サブグループ解析における統計学的留意事項

馬場　亜沙美 (*BABA Asami*)[1*], 鈴木　直子 (*SUZUKI Naoko*)[1], 野田　和彦 (*NODA Kazuhiko*)[1],
波多野　絵梨 (*HATANO Eri*)[1], 髙橋　徳行 (*TAKAHASHI Noriyuki*)[1], LIU XUN [1],
新林　史悠 (*SHINBAYASHI Fumiharu*)[1], 板橋　怜央 (*ITABASHI Reo*)[1], 松田　洋志郎 (*MATSUDA Yojiro*)[1],
柿沼　俊光 (*KAKINUMA Toshihiro*)[1], 山本　和雄 (*YAMAMOTO Kazuo*)[1]

Key Words: ヒト臨床試験，ヒト試験，特定保健用食品，機能性表示食品，安全性，目標参加者数

Statistical Considerations for Subgroup Analysis.

Authors: Asami Baba [1*], Naoko Suzuki [1], Kazuhiko Noda [1], Eri Hatano [1], Noriyuki Takahashi [1], Xun Liu [1],
Fumiharu Shinbayashi [1], Reo Itabashi [1], Yojiro Matsuda [1], Toshihiro Kakinuma [1], Kazuo Yamamoto [1]

[*] **Corresponding author:** Asami Baba [1]

Affiliated institution:

[1] ORTHOMEDICO Inc. [2F Sumitomo Fudosan Korakuen Bldg., 1-4-1 Koishikawa, Bunkyo-ku, Tokyo, 112-0002, Japan.]

Keywords: clinical trials, clinical research, Foods for Specified Health Uses (FOSHU), Foods with Function Claims (FFC), safety, Target sample size

1．はじめに

機能性表示食品制度は，食品表示法（平成 25 年法律第 70 号）の施行に伴い創設され，9 年が経過した。この制度により事業者は，国の定めるルールに基づき，食品の安全性と機能性に関する科学的根拠などの必要な事項を販売前に消費者庁長官に届け出れば，機能性を表示することができるようになった。「機能性表示食品の届出等に関するガイドライン」[1] によれば，最終製品（販売製品）を用いたヒト臨床試験を実施するか，一定のルールに基づき文献を検索し，総合的に評価（systematic review; SR）するかのいずれかを，機能性を評価する際の科学的根拠とする必要がある。いずれの届出方法にせよ，ヒト臨床試験（ヒト試験）を基にした有効性の評価が必須である。

ヒト試験において，同じ介入でも参加者によって反応が異なること，また同じ参加者でも介入によって反応が異なることが知られている。同じ参加者におけるこの後者の反応のばらつきは，通常，説明のつかないままであるが，参加者間の反応のばらつきの一部は，人口統計学的，環境的，ゲノム的，あるいは他の介入に関連する特性によって引き起こされることがもっともらしく，広く受け入れられている。日米 EU 医薬品規制調和国際会議（International Conference on Harmonization; ICH）E5 [2] では「集団の遺伝的，生理学的（内因性）特徴及び文化的，環境的（外因性）特徴」について記述し，臨床試験における多重性の問題に関する欧州医薬品委員会（Committee for Medicinal Products for Human Use; CHMP）の Points to consider（PtC）[3] では「性別，年齢，地域，疾患の重症度，民族的出身，腎障害，吸収や代謝の違いなど，介入効果の不均一性の原因となることが知られている因子がある」とし，「これらの重要なサブグループの分析は，臨床試験の評

[*] **責任著者**：馬場　亜沙美 (Asami Baba) [1]
[1] 株式会社オルトメディコ
〒 112-0002 東京都文京区小石川 1-4-1 住友不動産後楽園ビル 2 階

価の定期的な一部であるべきである」と示している。従って，これらの因子の 1 つ以上において類似した特徴を有する参加者をグループ化することは，臨床試験データセット内の異なる参加者グループ間の介入に対する反応のばらつきを調査する直感的な方法である。

実際に，研究者は可能な限り多くの情報を引き出すために，試験参加者のサブグループの解析を頻繁に用いており，これまでに SR によって届出された機能性表示食品においても，サブグループの結果をもって機能性を表示している商品が見受けられる。上述したように，サブグループ解析は，異なる試験参加者グループ間の介入に対する反応のばらつきを調査する直感的な方法であり，介入効果の不均一性を評価し将来の研究に有用な情報を提供することができるメリットがある一方で，サブグループによってサイズが小さくなることで，要約統計量の不確実性を伴うことが多くなるなど，解析上の課題をもたらし，過大評価や誤解を生む結果を招くデメリットもある。そこで本稿では，サブグループ解析実施における留意点をまとめる。

2. 一般的な考慮事項

2-1. サブグループの定義

本稿において「サブグループ」という用語は，通常ベースライン時に測定される試験参加者の 1 つ以上の内因性（年齢，性別，体重など）及び外因性要因（食事，喫煙，飲酒習慣など）によって定義される臨床試験集団のサブセットを指す。ベースライン時後の測定値は，受けた介入によって影響を受ける可能性があり，介入効果を調査するためのサブグループを定義することは通常適切ではない。また，特定のサブグループから除外された試験参加者は，補完的サブグループとして記述される。

サブグループの構築方法としては，二分法（例 : 男性 / 女性），カテゴリー法（例 : 地域），順序付きカテゴリー法（例 : ベースライン時のスコア等），連続因子（例 : 年齢）に基づいて定義することができる。また，その定義は，臨床における意思決定に関連する参加者集団の意味のある実体（亜集団）を反映したものでなければならない。そのため，カテゴリー的因子と連続的因子の調査は，単一因子の複数のレベルにわたるプーリングや，連続的因子のカテゴリー化のためのカットオフポイントが必要な場合，慎重な検討が必要となる。

バイオマーカーや遺伝子検査に関する研究は，2 つの側面からサブグループの定義に問題を生じさせる。第一に，複数のバイオマーカーの組み合わせと分類に基づいて，ほとんど無限の数のサブグループを定義することができる。第二に，バイオマーカーの予後的価値とその臨床的有用性に関する知識は通常不足している。したがって，このような方法でサブグループを定義することを検討する際には，特に注意が必要である。

2-2. サブグループ解析の重要性

臨床試験集団における異質性の程度は，主に臨床試験集団が反映する標的集団における異質性の程度によって決定され，検証的臨床試験に組み入れられる参加者集団の異質性の程度は，試験が実施される地域や，場合によっては介入などによって決まる。研究対象集団が異質であればあるほど，推定された全体的効果が関連するサブグループに広く適用されることを確認するためのサブグループ解析の重要性が増す。臨床試験において，特定の関連因子によるサブグループによって介入効果を検証することは，効能効果の文言だけでなく，介入効果を修飾する因子が存在することや介入効果が不確実であるという情報を消費者に提供する可能性がある。しかし，何が関連因子とみなされるべきかについて検証されることは，今後の重要な知見となる。

サブグループ解析は，事前計画なしに実施されることがほとんどだが，本来は，対象集団内の異質性が予想され，介入効果に一貫性がない可能性があると予想されたうえで，試験集団全体を評価した後に実施されることが正当な手順である。

2-3. サブグループの評価における重要な概念

ここでは，サブグループの評価において重要な概念を紹介する：

1. 異質性（Heterogeneity）は，介入効果の予測因子における，対象介入集団または臨床試験集団内の差異の程度に関係する。集団の不均一性が高ければ高いほど，明確に定義されたサブグループにおける介入効果の調査がより重要になる。

2. 一貫性（Consistency）とは，関連するサブグループにおいて推定された介入効果が，広範な範囲に適用されることを保証する程度を表す。
3. 信頼性（Credibility）とは，サブグループの所見が十分に立証されていると結論づけられ，それゆえ意思決定に依拠できる程度を表す。信頼性は，十分な根拠のある先験的定義，特定の所見の生物学的妥当性，再現性（下記参照）の程度に依存する。
4. 生物学的妥当性（Biological plausibility）とは，臨床的，薬理学的，機序的考察及び他の関連する外部データ源の考察に基づいて，特定の効果（この場合，サブグループ間の介入効果の差）がどの程度予測されるか，または予測されたかを表す概念である。妥当性は，主に臨床的・薬理学的判断であり，計画段階ですでに考慮されていない限り，通常，直接定量化・測定可能な概念ではない。
5. 再現性（Replication）とは，特定の共変量の効果や特定のサブグループ間の差異効果が，複数のデータ源で認められるかどうかをいう。

2-4. 複数のサブグループ解析を行う際の問題点

サブグループ探索の重要な問題は，多重検定の問題と密接に関連している。サブグループを同定できる複数の因子が利用可能であり，サブグループをどのように構築するかを選択する機会（たとえば，連続因子の異なる分類など）があると，両方とも多重性をもたらし，これらのサブグループの分析は，単に偶然の遊びのために矛盾した結論につながる可能性がある。

複数のサブグループを考慮すると，偽陽性所見の確率が高くなる。ここで定義する偽陽性所見とは，効果が一次分析集団で見られた効果と異なると結論づけられるが，実際にはそうではないサブグループのことである。偽陰性の結論も可能であり，同様に重要である。これらは，効果が全体的な介入効果と本当に異なることが検出されなかったサブグループと定義される。

偽陽性所見の可能性は，サブグループとその補完群における差異効果を無視または棄却する理由としてしばしば引用される。批判的に言えば，これは異なるサブグループ間の効果が試験全体の結果と一致しているという根本的な仮説を調査しないことを意味する。さらなる調査や議論なしに，重要なサブグループに一貫した効果があると仮定することは容認できない。

臨床試験の結果を評価する際，無作為化によって得られたバランスがサブグループに分けて検討する際に完全には保たれず，複数のサブグループの 1 つにおける所見が，介入の効果よりも介入群間の共変量におけるベースラインの不均衡によってもたらされる可能性が高くなるという追加的なリスクがある。試験全体の結果と矛盾するサブグループの所見は，まず共変量の不均衡を適切に調整することで，介入効果の全体推定値とサブグループ推定値の差を説明できるかどうかを調べる統計的手法で検討されることが期待される。

2-5. サブグループにおける介入効果の一貫性の評価方法と関連データの提示

サブグループ解析における介入効果の一貫性を評価することは，広範な試験参加者集団を組み入れたことを証明するために重要である。しかしながら，介入効果の一貫性はそれ自体で価値があるわけではないことに注意すべきである。また，全体の介入効果とサブグループ解析における介入効果の違いが認められた場合もあるが，その際は，そのサブグループには介入は有効ではない可能性があるため，別の議論が必要になる。

交互作用の統計学的検定は，試験集団のサブグループの一貫性を評価するために用いられてきた。しかし，現在利用可能な検定では，潜在的に臨床的に重要な介入効果の矛盾を検出する検出力が不足しており，第一種の過誤のレベルを上げて評価すると特異性が失われることが証明されている。このことは，メタアナリシスにおける異なる試験間の異質性，または 1 つの試験内のサブグループ間の異質性の他の尺度にも当てはまる。サブグループのサイズ大きさは，サブグループを形成する因子に依存し，サブグループとその補集合の大きさが異なると，異質性の検定はさらに検出力を失うという事実によって，状況はさらに複雑になる。したがって，交互作用の検定の役割は限定的であり，統計的な側面だけをカバーし，さらなる検査のためのシグナルを生成する場がある。交互作用の検定は，臨床的意思決定に関

連する可能性のあるサブグループにおいて，一貫性のない所見を同定したり除外したりする唯一の手段とはなりえない。シグナル生成におけるその役割から，これらの検定を実施する場合は，5% 以上の有意水準を採用すべきである。また，今でも一般的に行われていることではあるが，交互作用の検定から得られた単独の有意確率を報告するだけでは，意思決定の根拠としては不十分である。サブグループにおける推定介入効果を評価し，観察された差の臨床的妥当性を議論することが重要である。

探索的サブグループ解析は通常，様々な因子について提示される。結果の提示には，推定値と信頼区間を含めるべきである。サブグループ解析を表示する場合は常に，補完サブグループの解析も表示する。二項変数またはカテゴリー変数がサブグループの定義に使用される場合は，Forest plot で結果が表示されることが期待される。介入効果の方向性と大きさの一貫性を示す Forest plot は，試験の結果が調査された患者集団に適用されるという全体的な結論に妥当性を加えるものとして一般的に受け入れられている。連続変数の場合は，推定される介入効果が因子の範囲にわたってどのように変化するかを特徴付けるプロットを提示する。複数の重要な臨床試験が提示される場合は，臨床試験間の一貫性を評価するだけでなく，プールされたデータセットの Forest plot を提示することもできる。

あるサブグループで介入効果が平均介入効果より大きい場合，補完的なサブグループは必然的に平均より小さい介入効果を示す。方向性の一貫性が明らかでなく，さらなる検査が必要な場合は，提示された異なるサブグループは独立しておらず，観察された複数の矛盾の原因が同じ根本的な要因である可能性があることに注意する。例えば，腎機能不全は年齢とともに増加し，介入効果と腎機能との潜在的な交互作用は，腎機能のレベルによって定義されたサブグループでも，年齢によって定義されたサブグループでも，また年齢と相関のあるサブグループでも現れるであろう。Forest plot は，矛盾の潜在的な原因を綿密に調査する必要性と，全体的な有効性の評価においてどこまで考慮する必要があるかを示している。

サブグループ解析や交互作用検定に関連する多重性は，検定の有意水準を変更することや信頼区間を表示することで対処すべきであると議論されてきた。しかし，これらの調査は主に，さらなる調査を開始する必要性の指標となるため，調整は直感に反することになり，推奨されない。介入による不都合な影響が見落とされるのを避けるために，複数のサブグループを調査するという事実は，その妥当性に関するシグナルを注意深く評価し，現在の適用内外の再現性を見つける能力によって補償する必要がある。

2-6. サブグループ発見の「信頼性」

2-3. で示したように，生物学的妥当性と再現性を見出すことは，サブグループの所見の信頼性を評価するための重要な要素である。どの因子が参加者の介入効果を予測するかについては，試験計画時にある程度のエビデンスが得られているであろうが，対象サブグループにおける所見の信頼性は，臨床試験で得られたデータ，及び試験期間中に得られた他の外部データや知見に基づいて再評価する必要がある。評価者は，特定のサブグループにおける所見が信頼できるとみなせるかどうかという問題に対して，持ちうるすべての証拠を総合的に判断しなければならない。

試験計画時に得られる知見は，層別化や解析にどの因子を用いるか，さらにどのサブグループの調査を計画するかの選択に役立つ。このような事前特定は，サブグループ間で結果に差が出る可能性があることを反映し，肯定的または否定的なサブグループの所見に信頼性を与えることができる。逆に，事前規定がないからといって，特定のサブグループでの結果が信頼性に欠けるという直接的な論拠にはならない。特に，サブグループにおける不利な所見については，計画段階で関連するすべてのサブグループを適切に検討し，事前に特定することを阻害すべきではない。その代わりに，信憑性の欠如の論拠は，生物学的妥当性と再現性の欠如に焦点を当てるべきである。

3. 研究計画への影響

検証試験を計画する際，被験品に関する知識は限られているため，サブグループの評価を事前に完全に特定できることはまれである。しかし，優れた試験計画では，試験解析と評価の戦略を改善するため

に，可能な限り計画段階で知識を傾ける。このような考察は，計画段階でサブグループについて適切に議論することを支援するものである。なぜなら，妥当性の考察は，通常，試験データの知識に影響されないように試験前に行うほうが，説得力があるからである。

3-1．対象集団内の異質性を考慮する

臨床試験の計画中，介入効果に影響を与える因子ついて議論することは，組入れや除外の基準を組み立てる際に非常に役に立つ。広範な患者集団を対象とした試験は，広範な効能・効果を支持するのに役立つが，介入に対する反応の一貫性を調査する重要性も増す。実施可能であれば，対象集団を幅広くリクルートした大規模試験が 1 つ以上あれば，意思決定に最適である。これにより，効果修飾因子を知る可能性が高まり，有効性に関する適切な結論を下すことができる。

検証試験が実施されるにつれて介入に関する知識は増加するため，異質性の潜在的な原因をすべて事前に予測できるわけではないことを認識しなければならない。試験の過程で得られた知見は，試験実施計画書に反映されているか否かに関わらず，解析段階で反映される必要がある。ブラインドレビューが実施される場合，現在の試験の結果を知らなくても，利用可能な証拠を再検討し，計画を修正する機会となりうる。

3-2．評価のためのサブグループの選択と定義のための戦略

計画された臨床試験で採用される集団を慎重に検討し，介入効果が一貫していないと予測できる強い理由が存在する部分集団を除外した後，介入効果が一貫していることを前提に計画を進めることはできるが，この仮定が正しいかどうかの調査を予見しなければならない。

通常，採用された集団が母集団への介入効果を反映していることであろう（試験の外部妥当性）。層別化の必要性は，第一に異なる因子レベルの参加者を介入群に不均衡に割り当てるリスクを低減するため，第二に，異なるリスクプロファイルの参加者が被験品の使用から同じ効果を得られるかどうかについて，精査の対象となるべき因子のいくつかを示すために考慮されるべきである。具体的には，層別化に使用される因子が，被験品の有効性を予測するものであるかどうかの調査が必要である。しかし，層別化は，限られた数の因子しかモデルに含めることができないため，計画段階で，最も重要な因子を特定するために，治験責任医師と十分に話し合うことが重要である。適切に議論され，文書化されれば，評価段階で特定のサブグループ所見の重要性を決定するのに役立つ。

計画段階で異なるサブグループ解析の相対的な重要性を明確に理解することは，事後的な議論を最小化するのに役立つ。また，重要なサブグループの評価に十分なエビデンスを作成することは，集団のサブセットにおける有効性またはその欠如に関する誤った結論のリスクを低減する試みとして推進される。

一般的に，ある因子がどの程度アウトカムに影響するかについて，因子の形式（二値，カテゴリー，連続など）を尊重した初期調査を計画すべきである。連続共変量の機能形式（例えば，線形関係）は，サブグループ定義のための適切な分類が可能であるように，よく理解されるべきである。臨床試験の結果，介入効果が特定の因子のレベルにより異なることが示された場合，その後の検討は，カテゴリー化などに基づく必要があるかもしれない。この可能性は，計画段階で注意深く考慮されるべきで，最終的にこの目的に役立つかもしれないカテゴリーをあらかじめ指定する。結論の頑健性に及ぼすカットオフの選択を調査する分析も計画すべきである。

臨床試験のサンプルサイズは，通常，全集団の分析に基づく臨床試験の目的を達成するために計画される。しかし，介入効果の一貫性を評価するために，主要なサブグループで十分なエビデンスが得られるかどうかも考慮することが推奨される。この点については，介入効果の一貫性が観察されないリスクを事前に考慮した上で，ケースバイケースで検討する必要がある。

4．まとめ

本稿では，サブグループ解析における留意点をまとめた。サブグループ解析にはメリットもあるがデメリットを併せ持つことを理解し，解析計画に事前に組み込んでおくことが重要である。

参考文献

1. 消費者庁 . 機能性表示食品の届出等に関するガイドライン . [Internet]. 2017 [cited 2024 Apr 9]. Available from: https://www.caa.go.jp/policies/policy/food_labeling/foods_with_function_claims/assets/foods_with_function_claims_230929_0002.pdf
2. 厚生労働省 . ICH E5 外国臨床データを受け入れる際に考慮すべき民族的要因についての指針 . [Internet]. 2017 [cited 2024 Aug 6]. Available from: https://www.pmda.go.jp/files/000156571.pdf
3. Committee for Proprietary Medicinal Products (CPMP). Points to Consider on Multiplicity Issues in Clinical Trials. [Internet]. 2017 [cited 2024 Aug 6].