. https://www.iqwig.de/methoden/general-methods_version-6-1.pdf (accessed 2023-02-22)
- Lipson DA, Barnhart F, Brealey N, Brooks J, Criner GJ, Day NC, et al. Once-Daily Single-Inhaler Triple versus Dual Therapy in Patients with COPD. *N Engl J Med*. 2018;378(18):1671-1680.
- Suissa S, Drazen JM. Making Sense of Triple Inhaled Therapy for COPD. *N Engl J Med*. 2018; 378(18):1723-1724.
- Institute for Quality and Efficiency in Health Care. Fluticasone furoate/umeclidinium/vilanterol (COPD) –Benefit assessment according to § 35a Social Code Book V. 2018. https://www.iqwig.de/download/a18-15_fluticasone-furoate-umeclidinium-vilanterol_extract-of-dossier-assessment_v1-0.pdf (accessed 2023-02-22)
- Pascoe S, Barnes N, Brusselle G, Compton C, Criner GJ, Dransfield MT, et al. Blood eosinophils and treatment response with triple and dual combination therapy in chronic obstructive pulmonary disease: analysis of the IMPACT trial. *Lancet Respir Med*. 2019;7(9):745-756.