

# MDQ における Pain を 主要アウトカムにした試験設計の考え方

馬場 亜沙美 (BABA Asami)<sup>1\*</sup>, 鈴木 直子 (SUZUKI Naoko)<sup>1</sup>,  
田中 瑞穂 (TANAKA Mizuho)<sup>1</sup>, 山本 和雄 (YAMAMOTO Kazuo)<sup>1</sup>

Key Words : ヒト臨床試験, 特定保健用食品, 機能性表示食品, MDQ, Pain, サンプルサイズ, 試験設計

## Considerations for Trial Design with the MDQ Pain Subscale as the Primary Outcome

**Authors:** Asami Baba<sup>1\*</sup>, Naoko Suzuki<sup>1</sup>, Mizuho Tanaka<sup>1</sup>, Kazuo Yamamoto<sup>1</sup>

**\*Correspondence author:** Asami Baba

Keywords: Human clinical trials, Foods for Specified Health Uses (FOSHU), Foods with Function Claims (FFC), MDQ, Pain, Sample size, Trial Design

**Affiliated institution:**

<sup>1</sup> ORTHOMEDICO Inc.

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

### 1. はじめに

Menstrual Distress Questionnaire (MDQ) は、月経関連症状を評価する尺度として広く用いられている多次元尺度である<sup>1)</sup>。MDQ は以下の 8 つのサブスケールから構成される。

- Pain
- Water Retention
- Autonomic Reactions
- Negative Affect
- Impaired Concentration
- Behavior Change
- Arousal
- Control

MDQ は月経関連症状の包括的な評価に有用である一方、ヒト臨床試験 (ヒト試験) においてどの評

価項目を主要アウトカムとして採用するか、どのように対象集団を定義するか、そしてサンプルサイズをどう設計するかは、試験の妥当性を担保する上で重要な課題となる。

そこで本稿では、MDQ のサブスケールである「Pain」を主要アウトカムとした場合において、機能性表示食品の科学的根拠として一般化可能性の高い試験を設計するための方法論的考え方を体系的に整理する。

### 2. Pain を主要アウトカムとして選択する根拠

#### 2-1. MDQ における評価時期の特性

MDQ は月経関連症状を月経前 (Premenstrual)、月経期 (Menstrual)、月経間期 (Intermenstrual) の 3 つの時期に分けて評価する点に特徴がある。それぞれの時期の特性は以下のとおりである。

\* 責任著者 : 馬場 亜沙美 (Asami Baba)

所属機関 :

<sup>1</sup> 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

### ①月経前 (Premenstrual)

月経開始前に出現・増悪する症状を捉える時期であり、いわゆる月経前症候群 (PMS) に関連する症状が反映されやすい。一方で、症状の出現様式や強さには個人差が大きく、データにばらつきが生じやすい。

### ②月経期 (Menstrual)

月経痛や下腹部痛、腰痛など、痛みに関連する症状が最も強く現れやすい時期である。試験参加者自身も変化を最も実感しやすく、臨床的な解釈が比較的明確である。

### ③月経間期 (Intermenstrual)

症状が最も軽減し、日常状態に近い時期である。個人内のベースラインとしての意味を持つ一方で、「Pain」に関しては床効果 (floor effect) が強く、主要アウトカムとしては変化を捉えにくい。

MDQ を用いたヒト試験においては、これら3時期すべてを測定することで、症状の時間的推移や個人内変動を把握することが可能となる。そのため、3時期すべての測定を実施することが望ましい。一方で、主要アウトカムとしてどの時期・どのサブスケールを用いるかについては、別途慎重な判断が必要である。

## 2-2. Pain を選択する根拠

MDQ には複数のサブスケールが存在するが、本稿では「Pain」を主要アウトカムとして位置づける。その理由は以下のとおりである。

- 症状概念が単一であり、解釈が明確である
- 月経関連症状の中で、臨床的および当事者にとって最も重要度が高い
- 介入による変化を試験参加者自身が実感しやすい  
さらに、評価時期としては月経期 (Menstrual) の「Pain」に着目することが最も合理的と考えられる。月経期 (Menstrual) の「Pain」は、症状の強度が最

大となりやすく、介入による改善を捉えやすい。同時に、月経前 (Premenstrual) に比べてデータが過度にばらつきにくく、月経間期 (Intermenstrual) に比べて床効果 (floor effect) の影響も受けにくいという統計学的な利点がある。

複数のサブスケールや複数の時期を同時に主要アウトカムとして設定すると、検定の多重性の問題や結果解釈の複雑化を招きやすい。そのため、主要アウトカムを明確に1つに絞り、その他を副次的・探索的な位置づけとする整理は、ヒト試験において現実的かつ合理的な選択肢といえる (表1)。

なお、本稿の試験設計において「Pain」を主要アウトカムとして選択していることは、機能的表示食品の届出表示 (ヘルスクレーム) として「痛みの緩和」が直接的に標榜可能であることを意味するものではない。あくまで試験の妥当性を担保するための、方法論的検討であることに留意されたい。

## 3. 機能的表示食品向けの対象集団の定義

### 3-1. 健常者定義の課題

機能的表示食品のヒト試験を計画する際は、原則として健常者 (軽症域や境界域を含む) を対象とする必要がある。しかし、MDQ のサブスケールには軽症・中等症・重症といった一義的な基準 (カットオフ値) が設けられていない。そのため、健常者 (軽症域や境界域を含む) の選抜基準を決定することは実務上の課題となる。そこで本稿では、既定のカットオフ値に依存しない方法として、MDQ の回答スケール自体の意味合いを手がかりに対象集団を定義するアプローチを提案する。

### 3-2. MDQ の回答スケールを用いた対象集団の定義

MDQ の各質問項目は、以下の5段階で評価される。

- 0: No experience of symptom (症状なし)
- 1: Present, mild (軽度)
- 2: Present, moderate (中等度)
- 3: Present, strong (強い)
- 4: Present, severe (重度)

表1 評価項目の構成と分類

分類	評価項目
主要アウトカム	月経期 (Menstrual) における MDQ の Pain スコア
副次的アウトカム	月経前 (Premenstrual) および月経間期 (Intermenstrual) における MDQ の Pain スコア
探索的アウトカム	その他の MDQ サブスケール (Water Retention, Negative Affect 等) のスコア

Pain スコアは、「Pain」に関する6つの質問項目の合計得点であり、得点範囲は0~24点である。各質問項目の回答が0で「症状なし」、1で「軽度」であることを考慮し、対象者を表2のように定義した。

表2 MDQのPainスコアの判定基準

合計得点	平均得点	解釈
0~2点	0.00~0.33	無症状と考えられる
3~11点	0.50~1.83	軽度(境界域)と考えられる
12点以上	2.00以上	中等度以上と考えられる

### 3-3. 下限値・上限値の設定根拠

#### ● 下限値の設定根拠

合計0~2点はほぼ無症状に近く、介入による改善余地が極めて限定的である。このような試験参加者を含めた場合、床効果 (floor effect) により群間差の検出が困難となるリスクがある。合計3点は、複数項目で軽度の症状が認められる「症状あり」と言える最小レベルであり、改善余地の確保と統計的な感度を両立する下限として合理的である。

なお、合計6点(平均得点1.00点)を下限とする案も検討したが、この場合、症状が明確な層に限定されるため、軽度症状者の一部が除外され、結果の一般化可能性(外挿性)が限定される懸念がある。したがって本稿では、床効果 (floor effect) を回避しつつ、一般女性集団への広範な適用性を確保する目的で、下限を3点とした。

#### ● 上限値の設定根拠

合計12点以上は項目平均で2.00点に相当し、複数項目で中等度以上の症状が持続している状態を指す。これは軽症の範囲を超え、中等症の領域に入り始めていると推測されるため、機能性表示食品の対象である「健常者(境界域)」を逸脱する可能性が高まる。

以上の根拠から、Painスコアが3点以上11点以下の集団を選抜することが、試験の妥当性と一般化可能性を確保する上で妥当であると考えられる。また、留意事項として、月経間期 (Intermenstrual) から月経前期 (Premenstrual) あるいは月経期 (Menstrual) にかけての症状の変動が30%以上に及ぶ場合、一過性の月経関連症状の範囲を超える可能性が指摘されているため<sup>1)</sup>、選抜条件の設定に際し

ではこの変動率も考慮することが望ましい。

## 4. サンプルサイズの計算

### 4-1. 標準偏差の推定

Painスコアが3~11点に限定した集団における標準偏差 (SD) を推定するため、R (Ver. 4.5.2) を用いて、3種の分布仮定に基づくシミュレーションを実施した(各仮定につき n = 100,000)。

#### 仮定1: 一様分布 (3~11点) が等確率で出現)

Painスコア3~11の各値がすべて等確率で出現すると仮定した。これは症状の強さに特定の偏りがなく、端の値(3や11)にも中間値(6や7)と同程度の試験参加者が存在する状況に相当する。この仮定のもとでシミュレーションしたコードは以下のとおりである。

```
set.seed(123)
n <- 100000
x_uniform <- sample(3:11, n, replace=TRUE)
sd(x_uniform)
```

シミュレーションの結果、SDは約2.6と推計された。以上より、この仮定を採用する場合、SDを2.5程度と見積もることが妥当と考えられる。

#### 仮定2: 二点分布 (下限3点と上限11点のみに集中)

試験参加者が下限(3点)または上限(11点)のいずれかのみ分布するという、数学的に分散が最大となる極端な状況を想定した。この仮定のもとでシミュレーションしたコードは以下のとおりである。

```
set.seed(123)
n <- 100000
x_extreme <- c(rep(3, n/2), rep(11, n/2))
sd(x_extreme)
```

シミュレーションの結果、SDは4.0となり、これは取りうる範囲が3~11点に固定された場合の理論的最大値に相当する。

#### 仮定3: 弱いU字型分布 (下限寄りと上限寄りがやや多い)

実際の軽症集団においては、症状が軽微な下限寄りの層(3~4点)と、中等症に近い上限寄りの層

(10~11点)が一定数存在する一方で、中央値付近(6~8点)は症状の進行や改善の過程で通過しやすく、相対的に頻度が低くなる可能性がある。この特性を反映した弱いU字型分布を仮定した。この仮定をもとにシミュレーションしたコードは以下のとおりである。

```
set.seed(123)
n <- 100000
values <- 3:11
prob_raw <- c(0.16,0.08,0.06,0.05,0.04,0.05,0.06,0.08,0.16)
prob_sd3 <- prob_raw / sum(prob_raw)
x_sd3 <- sample(values, n, replace = TRUE, prob = prob_sd3)
sd(x_sd3)
```

シミュレーションの結果、SDは約3.1と推計された。丸め処理を考慮し、この仮定におけるSDは3.0と見積もることが妥当と考えられる。

#### 4-2. 標準偏差の選択

以上の結果から、現実的な分布を踏まえつつやや保守的に見積もった値として、SD=3.0を採用した。この値は、一様分布仮定(SD=2.5)より大きく、非現実的な二点分布(SD=4.0)より小さい中間値であり、恣意的に選択したのではなく、分布仮定から自然に導かれる保守的推定値として合理性を有する(表3)。

#### 4-3. 必要例数の算出と群間差の検討

##### ●前提条件

例数計算は、2群比較の並行群間試験デザインを前提とし、両側検定、有意水準 $\alpha = 0.05$ 、検出力90% ( $1 - \beta = 0.90$ )の条件を設定した。検出対象とする群間差は1.0~3.0点の範囲を0.5点刻みで変化させ、SDについては4.1節で導出した3つの仮定

表3 各分布仮定における標準偏差(SD)の推定値

分布仮定	標準偏差(SD)	特徴
一様分布	2.5	下限的な推定値
弱いU字型分布	3.0	現実的な中間的推定値
二点分布	4.0	理論的上限(非現実的)

値(2.5, 3.0, 4.0)をそれぞれ用いた。

##### ●Rコード(pwrパッケージ使用)

以下のRコード(pwrパッケージ)により、標準偏差と群間差の組み合わせごとに1群あたりの必要例数を算出した。

```
library(pwr)
library(dplyr)
sd_vec <- c(2.5, 3.0, 4.0)
delta_vec <- seq(1.0, 3.0, by = 0.5)

res <- expand.grid(SD=sd_vec, Delta=delta_vec)%>%
  mutate(effect_size = Delta/SD,
         n_per_group = ceiling(mapply(function(d){
           pwr.t.test(d=d, sig.level=0.05, power=0.90,
            type='two.sample')$n
         }, effect_size)))

res
```

シミュレーションの結果、SDの仮定がサンプルサイズに与える影響は、群間差が小さい領域(1.0~1.5点)において特に顕著であることが確認された。例えば、群間差1.0点の条件下では、SD=2.5とSD=4.0で必要例数に大きな乖離が生じる。一方、群間差が2.0点以上になると、SDの差による影響は相対的に減弱し、設計が安定する傾向がみられた(図1)。

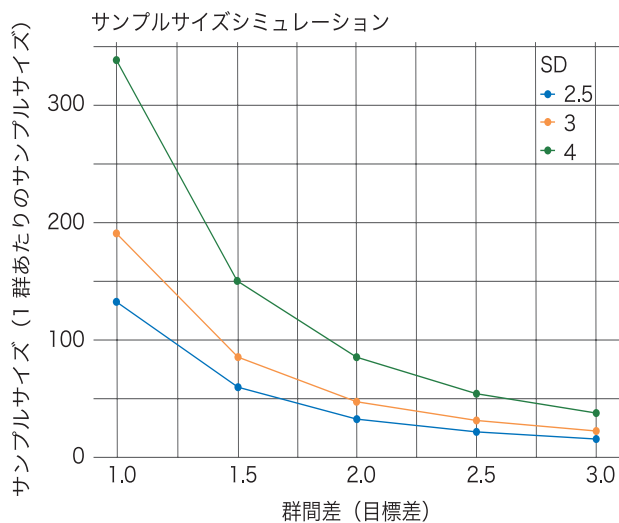


図1 群間差および標準偏差の仮定が必要例数に与える影響

●群間差と臨床的意義の解釈

「Pain」における Minimally Important Difference (MID: 最小臨床的有意差) に関する国際的なコンセンサスは、現時点において確立されていない。そのため本稿では、尺度得点の構造と試験参加者の主観的体感という観点から、各群間差の臨床的意義を表4に整理した。

4-4. 推奨サンプルサイズ

以上の検討を総合し、「Pain」を主要アウトカムとする機能性表示食品のヒト臨床試験において、推奨されるサンプルサイズを表5に示した。

これらの設計は、対象集団の一般化可能性、群間差の臨床的意義、および実施可能な試験規模という3つの観点のバランスを考慮し、保守的なSD(3.0点)の仮定のもとで導出された。探索的試験・検証的試験のいずれのフェーズにおいても、その例数設定の根拠を論理的に説明可能であることが、試験デザインの妥当性を担保する上で極めて重要である。

5. まとめ

本稿では、MDQの「Pain」を主要アウトカムとした機能性表示食品向けヒト臨床試験の設計において、以下の3点を体系的に検討した。

- 評価時期・サブスケールの選択: 月経期(Menstrua)

の「Pain」を主要アウトカムとすることが、臨床的意義・検出感度・多重性回避の観点から最も合理的である。

- 対象集団の定義: Painスコア3~11点を選抜基準とすることで、床効果(floor effect)を回避しつつ、一般化可能性を確保できる。

- サンプルサイズの計算: SD = 3.0 (現実的かつ保守的仮定)のもと、探索的試験では群間差2.0点・1群50例、検証的試験では群間差1.5点・1群90例が推奨される。

サンプルサイズの計算は単なる計算問題ではなく、対象集団の定義、臨床的意義のある変化の定量化、および検出の確実性という一連の判断の積み重ねである。また、例数計算によって得られる目標症例数は「必ず有意差を出すための人数」ではなく、「想定した大きさの介入効果が本当に存在するかどうかを、統計学的に十分な確実性をもって判断できる人数」を意味する。

標準偏差と群間差を系統的に動かしながら検討することで、「なぜこのサンプルサイズなのか」を論理的に説明できる強固な試験設計に近づく。本稿で提示した方法論的アプローチが、MDQを用いた機能性表示食品の臨床試験設計を検討する際の一助となれば幸いである。

表4 MDQのPainスコアにおける群間差の臨床的意義と実務上の評価

群間差	臨床的解釈	実務上の評価
1.0点	1項目1段階改善と同程度の差であり、複数項目での微小変化の積み重ねと区別困難である。	検証的試験の目標値として設定しにくい
1.5点	複数項目で軽度の改善が重なった状態であり、「痛みが和らいだ」と説明可能である。	境界域集団における最小臨床的有意差の候補
2.0点以上	複数側面で明確な改善が生じていると解釈可能である。	探索的試験において最も説明しやすい改善幅

表5 探索的試験・検証的試験における推奨サンプルサイズ

分類	標準偏差 (SD)	群間差 (目標差)	1群あたりのサンプルサイズ
探索的試験	3.0	2.0点	50例
検証的試験	3.0	1.5点	90例

参考文献

1. Moos, RH. (2010). Menstrual Distress Questionnaire (MDQ): Manual, instrument, and scoring guide (4th ed.). Mind Garden, Inc. (Original work published 1968; revised editions 1991, 2000)