

機能性表示食品における「鼻目のアレルギー反応」 評価試験：手引きの実践的解釈と試験デザインの最適化

馬場 亜沙美 (BABA Asami)^{1*}, 鈴木 直子 (SUZUKI Naoko)¹,
田中 瑞穂 (TANAKA Mizuho)¹, 山本 和雄 (YAMAMOTO Kazuo)¹

Key Words：ヒト臨床試験, 特定保健用食品, 機能性表示食品, 鼻目のアレルギー反応

Clinical Evaluation of Nasal and Ocular Allergic Symptoms for Foods with Function Claims: Practical Guidelines and Trial Optimization

Authors: Asami Baba^{1*}, Naoko Suzuki¹, Mizuho Tanaka¹, Kazuo Yamamoto¹

***Correspondence author:** Asami Baba

Keywords: Human clinical trials, Foods for Specified Health Uses (FOSHU), Foods with Function Claims (FFC), Nasal and ocular allergic reactions

Affiliation:

¹ ORTHOMEDICO Inc.

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

1. はじめに

機能性表示食品制度においては、疾病に罹患していない者を対象に、機能性関与成分によって特定の保健の目的が期待できる旨を表示することが認められている。そのため、機能性の科学的根拠となるヒト臨床試験（ヒト試験）は、原則として「健常者」を対象とし、対象とする摂取者層において機能が確認されていることが求められる。

一方、「鼻目のアレルギー反応（花粉やハウスダスト等によるアレルギー性鼻炎・結膜炎の症状）」は、症状が季節やアレルゲン飛散量、生活環境に大きく左右され、日々の変動も小さくない。さらに、症状評価の多くは主観的指標であり、レスキュー薬（治療薬）の使用状況によっても主観的指標のスコアが

変化する。このような領域では、「健常者のみ」に限定した試験デザインが現実的に困難な場面も多く、軽症者が含まれるデータの取扱いがしばしば論点となる。

こうした背景を踏まえ、消費者庁は機能性表示食品の届出等に関する手引き（以下、手引き）の別紙2「軽症者が含まれたデータの取扱いについて」において「鼻目のアレルギー反応関係」などの一部の表示領域に限り、一定の条件のもとで健常者に加えて軽症者が含まれるデータを例外的に使用できる枠組みを示している¹⁾。本稿では、この手引き（別紙2）の要点を整理した上で、推奨される試験方法論を提案し、あわせて解釈における注意点をまとめる。

*** 責任著者：**馬場 亜沙美 (Asami Baba)

所属機関：

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

表1 手引き（別紙2）が示す「鼻目のアレルギー反応関係」の要求事項1)

項目	要求・推奨される内容
試験方法	原則としてランダム化比較試験（RCT）とする。 ※並行群間比較試験，クロスオーバー比較試験のいずれも可。
盲検性	二重盲検試験を推奨。 ※機能性関与成分の特性によりプラセボ作製が困難な場合等は，単盲検試験も許容される。
評価指標	「鼻目症状」および「日常生活の支障度」を併せて評価する。 ※支障度は鼻目症状の変化に伴うことが条件。
評価指標の例	「鼻アレルギー診療ガイドライン 2020 年版」の重症度分類，「日本アレルギー性鼻炎標準 QOL 調査票」，またはそれに類似した海外の指標や妥当性が検証されたもの。
判定条件の特例	- 一部の指標のみで機能が確認された場合でもデータとして使用可能。 - アレルギー飛散量との関係等が適切に考察されていれば，一部の測定ポイントのみの有意差でも使用可能。
摂取（試験）期間	明示規定なし。
対象者	「健常者」または「健常者および軽症者」を対象とする。 【健常者の定義】 鼻目のアレルギー反応を有し（過去に有していた者を含む），かつ，試験前および試験期間中にアレルギー治療薬を摂取していない者。 【軽症者の定義】 鼻目のアレルギー反応を有し（過去に有していた者を含む），かつ，試験前および試験期間中にアレルギー治療薬を時々摂取している（常用していない）者。
機能性の確認方法	「健常者」または「健常者と軽症者全体」で機能を確認する（有意水準は両側 5%）。 ※全体で確認する場合，対象者におおむね半数以上の健常者が含まれていることが条件。割合が不明な場合は，評価指標の数値等から半数以上と推測できる適切な理由が必要。
他領域の疾病者が含まれるデータの取扱い	論文において対象者が健常者として扱われていれば，目的とする機能性以外の検査値等が疾病域に該当しても使用可能 ※ただし，該当する検査値等が「疾病域」「非疾病域」「不明」のいずれの内訳に該当するかを確認すること。

2. 手引き（別紙2）の要点整理

機能性表示食品制度は「疾病に罹患していない者」を対象に，機能性関与成分による保健の目的を表示する仕組みである。したがって，機能性の科学的根拠となるヒト試験も，想定される摂取者層における機能確認を目的として実施されることが前提となる。原則として特定保健用食品（トクホ）のヒト試験の考え方に準拠しつつも，例外的に軽症者等を含むデータの使用が一定の範囲で認められており，その領域の一つが「鼻目のアレルギー反応関係」である。当該領域に関して，手引き（別紙2）が求めている要求事項の要点を表1に示す。

3. 実務に適した試験デザインと方法論

手引き（別紙2）の要件を十分に理解したうえで，統計学的にブレが生じにくいプロトコル設計に落と

し込む必要がある。本章では，実務上重要となるポイントをいくつかの要素に分けて解説する。

3-1. 試験デザイン

基本として，並行群間比較デザインの採用を推奨する。手引き上ではクロスオーバー比較デザインも許容されているが，「鼻目のアレルギー反応」領域では期（Period）の違いによりアレルギー曝露量が大きく変動するため，交絡が避けられない。

仮にクロスオーバーデザインを選択する場合，ウォッシュアウト期間の設計根拠と，期間効果（Period effect）の検討について統計解析計画書（SAP）に事前記載することが必須である。この事前の取り決めがない場合，実施したクロスオーバーデザインによる試験の妥当性を第三者から指摘されるリスクが高まる。

また、主要評価項目は主観的評価が中心となるため、二重盲検法が原則である。プラセボの作製困難等により例外的に単盲検が許容される場合であっても、以下の要件を満たす設計にしなければ客観性の担保を指摘されやすい。

- 電子日誌の導入による評価の標準化
- 入力タイミングの固定（毎日同時刻の入力徹底など）
- 説明文言の統一（試験参加者へのバイアスを排除した説明手順の確立）
- 主要解析の事前定義（臨床試験登録システムおよびSAPでの公開）

実務上の工夫として、完全な二重盲検が困難な場合でも、エンドポイント評価（集計・判定・解析）を盲検化する設計を組み込むことで、主観的指標に伴うバイアスを抑制できる。たとえば、PROBE法（Prospective, Randomized, Open-label, Blinded-Endpoint）の概念を応用し、試験参加者への割付自体はオープン（または試験参加者が推測し得る状態）であっても、データ回収、集計および統計解析の担当者を割付情報から遮断する運用を徹底する。前述の電子日誌等によって評価者の介入を最小化し、事前に定めたプロトコルどおりに実施すれば、QOL評価の解釈可能性とデータの信頼性を実務的に補強することが可能である。

3-2. 割付方法

3-2-1. 層別ブロックランダム化（Stratified Block Randomization）

健常域と軽症域の両方を試験に組み入れる場合は、層別ブロックランダム化が標準となる。層別因子を明示し、割付の予測不可能性を担保するため、置換ブロック法（Permuted Block）を用いることが

望ましい。実務上想定される主な層別因子は以下のとおりである。

- 健常者 / 軽症者の区分（※本要件においては必須）
- 施設（多施設共同試験の場合）
- 性別（症状の出方や申告傾向に性差が想定される場合）

ブロックサイズは、4 または 6 が一般的であるが、ブロックサイズを固定すると試験現場で割付予測のリスクが生じるため、ランダムブロックサイズ（4, 6, 8 の混合など）を採用することが統計学的に推奨される。ただし、小規模試験（総サンプルサイズがおおむね 40 未満）においては、各層における群間の例数均衡を優先するため、あえてブロックサイズを固定運用することもある。

3-2-2. 最小化法（Minimization）

結論から言えば、当該領域において最小化法を積極的に採用する理由は乏しい。

最小化法は、割付時点において群間のバランスを動的に最大化できる利点があるものの、以下の理由から、機能性表示食品の届出を企図したヒト試験（特に「鼻目のアレルギー反応関係」の領域）においては推奨しにくい。

- 健常者 50% 以上の制約との相性の悪さ

手引き（別紙 2）は「対象者におおむね半数以上の健常者が含まれること」を要件としている。この要件を満たすには、参加者を募集する段階で健常者枠と軽症者の枠（層）をそれぞれ独立して設定し、登録比率をコントロールする必要がある（表 2）。

つまり、健常者比率の制約が存在する時点で、最小化法特有の「動的にバランスを取る」という利点を活用できず、層別ブロックランダム化で十分に目的を達成できることとなる。

表 2 層別ブロックランダム化と最小化法の違い

方式	運用上の特徴と実務への適合性
層別ブロックランダム化	健常者層・軽症者層それぞれに募集枠（例：健常者 50 名、軽症者 30 名）を設定し、各層内で独立にブロック割付を行う。層ごとの登録状況を見ながら募集を調整するため、最終的な比率要件を確実にコントロールできる。
最小化法	割付時点で「群間バランスが最も良くなる群」に割り付けるアルゴリズムのため、層ごとの人数枠という概念を持たない。健常者と軽症者の最終比率は、登録された試験参加者の構成に依存するため、募集途中で軽症者の登録を停止などの追加運用が発生し、手続きが煩雑になる。

表 3 状況に応じた割付方法の選択基準

状況	推奨する割付方法
健常者 50% 以上の制約あり（本領域の標準）	層別ブロックランダム化
層別因子が 2 から 3 個以内	層別ブロックランダム化
層別因子が 4 個以上かつ比率制約なし	確率付き最小化法を検討する

● 割付の予測可能性の問題

純粋な最小化法は決定論的（deterministic）なアルゴリズムであるため、次の割付が予測可能になるリスクがある。手引きでは「ランダム化」を明示的に要求しているため、予測可能性に対する説明責任が生じる。確率付き最小化法（Biased Coin Minimization）を用いて確率的要素を加えれば予測可能性は低減するが、その確率設定の根拠を SAP に明記する必要性が生じ、結果として説明コストが増大する。

● 層の細分化問題が起きにくい領域特性

一般的に、層別因子が 4 つ以上存在し、層別ブロックランダム化では層が細分化しすぎてしまう場合に最小化法が検討される。しかし、「鼻目のアレルギー反応」領域では、層別因子は通常 1 から 3 個（健常 / 軽症、施設、性別など）で収まるため、層の細分化問題自体が起きにくい。

実務上は、表 3 に示す状況に合わせて割付方法を判断することが望ましい。

なお、本領域においてあえて最小化法を採用する場合は、以下の点が指摘されやすいため、プロトコルおよび SAP の記載には注意が必要である。

- 最小化法を使用しているにもかかわらず、「ランダム化した」とだけ記載し具体的なアルゴリズムを明記していない
- 最小化法を選択した医学的・統計学的な理由が不明確である（層別ブロック法で対応可能であるにもかかわらず選択している）
- 健常者比率の制約（50% 以上）を満たすための具体的な運用方法がプロトコルに記載されていない
- 確率付き最小化法を採用した場合の、確率設定根拠が不明確である
- 割付予測リスクに対する具体的な対策（中央登録システムの導入など）が記載されていない

3-3. 評価指標の設定

手引き（別紙 2）が求めている評価指標は、本質的に「症状」と「生活への影響」の 2 要素である。具体的には「鼻目症状」と「日常生活の支障度（QOL 等）」を用いることが明記されているが、ここで重要なのが「支障度は鼻目症状の変化に伴うこと」という条件である。

この「伴うこと」は、症状の改善と QOL の改善が同方向に動く（連動性を示す）ことを要求していると解釈するのが自然である。換言すれば、QOL 単独の有意差をもって機能性を主張する設計にはしないこと、という行政側の設計思想が読み取れる。

3-3-1. 主要評価項目（Primary Endpoint）の単一化

鼻目のアレルギー領域のヒト試験は、評価指標が多岐にわたりやすいため、プロトコル設計の第一歩は「主要評価項目を一つに固定する」ことである。手引きには「一部の指標でも機能が確認できればデータとして使用できる」という柔軟な文言があるが、これを曖昧な試験設計の救済措置として認識してしまうと、後からの科学的説明が困難になる。実務上は、主要評価項目が曖昧な試験ほど指摘を受けやすいと捉えておくのが安全である。

基本方針は、「鼻症状」と「目症状」をまとめた総合スコアを一本化することである。具体案としては以下のいずれかが現実的である。

- 合計鼻症状スコア（Total Nasal Symptom Score; TNSS）および合計眼症状スコア（Total Ocular Symptom Score; TOSS）の合計：鼻と目の症状を同時に捉えやすく、解釈が直感的である。「日本アレルギー性鼻炎標準 QOL 調査票」の症状項目から合成することも、毎日の症状日誌から算出することも可能である。
- 鼻アレルギー診療ガイドライン（2020 年版）の重症度分類を連続スコア化：ガイドライン準拠を前面に出せる点が強みである。
ポイントは、主要評価項目を必ず 1 つに絞ること

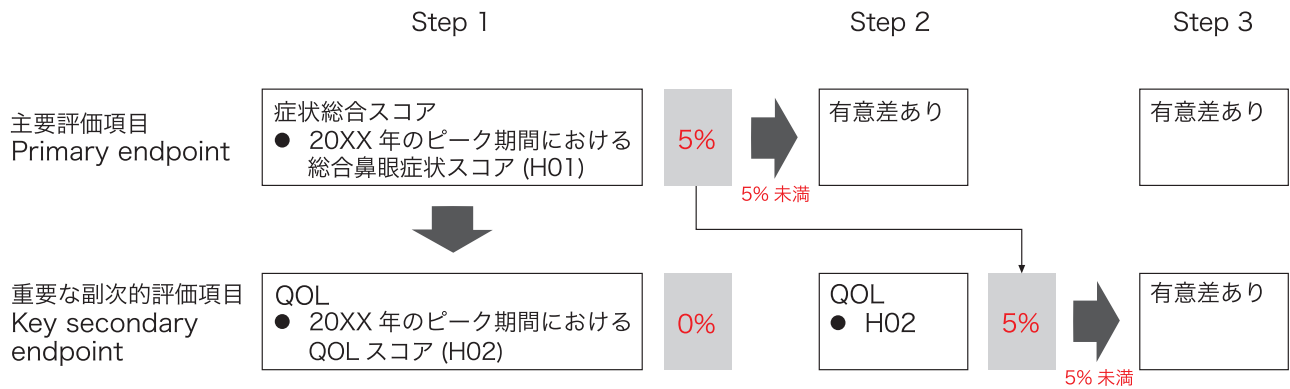


図1 主要評価項目から重要な副次評価項目（QOL）への階層的検定（ゲートキーピング）の例

である。主要評価項目が複数存在する、「どれが真の主要評価であったのか」「検定の多重性はどう制御したのか」という議論に収束しやすく、機能性の主張が散逸するリスクが生じる。

3-3-2. QOLの重要な副次評価項目（Key Secondary Endpoints）としての位置づけ

QOLは患者にとって重要なアウトカムである一方、手引きの要件では「日常生活の支障度は鼻目症状の変化に伴うこと」が明示されている。したがって、本領域のヒト試験では、QOLを独立した主要評価項目として機能性を主張するのではなく、症状スコアで捉えた改善が日常生活上の便益として整合的に現れていること（＝構成概念妥当性や臨床的整合性）を補強する指標として位置づけるのが合理的である。

具体的には、重要な副次評価項目としてQOLを設定し、主要評価項目と同方向の変化が観察されること、あるいは両者の変化量に一定の連動性があることを示すことで、効果の解釈可能性と説得力を高める役割を担わせる。考察においては、以下の根拠を添えると説得力が増す。

- 症状スコア変化量とQOL変化量の散布図（同方向の変化が視覚的に確認できるか）
- 相関係数（Spearman / Pearson など）による連動性の定量化

【多重性の制御について】

QOLをもう一つの主要評価項目に格上げしないことである。主要評価項目はあくまで症状スコアに一本化し、QOLはその結果の解釈可能性を高める重要な副次評価として整理する方が、別紙2の要求

（「支障度は症状変化に伴うこと」と最も整合する。

しかし、主要評価項目（症状スコア）とは別の仮説検定を行う以上、Key Secondaryとして結論に用いる場合は多重性の制御が必須である。したがって、QOLを重要な副次評価項目として扱う場合には、主要評価項目で有意が確認された場合に限ってQOLを検定する、といった階層的検定（gatekeeping / hierarchical testing）をSAPで事前規定し、家族内第1種過誤率（FWER）を管理する。逆に、この事前規定がないままQOLの有意差を強調すると、後付け解釈（選択的報告）と見なされやすくなる。

主要評価項目で有意差が確認された場合にのみ、重要な副次評価項目（QOL）の仮説検定を行う手順を示す。これにより、主要とQOLを別々に検定しても家族内第1種過誤率（FWER）を5%に管理できる。

3-3-3. その他の副次評価項目（Secondary Endpoints）

副次は「解釈の補助」に徹する。特に軽症者を含む設計なら、レスキュー薬（治療薬）の使用状況を把握することはほぼ必須である。

- 個別症状（くしゃみ、鼻水、鼻づまり、目のかゆみ等）
 - レスキュー薬（治療薬）の使用状況：使用有無・回数・用量
 - 軽症者では薬使用が症状スコアを“押し下げる”方向に働くため、取得していないと結果解釈が崩れる
 - SAPにそのデータを「どう扱うか（集計方法・欠測の扱い・解析の位置づけ）」まで書いておくのが実務的に重要である
- 別紙2が求めている評価指標は、「鼻目症状」と「日

常生活の支障度（QOL）」のセットであり、QOLは症状変化に伴うことが前提である。したがって、設計の基本は主要評価項目を鼻目症状の総合スコアに一本化し、QOLはその改善が症状改善と連動していることを示す重要な副次評価として位置づけることになる。

また、軽症者を含む場合はレスキュー薬（治療薬）の使用が結果の解釈を左右するため、使用状況データの取得と取り扱いをSAPにて明確化しておくことが不可欠である。主要のブレ、多重性の曖昧さ、QOL単独の有意、レスキュー薬の未定義——これらは審査で指摘されやすい典型的な落とし穴であり、事前の設計で回避するとした場合に最もコストが低い対応となる。

3-4. 解析

「鼻目のアレルギー反応」領域のデータは、日々の症状スコア（連続～順序尺度）が中心で、しかも欠測が出やすいという実務的に厄介な条件が揃っている。ここで解析手法の選び方と、SAPへの書き方が曖昧だと、結果そのものより先に「設計・解析の筋」が問われる。この章では、審査で説明しやすく、かつ統計的にもブレにくい解析パッケージを提示する。

3-4-1. 推奨する解析手法

A. 反復測定混合モデル（MMRM）

日々または週次の症状スコアを縦断データとして扱う場合の標準的な手法である。

$$\text{モデル式: } Y_{ij} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \beta X_{i0} + \delta S_i + \varepsilon_{ij} \quad (1)$$

- Y_{ij} : 試験参加者 i の時点 j における症状スコア
- τ_i : 群効果（介入 vs プラセボ）
- γ_j : 時点効果
- $(\tau\gamma)_{ij}$: 群 × 時点
- X_{i0} : ベースライン（共変量）
- S_i : 層（健常 / 軽症）
- ε_{ij} : 誤差（同一試験参加者内の相関を考慮）

共分散構造は、試験規模と測定頻度で選ぶ。

- 非構造（Unstructured）：柔軟だがパラメータ数が多い。中規模以上向け。

- 1次自己回帰構造（AR（1））：近い時点ほど相関が高いと仮定。日誌データと相性が良い。
- 複合対称（CS）：最も単純で説明しやすいが、現実合わないことも多い。

実務の安全策としては、SAPに「第一選択は非構造、収束しない場合はAR（1）に切替」と明記しておくのが無難である。こうしておく、解析時のトラブル（収束不良）を後出し変更と見られにくくなる。

B. 共分散分析（analysis of covariance; ANCOVA）

もう一つの有力な選択肢が、主要評価期間をあらかじめ定め、日々の症状スコアを要約値（平均またはAUCなど）に落とし込んだ上で、ベースライン調整付きで群間比較するANCOVAである。

MMRMは縦断全体を同時にモデル化する一方、ANCOVAは主要期間の要約値に集約して1回の比較に落とし込むため、期間平均差が明確で、結果の提示が単純化される。

$$\text{モデル式: } Y_i = \mu + \tau_i + \beta X_{i0} + \delta S_i + \eta C_i + \varepsilon_i \quad (2)$$

- Y_i : 主要評価期間における平均スコア
- τ_i : 群効果（介入 vs プラセボ）
- X_{i0} : ベースラインスコア（共変量）
- S_i : 健常者 / 軽症者の区分（層別割付を行った場合は調整する）
- C_i : その他の層別因子（例：施設 / 地域、必要に応じて）
- ε_i : 誤差項

実務上は、MMRMと同様に層別割付に用いた因子（少なくとも健常 / 軽症、必要なら施設 / 地域）を共変量として含めると、設計と解析の整合性が保たれ、後から「解析が恣意的」と見られにくくなる。

3-4-2. 飛散量の取り扱い

別紙2は「一部時点のみ有意でも、考察が適切なら可」という柔軟性を残しているが、統計の観点では、考察で後追いするよりも事前に計画しておく方が良い。飛散量は、解析でいじるほど説明が難しくなる。

設計段階での対策（推奨）：

- 主要解析期間の事前定義：過去の飛散データを根

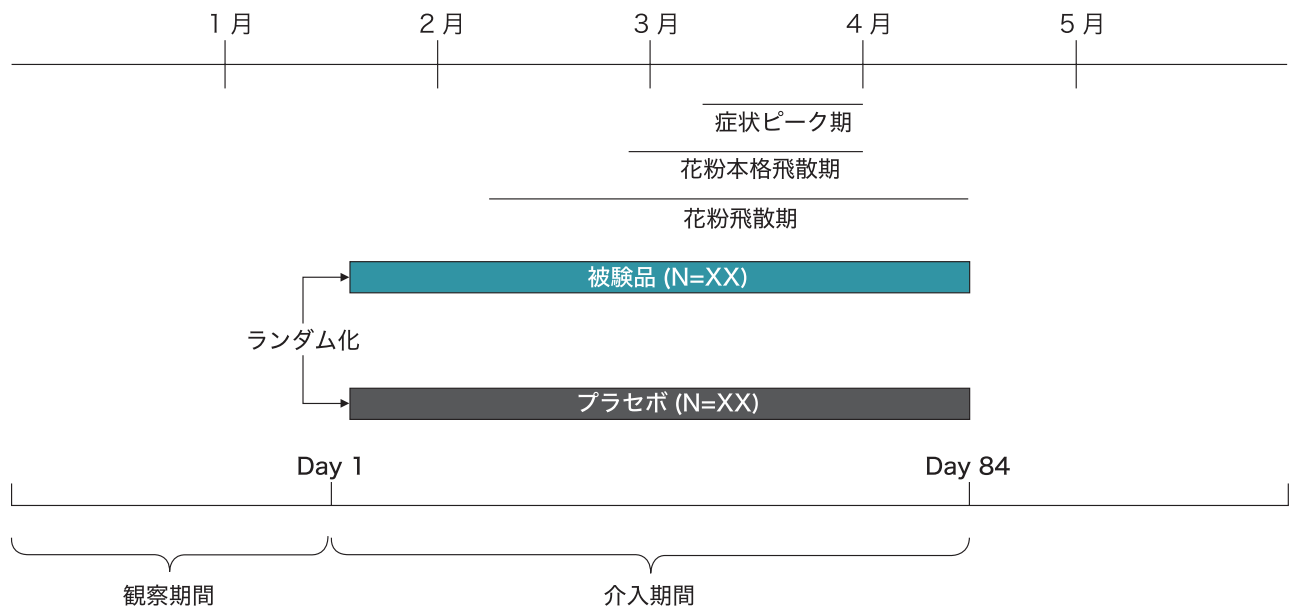


図2 鼻目アレルギー試験における試験期間と主要評価期間（ピーク）の位置づけ（例）

抛に、ピーク期間をプロトコルに記載

- 地域の限定：同一気象圏に絞るほど、ばらつきが抑えられる
- 地域（施設）で層別割付：広域実施なら地域を層別因子として設計に組み込む

解析段階での対策：

飛散量を日単位の時変共変量としてモデルに入れること自体は可能であるが、解釈が複雑化しやすく、基本は推奨しない。それよりも実務的には、「飛散量が一定以上の日に限定した感度分析」をSAPに規定しておく方が、説明可能性が高い（例：主要解析は全期間、感度分析として「高飛散日限定」を追加など）。

図2は、飛散期の年変動に対応するために、介入期間を一定の幅で確保しつつ、主要評価は「症状ピーク期（例：連続して飛散量が閾値以上の期間）」など事前定義された時間窓で行う設計を示す。主要評価期間を先に固定しておくことで、時点選択の恣意性や多重性の問題を抑え、解析（MMRM/ANCOVA）の解釈を一貫させやすい。

3-4-3. 欠測データの取り扱い

日誌データは、入力忘れや体調不良、外出の有無などで欠測が起きやすく、欠測の扱い次第で結論が簡単に揺れる。したがって欠測対応は「解析の付録」ではなく、試験計画の段階から骨格として組み込む

でおくべき論点である。

主解析の設計としては、MMRMとANCOVAで考え方が少し異なる。MMRMは、MAR（Missing at Random）仮定のもとで欠測をモデルの枠組みに取り込み、観測されたデータから推定を行うため、日々の縦断データと相性が良い手法である。一方、主要期間の要約値（平均やAUC）で評価するANCOVAでは、要約値の計算そのものが欠測に影響される。したがって「主要評価期間内の欠測をどこまで許容するか」（例：欠測が一定割合を超える場合は解析対象外など）を、プロトコルやSAPで事前に固定しておく必要がある。ここが曖昧だと、要約値の作り方が後付けに見えやすく、結果の信頼性が落ちる。欠測が一定程度以上に増える可能性がある領域では、感度分析は「念のため」ではなく、結論が欠測の仮定や扱いに過度に依存していないことを示すための必須の補強になる。まず最終観察値繰り越し（Last Observation Carried Forward; LOCF）法は前提が強く、推定を歪める可能性もあるため主解析に用いるべきではないが、欠測の扱いを変えたときに結論がどの程度動くかを示す参照ケースとして、位置づけを明確にした上で併記されることがある。欠測メカニズムが複雑で、単純な補完では説明が難しい場合には、ベースラインや過去の経過などを用いて欠測値を補完するMultiple Imputationを感度分析の選択肢として検討する。さらに、Tipping Point Analysisにより、介入群の欠測を不利な方向に段階

的に動かした場合でも結論が維持されるか（どの程度の仮定変更で結論が反転するか）を確認しておく、欠測に対する頑健性をより直接的に示すことができる。

第三者から指摘されやすいのは…

- 欠測の取り扱いがSAPに明記されていない
- 欠測が多い（目安として20%以上）にもかかわらず感度分析がない
- そして軽症者を含む設計で特に致命的になりやすいのが、レスキュー薬使用後の症状スコアをどう扱うかが不明なケース

使用後のスコアをそのまま用いるのか、欠測扱いにするのか、あるいはフラグを立てて別途解析するのか——この扱いが曖昧なままだと、「軽症者を含めて全体で検証する」という設計の整合性そのものが疑われやすくなる。

3-5. 運用

別紙2の要件は、研究デザインの理想論というより運用に強く依存する実務要件である。特に重要なのは、健常者と軽症者をまとめて全体解析する場合、健常者が概ね半数以上であることが条件になる点である。さらに「健常者の割合が不明な場合は、スコア等から推測できる合理的理由が必要」とされており、記録が不十分だと結果の良し悪し以前に「定義と証拠の不足」で問題が生じるリスクがある。つまりこの領域では、設計や解析よりも前に、スクリーニングと割付で決まると言ってよい。

3-5-1. 健常者と軽症者の定義と割付

実務上の最適解は、最初から「健常 / 軽症を確定できる定義」をプロトコルに書き、割付の層に組み込むことである。後から「だいたい半分くらい健常者だった」と説明するのは根拠不十分で、別紙2の要件充足（割合不明時の合理的推測）を自ら難しくしてしまう。

定義のコツは、「時々」という曖昧な用語を回数で定義することである。たとえば、

- 健常者：過去4週間に花粉症治療薬を使用していない
- 軽症者：過去4週間に花粉症治療薬を時々使用（例：週2回以下）

といった形で明文化し、スクリーニング時に質問票・記録で確認する。ここが残っていないと、健常者比率を推測で説明することになり、説明の難易度が一気に上がる。

割付は、健常 / 軽症を層別因子とした層別ブロックランダム化が基本である。加えて、健常者比率の条件を確実に満たすために、募集段階で健常者枠を優先する（例：目標症例数が80名なら、健常者50名・軽症者30名を目標に募集）。この運用を取っておけば、別紙2の要件は「解析で工夫して満たす」のではなく、運用で自然に満たす形になる。

3-5-2. 解析の層別

解析の立て付けはシンプルが安全である。主解析は全体（健常 + 軽症）で行い、健常のみ・軽症のみは事前定義したサブグループ解析として添える。群 × 層（健常 / 軽症）の交互作用も確認はするが、この手の交互作用は検出力が不足しやすいので、有意 / 非有意で結論を作るより、点推定と信頼区間で方向性を議論する整理が実務的である。

重要なのは、層別解析を後から都合よく採用しないことである。別紙2が軽症者データを例外的に認める枠組みである以上、解析方針の透明性 = 信頼性になる。

3-5-3. 第三者から指摘されやすいパターン

- 運用面の典型的な指摘をまとめると次の3つになる。
 - 全体で有意でないため、後付けで健常者のみ解析を主要結論にする
 - 健常者と軽症者の比率が不明（スクリーニング記録が残っていない）
 - 軽症者の定義が曖昧（「時々」の回数定義がない）
- 結局のところ、この領域の運用は「解析の工夫」で救うより、定義・記録・割付を最初から固めておくことが最短ルートである。

4. 問題点を考える

ここまで、別紙2から読み取れる内容から最適な方法論を示してきた。しかし、別紙2の記述には柔軟性がある一方、解釈を誤ると後出しに見えやすいポイントも散りばめられている。

この章では、第三者からの指摘につながりやすい

点を、何が問題で、どう先回りすべきかを整理する。

4-1. 「摂取期間を設定しない」は、放置しない

別紙2は摂取期間を明示的に規定していない。この文言は、花粉飛散期の年変動に対応するための柔軟性として理解するのが自然である。ただし、そのまま「期間を決めなくてよい」と解釈してしまうと、主要評価の時間軸が曖昧になり、再現性と比較可能性が落ちる。「鼻目のアレルギー反応」領域は季節とピークが重要であるため、「期間を決めない＝主要評価が決まらない」と直結しやすいからである。実務的には、摂取期間と主要評価期間を次のように運用可能な形で定義しておく安全である。たとえば摂取期間は「飛散開始予測日の1週間前から飛散終了まで(最長12週間)」のように上限付きで置き、主要評価期間は「飛散量が一定閾値以上の日が連続した2週間」など、飛散量と連動して事前定義する。こうしてプロトコルとSAPに書いておくことで、摂取期間を設定しないという別紙2の趣旨(柔軟性)と矛盾せず、むしろ科学的に筋の通った説明になる。

指摘されやすい例は…

- 摂取期間が「飛散期全体」とだけ書かれている
 - 主要評価期間の選択根拠がない
- といったケースである。

4-2. 「一部ポイントのみ有意でも可」は、p-hacking 疑念と表裏一体

別紙2は「一部の測定ポイントのみで有意差が付いた場合でも、適切に考察されていればデータとして使用できる」としている。これは実務的な救済条項にも読み取れてしまうが、統計学的には多重性の問題を含む。時点を増やせば、いずれかの時点で有意になる確率が上昇し、「飛散量が低かった時点は除外」といった後付け説明も可能としてしまう。つまり、柔軟性に見える面はそのまま恣意性に見えるリスクでもある。

このリスクを抑えるため、「救済条項」に頼らなくて済む設計を先に作るのが本筋である。具体的には、主要評価時点(または主要時間軸)を1つに絞り、その他の時点は副次的・探索的に扱う。どうしても複数時点を正式に検定する必要性が生じのであれば、階層的検定(順序を事前定義し、有意でなくなっ

た時点で停止)をSAPに書いておく。また飛散量については、主解析期間の定義を動かすのではなく、「高飛散日限定」のような感度分析を事前規定し、頑健性の確認として位置づける方が説明しやすい。

指摘されやすい例は…

- 複数時点で検定して多重性調整がない
 - 事後的に有意でない時点を「飛散量が低かったから」と外している
 - プロトコルやSAPにない解析が主要解析として報告されている
- などである。

4-3. 「一部の指標でも可」は、主要評価がブレやすい

評価指標についても、別紙2には「一部の指標でも機能が確認できれば使用できる」という柔軟な表現がある。これも救済のように見えるが、主要評価項目が複数あると「どれを採用するか」の選択が恣意的に見えやすく、届出後に「なぜこの指標なのか」という問いが来やすくなる。

実務対応はシンプルで、主要評価項目は一本化(例:TNSSとTOSSの合計)し、QOLは症状変化に伴うことを示す副次的な位置づけに限定する。ストーリーとしては「症状スコアが改善し、それに伴ってQOLも改善した」という形を描けるのが最も自然である。

どうしても一本化が難しい場合は、(1) co-primaryとして両方で有意が必要という保守的設計にする、または(2) gatekeepingなどの閉手順で多重性を制御する、という先にルールを設定する方向で整理する。

指摘されやすい例は…

- 主要評価項目が2つあるにもかかわらず、1つだけが有意な状態で機能性を主張する
 - QOLのみ有意で症状との連動性が示されていない
- などである。

4-4. 軽症者の「時々使用」は、曝露と行動の差が混ざる

「軽症者」を「治療薬を時々使用」と定義すると、薬の影響だけでなく、回避行動(マスク・外出控え)によって曝露自体が低い可能性も出てくる。さらにレスキュー薬(治療薬)で症状がマスクされ、群間

差が縮むことも起こり得る。そして最も厄介なのが、レスキュー薬（治療薬）使用後のスコアをどう扱うかである。ここを曖昧にしたままだと、解析の前提が崩れる。

実務上は、レスキュー薬（治療薬）の扱いをプロトコルや SAP で選択しておく必要がある。たとえば、使用日のスコアをそのまま使えば現実に近い一方、薬効で抑えられた値を評価に使うことになる。使用日のスコアを欠測扱いにすれば純粋な機能評価に寄る一方、欠測が増える。あるいは症状スコアと薬使用回数の複合指標にする方法もあるが、重み付けを含めて事前定義が必須で、解釈はやや複雑である。

曝露の標準化（飛散量データから曝露日数を共変量にする等）は技術的には可能だが、説明が難しくなるため、主解析ではなく感度分析の位置づけが無難である。

指摘されやすい例は…

- レスキュー薬使用後スコアの扱いが SAP にない
- 軽症者だが薬使用が多く症状スコアが実質的に読めない
- 健常 / 軽症で効果の方向が逆転しているのに全体の結果のみ報告

などである。

4-5. 単盲検の容認

症状スコアも QOL も主観評価である以上、盲検性は効果推定に直結する。単盲検が許容されとしても、試験参加者が割付を推測できる状況では、プラセボ効果や期待効果が群間差を歪める可能性がある。

二重盲検が困難な場合は、せめて盲検性の穴を埋める工夫が必要である。たとえば PROBE（割付はオープンでも、評価・集計は盲検）に近い運用や、電子日誌の導入による自動収集によって評価者の介入を最小化する。あるいは「自分がどちらの群だと思うか」を途中で尋ね、推測の偏りを間接的に示す。客観指標（鼻腔通気度、鼻汁好酸球など）を副次として併用するのも一案だが、症状との相関は常に高いとは限らないため、主役に据えないような整理が必要である。

指摘されやすい例は…

- 単盲検の理由が「プラセボが難しかったから」というだけで、盲検維持の工夫がない
- 群推測の正答率が極端に高い（実質オープン）
- 主観指標のみで効果が非現実的に大きいなどである。

4-6. 「他領域の疾病域データ OK」は、外的妥当性と安全性の説明が残る

「論文上、健常者として扱われていれば、他領域の検査値が疾病域でも使用できる」という整理は、実務上は助かる。ただし疾病域の試験参加者が多いと「本当に非疾病者か？」という疑義が生まれやすく、外的妥当性と安全性の説明が求められる（例：高血圧で治療中の試験参加者が花粉症試験に参加している場合など）。

ここでは透明性が重要で、除外すべき疾病（重大疾患、評価に影響する疾患）を明示し、他領域の疾病を持つ試験参加者の人数と内訳を報告書に記載し、必要ならその試験参加者を除いた解析も併記する。安全性についても、当該試験参加者で有害事象が偏る場合は追加確認が必要になり得る、という整理が現実的である。

指摘されやすい例は…

- 疾病を持つ試験参加者の割合が不明（いないなら 0 と明記する）
 - 疾病域の定義が曖昧
- などである。

まとめ

「鼻目のアレルギー反応」領域は、症状が季節・飛散量・生活環境に左右され、評価が主観指標中心で欠測も生じやすいという、設計・解析の難所が揃っている。別紙 2 が提示する柔軟な許容（軽症者の混在、時点や指標の一部での有意など）は、現場を助ける一方で、解釈を誤ると多重性や p-hacking への疑念につながり得る。

だからこそ、この領域で重要なのは「高度な解析」よりも、事前定義で恣意性を排除する設計である。具体的には、主要評価項目を症状の総合スコアに一本化し、QOL は“症状改善に伴う”ことを補強する重要な副次評価項目として位置づける。主要

の時点（または主要期間）も1つに絞り、その他は副次・探索に回す。飛散量の取り扱いは事前に計画し、どうしても不確実性が残る部分は感度分析としてSAPにあらかじめ規定する。

さらに、全体解析を行うなら健常者比率 $\geq$ 約50%

を運用で担保し、健常/軽症の定義（「時々」を回数で定義）とスクリーニング記録を残すことが不可欠である。結局のところ、審査で通りやすい試験とは「結果が良い試験」ではなく、主要な結論へ至る筋道が一貫している試験である。

参考文献

1. 消費者庁. 機能性表示食品の届出等に関する手引き [Internet]. 東京: 消費者庁; 2025年3月25日(消食表第273号). 2025年10月1日一部改正(消食表第711号) [2026年8月6日引用]. Available from: https://www.caa.go.jp/policies/policy/food_labeling/foods_with_function_claims/notice/assets/food_labeling_cms205_251001_41.pdf