



# 患者選好研究入門

日本製薬工業協会 医薬品評価委員会

2026 年度タスクフォース 9-5

2026 年 10 月

## 目次

|                                                         |    |
|---------------------------------------------------------|----|
| 1. はじめに .....                                           | 5  |
| 2. 患者選好研究概論.....                                        | 8  |
| 2.1. 定性的研究・定量的研究の使い分け/組み合わせ .....                       | 8  |
| 2.2. 医薬品開発における患者選好情報の活用目的 .....                         | 9  |
| 2.2.1. 医薬品のベネフィットとリスクの評価 .....                          | 10 |
| 2.2.2. 医薬品の Target Product Profile (TPP) やセグメントの検討..... | 12 |
| 2.2.3. 臨床試験デザインの最適化 .....                               | 12 |
| 2.2.4. リスク軽減策の検討 .....                                  | 12 |
| 2.2.5. 患者と医療関係者への情報提供 .....                             | 13 |
| 2.2.6. 医療技術評価 (HTA) や保険償還判断 .....                       | 13 |
| 2.2.7. 市販後の継続的な意思決定 .....                               | 14 |
| 2.3. 患者選好研究の時期と目的の関係.....                               | 14 |
| 3. 患者選好研究を実施する際の留意事項.....                               | 15 |
| 3.1. 倫理的枠組み.....                                        | 15 |
| 3.2. 外部専門家 (ステークホルダー) との早期からの連携 .....                   | 15 |
| 3.3. 倫理審査 (IRB/EC) .....                                | 16 |
| 3.4. インフォームド・コンセント .....                                | 16 |
| 3.5. プライバシー・データ保護.....                                  | 16 |
| 3.6. 研究の透明性確保 .....                                     | 17 |
| 3.7. 患者負担の最小化 .....                                     | 17 |
| 3.8. 患者以外の回答者の選択 .....                                  | 18 |
| 4. 定性的研究について .....                                      | 19 |
| 4.1. 定性的研究の目的と意義 .....                                  | 19 |
| 4.2. 定性的研究の進め方.....                                     | 19 |
| 4.3. 定性的研究の枠組み.....                                     | 20 |
| 4.3.1. サンプルング方針 .....                                   | 20 |
| 4.3.2. 研究参加者との主なコミュニケーション方法 .....                       | 22 |
| 4.3.3. データ収集の主な方法.....                                  | 23 |
| 4.3.4. データ分析の準備 .....                                   | 25 |
| 4.4. 定性的研究におけるデータハンドリングと解析.....                         | 26 |
| 4.4.1. コード化によるテキスト分析.....                               | 26 |
| 4.4.2. ソフトウェアの活用 .....                                  | 30 |
| 4.4.3. 定性的研究の解析手法.....                                  | 31 |
| 4.5. 成果物のイメージと医薬品開発への活用 .....                           | 36 |

|                                         |    |
|-----------------------------------------|----|
| 5. 定量的研究について .....                      | 49 |
| 5.1. 定量的研究の概要 .....                     | 49 |
| 5.1.1. 定量的研究の目的と意義 .....                | 49 |
| 5.1.2. 定量的研究の標準的な進め方 .....              | 49 |
| 5.2. 定量的研究実施の留意事項 .....                 | 50 |
| 5.2.1. リサーチクエスションの明確化 .....             | 50 |
| 5.2.2. 対象集団の定義・募集方法 .....               | 51 |
| 5.2.3. 質問票の設計 .....                     | 53 |
| 5.3. 各定量的方法の考え方・実施方法と使い分け .....         | 68 |
| 5.3.1. Discrete Choice Experiment ..... | 68 |
| 5.3.2. その他の手法の紹介と比較 .....               | 78 |
| 5.4. 選好が均一でない集団・部分集団の考え方 .....          | 84 |
| 5.5. 国や文化の違いの考え方 .....                  | 87 |
| 6. まとめ .....                            | 94 |
| 7. 参考文献 .....                           | 95 |

## 略語一覧

| 略語表記          | 英語表記                                                                                                | 日本語表記              |
|---------------|-----------------------------------------------------------------------------------------------------|--------------------|
| BR            | Benefit-Risk                                                                                        | ベネフィット・リスク         |
| BWS           | Best-Worst Scaling                                                                                  | -                  |
| COPD          | Chronic Obstructive Pulmonary Disease                                                               | 慢性閉塞性肺疾患           |
| COA           | Clinical Outcome Assessment                                                                         | 臨床アウトカム評価          |
| CAQDAS        | Computer Assisted/Aided Qualitative Data Analysis Software                                          | -                  |
| CRO           | Contract Research Organization                                                                      | 開発業務受託機関           |
| CIOMS         | Council for International Organizations of Medical Sciences                                         | 国際医学団体協議会          |
| DCE           | Discrete Choice Experiment                                                                          | 離散選択実験法            |
| EC            | Ethics Committee                                                                                    | 倫理委員会              |
| EMA           | European Medicines Agency                                                                           | 欧州医薬品庁             |
| EQ-5D-Y       | EuroQol-5 Dimensions, Youth Version                                                                 | -                  |
| FDA           | Food and Drug Administration                                                                        | 米国食品医薬品局           |
| GDPR          | General Data Protection Regulation                                                                  | EU 一般データ保護規則       |
| GCP           | Good Clinical Practice                                                                              | 医薬品の臨床試験の実施に関する基準  |
| GTA           | Grounded Theory Approach                                                                            | グラウンデッド・セオリー・アプローチ |
| HTA           | Health Technology Assessment                                                                        | 医療技術評価             |
| HRQoL         | Health-Related Quality of Life                                                                      | 健康に関する生活の質         |
| IMI<br>PREFER | Innovative Medicines Initiative Patient Preferences                                                 | -                  |
| IRB           | Institutional Review Board                                                                          | 倫理審査委員会            |
| ICH           | International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use | 医薬品規制調和国際会議        |

| 略語表記  | 英語表記                                                              | 日本語表記            |
|-------|-------------------------------------------------------------------|------------------|
| ISPOR | International Society for Pharmacoeconomics and Outcomes Research | 国際医薬経済・アウトカム研究学会 |
| jRCT  | Japan Registry of Clinical Trials                                 | 臨床研究等提出・公開システム   |
| MCMC  | Markov Chain Monte Carlo                                          | マルコフ連鎖モンテカルロ     |
| MAR   | Maximum Acceptable Risk                                           | 最大許容リスク          |
| MCID  | Minimal Clinically Important Difference                           | 臨床的に意味のある最小変化量   |
| MRB   | Minimum Required Benefit                                          | 最小必要ベネフィット       |
| MCDA  | Multi-Criteria Decision Analysis                                  | 多基準意思決定分析        |
| NLP   | Natural Language Processing                                       | 自然言語処理           |
| PED   | Patient Experience Data                                           | 患者経験情報           |
| PPI   | Patient Preference Information                                    | 患者選好情報           |
| PPS   | Patient Preference Study                                          | 患者選好研究           |
| PRO   | Patient Reported Outcome                                          | 患者報告アウトカム        |
| PFDD  | Patient-Focused Drug Development                                  | 患者中心の医薬品開発       |
| PMDA  | Pharmaceuticals and Medical Devices Agency                        | 医薬品医療機器総合機構      |
| QALY  | Quality-Adjusted Life Years                                       | 質調整生存年           |
| RPLM  | Random Parameters Logit Model                                     | ランダムパラメータロジットモデル |
| SE    | Standard Error                                                    | 標準誤差             |
| SW    | Swing Weighting                                                   | -                |
| TPP   | Target Product Profile                                            | 目標製品プロファイル       |
| TT    | Threshold Technique                                               | -                |
| UMIN  | University Hospital Medical Information Network                   | 大学病院医療情報ネットワーク   |

## 1. はじめに

近年、医療や医薬品を考える際に患者の経験に着目する流れが起きており（図 1）、特に近年では規制当局を中心に患者経験情報（Patient Experience Data、以下 PED）を意思決定に用いる考え方が広がりを見せている<sup>1-4)</sup>。（詳細は製薬協医薬品評価委員会データサイエンス部会の 2024 年度 TF6（以下、本 TF）作成の「Patient-Focused Drug Development への患者経験情報の活用 -患者の声に基づいた意思決定のために-」<sup>1)</sup>（以下、1 報目の報告書）を参照）

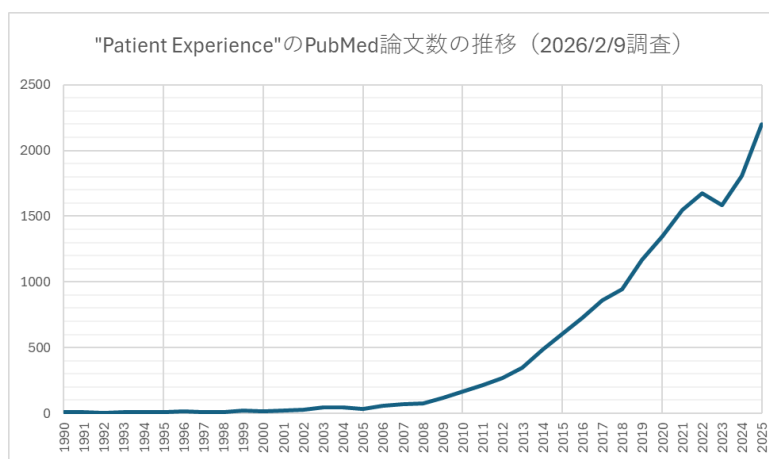


図 1 患者の経験に関する論文数の推移

PED は、臨床アウトカム評価や患者選好情報(Patient Preference Information、PPI。  
注：本書内では FDA に倣い PPI を Patient and Public Involvement の意味としては使用しない) などを含む概念であるが、その中でも患者選好についての研究が急速に増え始めており、近年重要な概念となりつつある。Nouwens らの研究(2025)<sup>5)</sup>によれば、代表的な選好研究手法である離散選択実験法（Discrete Choice Experiment、以下 DCE）を用いた患者選好研究（Patient Preference Study、PPS）は直近 5 年間（2018-2023）の論文数がそれ以前の論文数の合計の 2 倍以上となっている。また、直近 5 年間の 1279 本の DCE を扱った論文のうち 35 本は日本での研究となっており、国内でも DCE の実施が増えてきていることがわかる。

医薬品の開発において医師が重視する医薬品の特性についてはこれまでも検討され、キーオピニオンリーダーの声として開発計画に反映されていたが、実際には医師と患者で選好が異なることがしばしばあることが示されている<sup>6-7)</sup>。患者が好まない医薬品ではアドヒアランスが低下することと言われており<sup>8-9)</sup>、投薬が続けられないことで結果的に期待される有効性が得られなくなってしまうことを考えると、PPI は医薬品の価値判断に影響を与える重要な情報である。そういった重要性を踏まえ、PPI を得るための PPS については、新規 ICH ガイドラインである ICH E22「患者選好研究に関する

る一般指針」が 2026 年 7 月現在作成中であり、ドラフト版が公開されている<sup>10)</sup>。本ガイドラインは医薬品開発・承認申請での PPS の活用に焦点を当てている。

患者から選好情報を得ることは医薬品開発の様々なプロセスにとって重要だが、活用する目的によって必要とされる方法論の科学的厳格さは変わりうる。製薬企業が自社製品の改善のために患者に自由記述式アンケートを実施し、製品の良い点や改善すべき点を収集する場合もある。一方で、ある疾患領域におけるアンメットニーズの理解を深め論文化する目的で、患者にとっての特性の重要性を明確にできる DCE などのより体系的な手法を用いた研究をする場合もあるだろう。特に規制当局への新薬承認を検討する際の価値評価のエビデンスとして PPI を使う場合には、より厳格な方法論の適用が必要となる。患者からの情報は様々な方法で得ることができ様々な目的に使われている(図 2)<sup>1)</sup>。しかし、医薬品の科学的な価値評価に活用する場合(図 2 左上側)には、より厳格な方法論に従う必要がある。

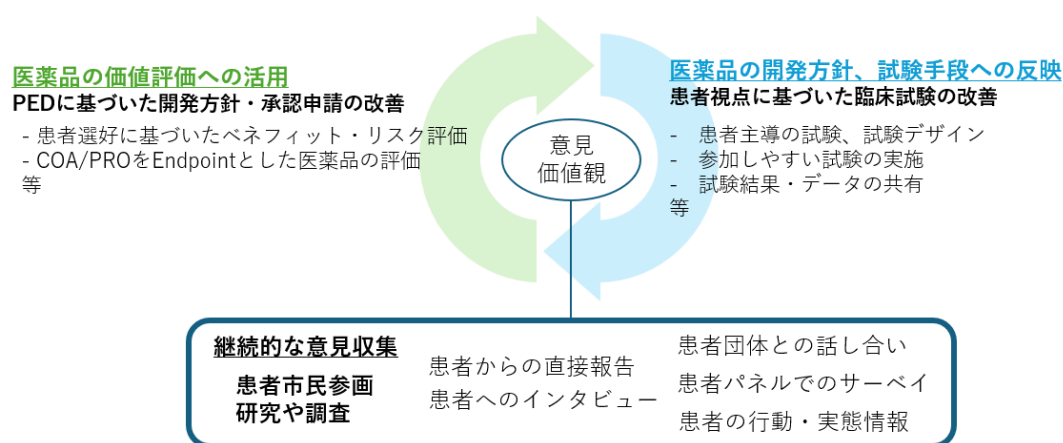


図 2 医薬品開発での PED の利用場面

製薬企業はこれまでも市場調査の目的で医師や患者へ選好を聞き取る調査は実施してきたが、承認申請など医薬品の価値を判断する意思決定のエビデンスとして使われることを意識した方法論に則った研究は限定的だった。2015 年ごろから海外規制当局及び関連組織からガイドラインが出始めた(図 3)が、国内に PPS に関する成書・報告書は、2026 年現在時点で 1 報目の報告書以外にはほとんど存在しない。また、実際に PPS を適切に立案・実行・報告するためには、より具体的な手順を理解する必要があり、上記の各組織が発出したガイドラインを読んでも詳細な手順までは書かれていないために、そのままでは実行できない難しさがある。

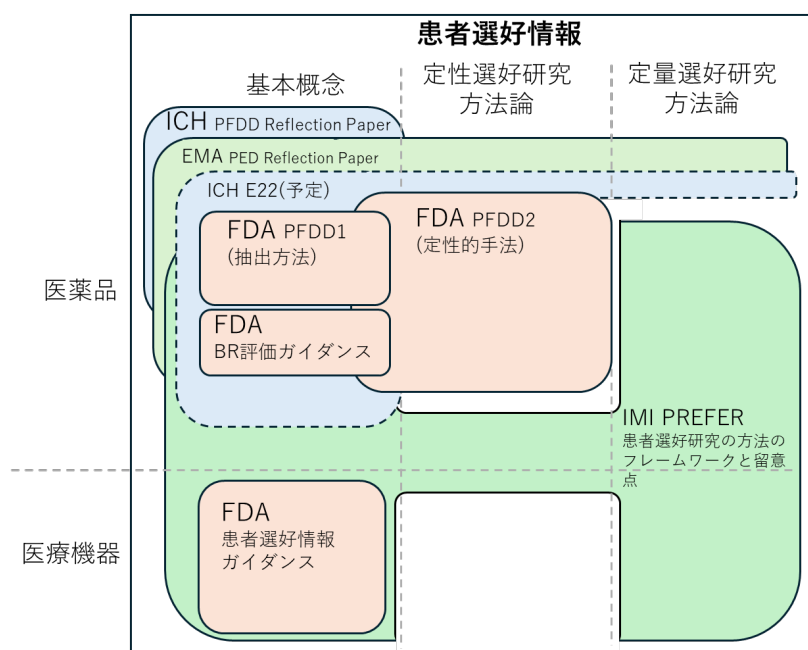


図 3 規制当局等から発出された患者選好情報に関連したガイダンスの記載範囲

そこで、必要に応じて実例に触れながら、本報告書を読むことで実際に PPS が実行できるようになること、また特にデータサイエンスの観点から PPS に関わる製薬業界の担当者が研究の立案から報告まで実行できるようになること、を目的として本報告書を作成した。PPS 自体は医薬品開発の目的に限らず、実際の医療現場で医療者と患者が話し合って治療方針を決めるシェアードデシジョンメイキング等での活用や、市場調査目的での簡易な調査として行われることもあるが、本報告書では医薬品開発における意思決定のエビデンスとして用いることを目的とした方法に焦点を当てて説明する。

本報告書では、1 章で作成の背景や記載内容について概説したのち、2 章では PPS の種類・目的・実施時期について整理し、3 章では倫理的な考慮事項等の全般的な留意点について説明する。また、4 章では定性的研究の特徴・方法の種類・手順について解説し、5 章では同様に定量的研究の特徴・方法の種類・手順について説明し、6 章はまとめとした。本報告書は主に FDA の PFDD ガイダンス 1・2<sup>11-12)</sup>、医療機器を対象とした PPI ガイダンス<sup>13)</sup>、BR 評価ガイダンス<sup>14)</sup>、EMA の PED に関するリフレクションペーパー<sup>15)</sup>、IMI PREFER の勧告書<sup>16)</sup>の記載に加え、実際に患者選好研究を行った本 TF メンバーの経験を踏まえ、記載を取りまとめている。そのため、本報告書の記載はすべての参考文献に完全に準拠した記載としていないわけではない。研究計画や手法選択においてより詳細な検討が必要な場合には、各参考文献の原文を併せて確認することもあるだろう。

## 2. 患者選好研究概論

本章では、定性的研究および定量的研究の位置づけと役割について概観するとともに、患者選好研究の目的や活用場面を整理する。

本章で示す考え方は FDA をはじめとする海外規制当局の各種ガイドラインを参照している。そのため、現時点の日本における制度や運用状況と必ずしも一致しない点も含まれるが、今後の医薬品開発において目指すべきベストプラクティスの一例として提示するものである。

### 2.1. 定性的研究・定量的研究の使い分け/組み合わせ

定性的研究は、どのアウトカム、エンドポイントまたはその他の特性が患者によって重要と認識されているか、さらに、どの要素がベネフィットやリスクに関する患者の見方に影響を及ぼすかを特定するのに有用である。（詳細は第 4 章参照）

一方、定量的研究は、異なるベネフィットおよびリスクの特性に対する相対的重要度を測定し、様々な治療選択肢のベネフィットとリスクに関する患者選好の不均一性や分布を定量的に理解するのに役立つ。（詳細は第 5 章参照）

医薬品開発においてよく用いられる手順としては、まず定性的研究で患者にとって重要な因子を特定し「何を測定すべきか」を明らかにした上で、定量的研究によって「それらがどの程度重要であるか」を数値化する方法が挙げられる。定量的研究のみを実施した場合、重要な特性を見逃すリスクがあり、定性的研究のみを実施した場合、各特性がどのくらい重要なのかを知ることができない。両方を組み合わせることで、より堅牢なエビデンスが得られることが期待できる。これは混合研究法と呼ばれ、この段階的アプローチにより、患者のニーズをより深く理解した上で、それを定量的に評価することが可能になる。必要に応じて、定量的研究で明らかになった傾向（なぜその数値になったのか）を、定性的研究によって深く解明することも可能である。

ただし、定性的研究に関しては、例えば既に患者選好研究の先行研究がある場合やよく知られている疾患の場合は、既に十分な情報が存在するため新たに定性的研究を実施する必要がないこともある。また、臨床試験のエンドポイント候補を考える場合や、どのような状況でどのような選好が発生するかの細かい状況を知りたい等定量化することが目的でない場合は定量的研究を割愛することもある。このように研究目的から明らかに一方のみで足りる場合には、必ずしも定性的研究と定量的研究の両方を実施する必要はなく、患者負担やリソースを考慮して、適切な方法を選択することが重要である。

定性的研究と定量的研究のよくある使い分けについては図 4 に示す。

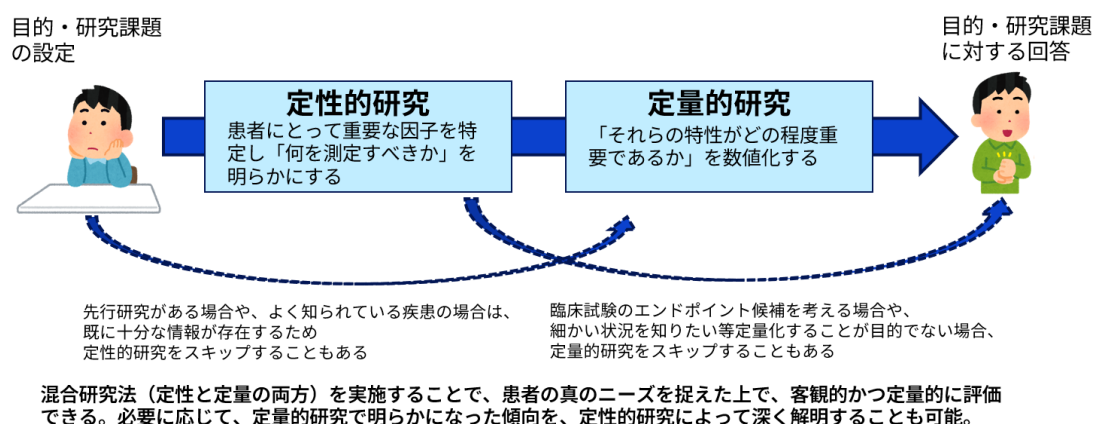


図 4 定性的研究と定量的研究の使い分け

## 2.2. 医薬品開発における患者選好情報の活用目的

PPI は疾患や治療に関する患者の価値観や優先順位を把握するための重要な情報であり、医薬品開発のさまざまな段階においてスポンサーにとって有用である。具体的には、疾患や治療を取り巻く背景を理解するための情報提供、開発戦略の検討・意思決定、患者にとって重要なエンドポイントの特定、さらには BR 評価における判断材料として活用される。

医薬品開発における患者選好研究の時期と方法と目的の関係を図 5 に示す。

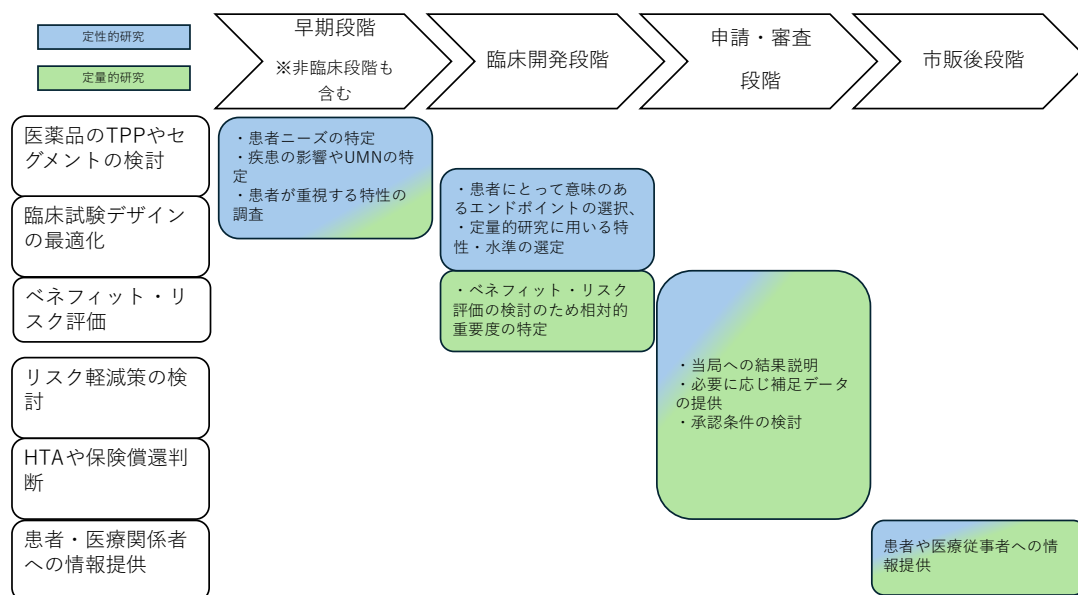


図 5 医薬品開発における患者選好研究の時期と方法と目的の関係

## 2.2.1. 医薬品のベネフィットとリスクの評価

PPI は医薬品の BR バランスを評価する上で重要な役割を果たす。BR 評価とは、CIOMS Working Group IV による報告書<sup>17)</sup>等で示されているように、ある医薬品が世に出る価値を有するか否かを判断するために、有効性などの期待されるベネフィットと、副作用等の潜在的なリスクを体系的に比較・評価するプロセスを指す。近年では、CIOMS Working Group XII により、BR 評価をライフサイクル全体で継続的に実施し、患者の視点を取り入れた構造化アプローチとして運用する重要性が示されている<sup>18)</sup>。

2010 年代半ばから医薬品開発や規制領域で多基準意思決定分析（Multi-Criteria Decision Analysis；MCDA）等の定量的 BR 評価手法が複数の治療特性を体系的に比較・統合する枠組みとして広く議論されてきた。近年では、この MCDA に定量的な PPI を組み込むことで、各特性の重要度を患者視点で重み付けし、より透明性と説明可能性の高い BR 評価を行うアプローチが注目されている。具体例としては 1 報目の報告書<sup>1)</sup>でも出典したリトレンチニブトシル酸塩の事例が挙げられる。このほかにも、2つの第3相臨床試験で収集された患者選好データを用いて、不眠症治療薬ダリドレキサントに対する患者中心の BR 評価を実施した研究も報告されている<sup>19)</sup>。当該研究では、患者が治療効果と副作用の間でどのようなトレードオフを受け入れるかを定量的に評価し、患者視点から見た BR バランスを解釈している。

下図<sup>16)</sup>のように患者選好の結果は提示した上で、別途臨床試験結果の BR 評価を示す等、PPI を用いた BR 評価には目的や文脈に応じて様々なアプローチが存在する。

| Endpoint |                                  | Rate Difference (95% CI)<br>/ 10,000 pat-years |                  | Preference Weight |
|----------|----------------------------------|------------------------------------------------|------------------|-------------------|
| Benefits | All-cause mortality              | -34.2                                          | (-71.62, 3.29)   | 1.000             |
|          | Disabling stroke                 | -11.8                                          | (-29.31, 5.81)   | 0.811             |
|          | Non-CNS systemic embolism        | -14.9                                          | (-24.29, -5.53)  | 0.420             |
|          | Non-disabling stroke             | 2.0                                            | (-21.16, 25.19)  | 0.359             |
|          | Myocardial infarction            | -20.9                                          | (-47.33, 5.59)   | 0.275             |
| Risks    | Major bleeding                   | 14.2                                           | (-35.30, 63.73)  | 0.267             |
|          | Non-major clin relevant bleeding | 43.8                                           | (-50.23, 137.77) | 0.159             |

Favors study drug  
 Favors comparator

Figure A3-13. Example table showing clinical data side-by-side with preference weights (mock data and weights).

図 6 BR 評価を PPI と組み合わせて評価するイメージ

PPI を BR 評価にどのように、またどの程度まで組み込むかについては、申請段階

に限らず、臨床開発の比較的早期から規制当局と議論を行い、目的や状況に応じた対応を検討することが望ましい。

PPI を BR 評価に組み込むことで、例えば以下のような価値が期待できる。

- ・ 患者にとって重要な特性の明確化
- ・ 特性間のトレードオフの可視化
- ・ 選好の不均一性の把握
- ・ 高リスク受容性の裏付け

「患者にとって重要な特性の明確化」に関しては、PPI を活用することにより患者が治療において重視する機能や特性といった要素を把握することができ、患者にとって重要なベネフィットとリスクの特定に役立つ。

「特性間のトレードオフの可視化」に関しては、各特性がどの程度重要であるか、また複数の特性間で患者がどのようなトレードオフを受け入れられるかを明らかにすることが可能となり、患者視点を反映した BR 評価への活用が期待される。FDA の BR 評価ガイダンスにおいても、適切に設計・実施された患者選好研究により、どの特性が患者にとって重要か、それらがどの程度重要か、そして患者が特性間でどのようなトレードオフを行う意思があるかを引き出すことができると述べられている。

「選好の不均一性の把握」に関して、不均一性とは例えば同じ疾患を持つ患者集団の中でも、個々の患者によって治療に対する価値観や優先度、すなわち効果、副作用、利便性等の特性に対する重要度が異なることを指す。PPI により、患者集団内での選好の不均一性を把握することで、BR プロファイルを良好とみなす可能性のある患者のサブグループの同定に役立つ。

「高リスク受容性の裏付け」に関しては、代替治療が存在しない重篤な疾患の場合など、患者がより高いリスクを受け入れる意思があることを示すエビデンスとして PPI を活用することができる。リスクに対する忍容性およびベネフィットに関する価値観の患者のばらつき(patient-centered variations)を確認することにより、患者の BR 選好に関する集団レベルでの評価を明らかにできる可能性がある。

PPI は、特に以下のような「患者の選好に左右されやすい (preference-sensitive)」状況において、規制上の意思決定に情報を提供する上で有用であると考えられている。

- (1) 期待されるベネフィットに対して、治療に伴う重大なリスク、またはリスクに関する不確実性が存在する場合
- (2) 最も重要なベネフィットとリスクに関する患者の見解が、患者集団内で大きく異なる場合
- (3) 最も重要なベネフィットに関する患者の見解が、医療専門家の見解と異なるこ

とが予想される場合

- (4) 複数の治療選択肢が存在するものの、すべての患者にとって明確に優れた選択肢が存在しない場合

なお、PPI は、安全性・有効性データを代替するものではなく、患者がどのような治療特性を重視し、どの程度のリスクを受容するかについての追加的なエビデンスを提供し、ベネフィット・リスク評価や医療上の意思決定を支援するものである。

## 2.2.2. 医薬品の Target Product Profile (TPP) やセグメントの検討

定性的または定量的な患者選好研究を通じて、疾患や治療状況に関するさまざまな懸念事項に関する定性的 PPI、さらに必要に応じて定量的 PPI を通じて TPP やセグメント設計の検討において有用な情報を得ることができる。具体的には、定性的 PPI では主に疾患の自然経過、現在の治療におけるアンメットニーズなどが挙げられ、定量的 PPI では主に治療において優先したい事項などが挙げられる。

これらの疾患・治療状況に関する情報は、目標とする製品プロファイルである TPP 検討に活用される。TPP には、期待される治療効果や、対象患者、既存治療と比較した位置づけや優位性等が含まれるが、PPI を踏まえることで、患者のニーズや価値観を考慮した、より適切な TPP および患者セグメントの設定が可能となる。

## 2.2.3. 臨床試験デザインの最適化

PPI は、臨床試験デザインの最適化にも活用され得る。具体的には、患者にとって最も重要なエンドポイントの特定およびその優先順位付けや、患者自身が「意味のある改善があった」と実感できる最小限の変化量 (Minimal Clinically Important Difference; MCID) の定義づけに活用できる。

また、訪問頻度、評価手法、PRO 項目の選択など、臨床試験のデザインを患者視点で見直すことで、臨床試験への参加率、継続率の向上につながる可能性がある。

## 2.2.4. リスク軽減策の検討

PPI は、リスク軽減策の検討においても有用な情報を提供し得る。例えば、頻繁なモニタリング検査に伴う患者負担といったリスク管理計画の実行可能性や、安全性モニタリングの方法 (例: デジタルツールの活用と診療所訪問の比較) に関する患者の選好を把握することが可能となる。

加えて、食事制限や活動制限など、リスク軽減策が患者の日常生活に及ぼす影響についての理解や、患者教育プログラムの効果および受容性の評価にも活用できる。これにより、患者の生活実態や価値観を考慮した、より実効性の高いリスク軽減策の検討が可能となる。

#### 2.2.5. 患者と医療関係者への情報提供

PPI は、患者および医療関係者による情報に基づく意思決定（informed decision-making）を支援するための情報提供にも活用され得る。例えば、添付文書に患者選好データを含めることで、患者と医療提供者の意思決定を支援し、異なる治療選択肢の比較をより容易にすることが期待される。

また、患者が自身の価値観や選好に基づいて意思決定を行えるよう支援するための患者支援ツールの開発にも活用可能である。具体的には視覚的な意思決定補助ツールやリスクをわかりやすく伝えるためのリスク・コミュニケーション・ツールの作成等が挙げられる。

#### 2.2.6. 医療技術評価（HTA）や保険償還判断

PPI は、医療技術評価（Health Technology Assessment : HTA）や保険償還判断においても、意思決定を補完する情報として活用され得る。これまでも QALY を算出するにあたり HRQoL を効用値に変換する際に、患者選好研究の結果を基に行われていることが多く、患者が治療効果や副作用といった複数の特性の間で、どの程度のトレードオフを受容可能と考えているかに関する情報を用いることで、新薬の相対的な価値や市場における需要を推定することが可能となる。これらの情報は、費用対効果評価や予算影響評価へのインプットとして利用される場合がある。

また、患者が特に重視する治療効果や、可能な限り回避したいと考えるリスクに関する情報は、医療政策や保険償還判断における優先順位設定を検討する際の参考情報となり得る。近年、各国の HTA 機関においても、患者視点のエビデンスは有用であるとの認識が広がりつつある。一方で、PPI を HTA や償還判断にどのように体系的に組み込むかについては、方法論や運用面での標準化が十分に確立されていないのが現状である。

### 2.2.7. 市販後の継続的な意思決定

PPI は、承認取得後の市販後段階においても、必要に応じて継続的な意思決定を支援するために活用され得る。具体的には、市販後に得られる新たな知見を踏まえつつ、BR 評価、TPP や患者セグメントの再検討、ならびにリスク軽減策の見直しなどに活用することが考えられる。

これにより、適応拡大、用量変更、剤形改良といったライフサイクルマネジメント上の意思決定を、患者の価値観やニーズを踏まえて検討することが可能となる。

## 2.3. 患者選好研究の時期と目的の関係

本節では、2.2 で整理した患者選好研究の目的について、医薬品開発の各時期において、どのような目的で、どのような方法（定性的研究／定量的研究）が用いられるかを整理する。

- 早期段階（非臨床段階も含む）：疾患および治療状況の理解や、TPP の検討に患者ニーズを反映させることを主な目的とし、主として定性的研究が用いられる。この段階で同様の目的で、定量的研究を実施することも可能である。なお、TPP は以降の臨床開発段階においても、TPP は新たな知見が得られる度に更新されるものであり、定性的研究および定量的研究の結果を踏まえて随時更新される。
- 臨床開発段階：患者にとって意味のあるエンドポイントの選択、臨床試験デザインの最適化、ならびに医薬品の BR 評価に向けた情報収集を目的とする。これらの目的に応じて、定性的研究および定量的研究が組み合わせられ、後続の定量的研究の設計や、規制当局との対話に活用される。
- 申請・審査段階：患者選好研究の結果を他の臨床データと統合し、規制上の BR 評価に活用することを目的とする。この段階では、主として定量的研究の結果が用いられるが、必要に応じて補足的に定性的研究が実施されることもある。なお、BR 評価への活用を想定する場合には、研究の計画段階から規制当局と相談することが望ましい。
- 市販後段階：継続的な BR 評価および患者とのコミュニケーションを通じた情報提供や意思決定支援を目的とする。これらの目的に応じて、定性的研究および定量的研究が、状況に応じて単独または組み合わせて用いられる。

### 3. 患者選好研究を実施する際の留意事項

患者選好研究（PPS）は身体的介入を伴わない場合が多いものの、患者の疾患経験や価値観といった繊細な情報を扱うため、倫理的な配慮と適切なデータ保護が欠かせない。不適切な倫理設計は、研究結果の信頼低下や規制当局への説明が困難になることも引き起こしかねない。本章では、PPS を実施するうえで考慮すべき倫理上のポイントを、理由と期待できる効果とともに整理する。

#### 3.1. 倫理的枠組み

患者選好研究は、「人を対象とする研究」に該当することが多く、研究の目的・リスク管理・参加者の権利保護について、以下の国内外指針に準拠する必要がある。

- 人を対象とする医学系研究に関する倫理指針  
GCP 省令に則った臨床試験に患者選好研究を組み込むこともできる
- 個人情報保護法/ GDPR（データ保護）  
データ収集にあたり個人情報保護法や、海外での利用が想定される場合にはデータ移転先の規制を遵守する
- ICH E22 ガイドライン（案）
- その他国際的な倫理原則：ヘルシンキ宣言、CIOMS など

患者の価値観や疾患経験は個人情報として非常にセンシティブであること、さらに、国際共同研究や複数国での提出を想定する場合には国際的に統一した倫理基準に準拠していることは重要である。こうした枠組みで進めることで、PPS 全体の信頼性と透明性が担保され、社内外のステークホルダーの理解も得やすくなる。

#### 3.2. 外部専門家（ステークホルダー）との早期からの連携

研究の妥当性と参加者保護の観点から、研究を計画する早い段階から外部専門家と連携することは極めて有用である。外部専門家の例として、患者団体、医療専門職（専門医、看護師）、倫理の専門家、行動科学・調査法の専門家、言語・コミュニケーションの専門家が挙げられる。PPS は患者が理解できる「特性の定義」「質問文のわかりやすさ」が重要であるものの、企業担当者だけで設計すると、患者にとっては不自然、もしくは理解困難な設問になりがちである。また、質問の表現（例えばリスク）、量が回答者に過度な負担とならないことも大切である。上記のような専門家との連携によって、特性の妥当性（患者にとって本当に重要か）、説明文の理解向上（読みやすさ・誤

解防止)を早期に発見できたり、調査設計の負担最小化、不適切・過剰な誘導リスクの低減、患者への配慮(例:希少疾患、小児)などリスク・負担への対応が可能となる。

海外のガイダンスにおいても早期段階での規制当局・患者連携が強く推奨されている<sup>10, 15)</sup>。多様な専門家との連携によって、回答の質の向上(理解しやすく、比較しやすい設計)、特性の誤解・過剰負担による回答の質の低下が回避可能となるばかりでなく、患者から見て「納得感のある研究」として参加促進と、研究の透明性・信頼性が向上することが期待できる。

外部専門家に意見を求める場合には、アドバイザー契約など適切な契約を結んだ上でコミュニケーションすることを推奨する。

### 3.3. 倫理審査 (IRB/EC)

PPS は身体的介入はないものの、健康に関わる個人情報や疾患情報を扱うこと、負担・心理的リスクがゼロとは限らないことなどから、倫理審査が原則必要である。審査を受けることで研究の透明性・信頼性が高まり、研究結果の発表、規制当局における意思決定にも利用可能となる。

### 3.4. インフォームド・コンセント

PPS に参加する前に、参加者は研究内容を理解し、自由意思で参加に同意する必要がある。説明文書には、研究目的、調査方法(オンライン/対面、所要時間、質問内容)、参加負担(心理的負担、認知負荷)、個人情報・データの取り扱い、断っても不利益はないこと、いつでも撤回できること、結果の取り扱い(公開方針含む)などを含む。また、特定集団への配慮事項として、小児を対象とする場合には保護者同意と本人のアセント、認知症・精神疾患を対象とする場合には代理人同意、本人の理解確認、重症患者を対象とする場合には短時間、分割実施などの負担軽減など対象に応じて検討が必要である。

### 3.5. プライバシー・データ保護

患者の疾患情報は個人情報の中でも要配慮個人情報に該当し、慎重な管理が求められる。研究で収集したデータを安全に分析・利用するためには、研究の計画段階で収集するデータの内容と取り扱いの管理手順を設計することが大切である。

保護の基本方針として考慮する事項として、収集は最小限(データミニマイゼーシ

ョン) とすること、匿名化／仮名化の徹底、アクセス権限の限定と管理、データ転送・保存の暗号化、データの保存期間・破棄プロセスの構築などが挙げられる。また、国際共同研究では GDPR (EU) や各国法規制に適合する必要がある。

### 3.6. 研究の透明性確保

PPS は非介入研究であるものの、透明性を確保することは重要である。その手段として、臨床試験、臨床研究と同様に、レジストリへの登録や結果の公開が期待される。

具体的には、公的レジストリ（例：jRCT、UMIN）に登録し、研究計画並びに結果を公開する。同じ内容の研究を繰り返し行うことは協力者に負荷となる可能性があるため避けるべきであり、そうした事前調査にも役立つ。また、出版バイアス回避のためにも、研究結果の発表・論文投稿も計画段階から考慮することが望ましい。治験で結果のレイサマリーの作成、公開が進んでいるのと同様に、PPS においても、協力者へのフィードバックは積極的に行うことを推奨する。

こうした研究の登録・公開は EMA・FDA のいずれにおいても重要項目として取り上げているだけでなく、透明性を確保することは社会的信頼、規制当局・患者団体からの信頼向上に繋がるとともに、結果の利用場面において研究の正当性を説明できることが期待される。一方、登録や公表のタイミングは開発戦略上の機密に関わることも想定され、適切なタイミングを社内で調整し、登録・公表を進めることを推奨する。

### 3.7. 患者負担の最小化

質問票の説明や質問が難しすぎたり情報量が多すぎると、疲労による直感的な回答や、中止・中断率が増えることが想定される。過度な負担はデータの信頼性と再現性を損なうため、回避すべきである。このため、回答者の負担最小化の検討は不可欠である。具体的な対策として、調査時間を適切に抑える、難しい用語を避ける、理解チェックのための練習問題を入れる、疲労を想定したタスク数設定、オプトアウトしやすい設計などが挙げられる。外部ステークホルダーの支援を得て負担の最小化を検討するとともに、本調査の前にプレテストを実施して回答者が正しく質問を理解できているか、負担の有無を確認することも対応策の 1 つである。患者負担の大きさは研究デザインや対象集団によって異なることから、患者を含むステークホルダーとの協働を通じて調査票や質問内容を設計し、プレテストや認知インタビューにより理解度や回答負担を確認することが望ましい。また、PPS の検討段階に、参照できる研究の有無を調査し、同じ内容の研究を繰り返さないことも患者負担を軽減する目的では重要である。

回答者の負担を最小化することは、患者にとって参加しやすい研究として協力が得

やすくなるとともに、無効回答や中止率が減りデータの質向上が期待できる。

### 3.8. 患者以外の回答者の選択

PPS は直接患者から選好情報を収集することが原則であるが、幼児や重度の認知障害のある患者のように患者本人が回答することが難しい場合、保護者や介護者、医療従事者に対して選好研究を実施することも有用である。そのメリットは患者本人の回答が困難な場合も研究を実施できること、患者の身体的・精神的負担を回避できることが挙げられる。一方、デメリットとしてこれら代理人の選好が患者本人の選好と異なる場合があり、患者の真の選好を反映していない可能性が挙げられる。しかしながら、保護者や介護者からの情報も有用である場合が多く、これら代理人を対象にした研究を実施可能である。この場合には、患者の選好とは区別して、代理人の選好研究であることを明記し、結果・解釈には注意を払う必要がある。

本章では、PPS を実施するうえで考慮すべき倫理上のポイントを概説した。PPS の結果を医薬品開発の意思決定に利用するにあたっては、研究の透明性確保やステークホルダーとの連携によって研究の信頼性と再現性を担保すること、患者保護に最大限配慮した研究設計が重要である。

## 4. 定性的研究について

本章では、FDA の PFDD ガイダンス 1・2 を基に、文献や実例を参照しつつ、未経験のデータサイエンティストにも理解しやすいよう、定性的研究の全体像や留意点の要旨をまとめた。

### 4.1. 定性的研究の目的と意義

定性的研究は、患者の経験や考え方、ニーズ、優先事項を体系的に収集し、その結果を医薬品開発や規制当局の意思決定に反映することを目的として実施される。症状や疾患が患者に及ぼす影響、治療に伴う負担、治療選好、アンメットニーズといった患者自身の声を詳細に把握できるという特徴がある。特に、新疾患領域や先行研究が限られている領域において、疾患特性を理解するうえで有用である。

具体的には、疾患や治療の影響、日常生活での困りごと、最も負担となっている症状、治療に対する期待などを把握したい場合や、新たな治療法や臨床試験計画に患者の声を反映させるための基礎情報の収集、既存の定量的研究では捉えにくい患者の価値観や優先事項を探索する場面で活用できる。

製薬企業では、新薬開発における治療コンセプトの検討や臨床試験デザインにおけるエンドポイントおよびアウトカムの選定など、各段階の意思決定の判断材料として活用できる。

### 4.2. 定性的研究の進め方

グラウンデッド・セオリー・アプローチ（GTA）の考え方を参考に、定性的研究の一般的な進め方を図 7 に示す。GTA の詳細は Appendix. 1 を参照。

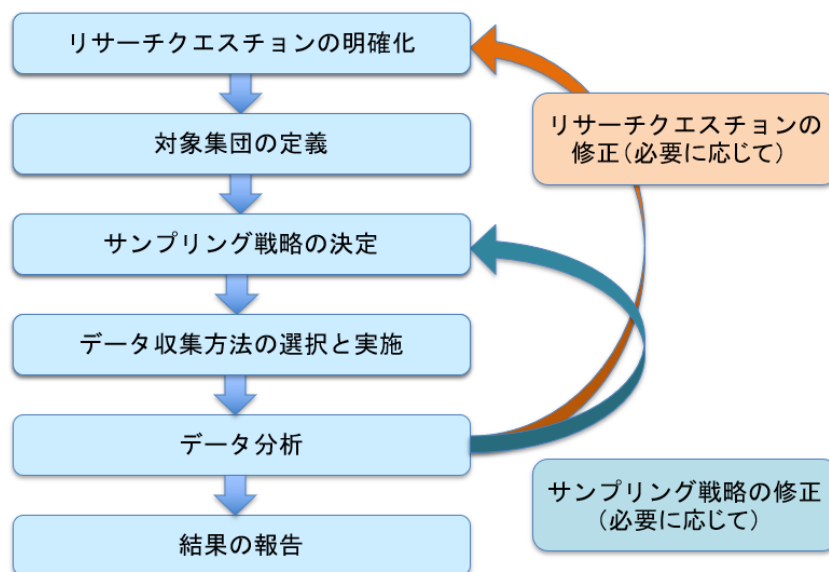


図 7 定性的研究の進め方

#### 4.3. 定性的研究の枠組み

本節では、定性的研究特有の注意が必要な内容について述べる。サンプリング方針、研究参加者とのコミュニケーション方法およびデータ収集方法、得られるデータの準備方法を説明する。

##### 4.3.1. サンプリング方針

###### 4.3.1.1. サンプリング方法の選択

目的に応じて適切なサンプリング方法を選ぶことが重要であり、例としては、特定の属性を持つ人を選ぶ目的サンプリング、参加者から他の参加者を紹介してもらうスノーボールサンプリング、来訪できる参加者を募集する利便性サンプリングなどがある。また、選好に影響すると考えられる年齢・性別・人種・疾患重症度・文化的背景・治療経験などの多様性を考慮して参加者を募集することが重要である。

理想的には理論的飽和に達するまで研究を続けるため、サンプリングおよびサンプルサイズの決定は反復的に行う。研究計画書には目標症例数とその根拠についても記載する。目標症例数を事前に明示できない場合は、その旨を記載する。ただし、実際に研究を行う場合の多くは、研究期間や予算の都合により、理論的飽和を目指した研究の実施は難しいことに留意する。理論的飽和の詳細は Appendix. 1 を参照。

## 4.3.1.2. 募集方法

研究参加者の募集方法としては、医療機関、アカデミック機関、研究コンソーシアム、調査パネルなどの患者リクルートに関する組織や患者団体を介する方法に加え、ソーシャルメディアを利用する方法などがある。選定した募集チャネルにより対象集団に偏りが生じる可能性はあるものの、複数のチャネルから参加者を募集することで、対象集団の多様性を反映した参加者を組み入れることができる。詳細は下表を参照のこと。

| 募集チャネル          | 強み                                                                                                                           | 課題                                                                                                                                                                                 |
|-----------------|------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 医療機関 / アカデミック機関 | <ul style="list-style-type: none"> <li>医師の診断に基づいて参加者を選定できる。</li> <li>適格基準を満たす参加者を特定しやすい。</li> <li>参加者の背景情報の精度が高い。</li> </ul> | <ul style="list-style-type: none"> <li>医療アクセスのある参加者に偏るなどして、多様な参加者を組み入れられないことがある。</li> <li>倫理審査や施設調整などが必要となり、参加者募集開始までの負荷および費用は高くなることもある。</li> <li>希少疾患などでは利用可能な施設が限られる。</li> </ul> |
| 調査パネル           | <ul style="list-style-type: none"> <li>登録会員を対象とするため、研究参加者の募集は比較的容易。</li> <li>適格基準を満たす参加者を効率的かつ計画的に募集しやすい。</li> </ul>         | <ul style="list-style-type: none"> <li>登録会員に限定されるため多様な参加者を組み入れられないことがある。</li> <li>自己申告が中心となるため、診断や臨床情報の正確性確認が限定的なことがある。</li> </ul>                                                 |
| ソーシャルメディア       | <ul style="list-style-type: none"> <li>低コストで広く参加者を募集できる。</li> <li>希少疾患などのアクセスが難しい患者集団へ到達できることがある。</li> </ul>                 | <ul style="list-style-type: none"> <li>自己選択バイアス*が生じやすい。</li> <li>自己申告が中心となるため、診断や臨床情報の正確性確認が限定的なことがある。</li> </ul>                                                                  |

\* 自己選択バイアス：参加意思のある人のみが自発的に参加することで、参加者が本来の対象集団全体の多様性を十分に反映しなくなること。

## 4.3.2. 研究参加者との主なコミュニケーション方法

## 4.3.2.1. 実施形態の選択と設計上のポイント

研究参加者とのコミュニケーションについては、リサーチクエスト（Research Question）や対象集団の特性に応じて選択する。FDA は、インタビューやフォーカスグループの実施形態について特定の方法を推奨しておらず、対象集団、研究の特性および研究目的に応じて適切な方法を選ぶべきとしている。FDA は、選択した実施形態の妥当性およびその根拠を評価する。詳細は下表<sup>12)</sup>を参照のこと。

| 実施形態               | 強み                                                                                                                                                    | 課題                                                                                                                           |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| 対面                 | <ul style="list-style-type: none"> <li>言語・非言語（表情・ジェスチャー等）両方の反応を観察できる。</li> <li>書面やブレインストーミングなどの演習を取り入れやすい。</li> </ul>                                 | <ul style="list-style-type: none"> <li>移動・会場などの費用がかかる。</li> <li>健康状態、地理的制約、経済的制約などで参加が難しい場合がある。</li> </ul>                   |
| 電話                 | <ul style="list-style-type: none"> <li>健康状態や地理的制約で対面が難しい場合も実施可能。</li> <li>移動・会場などの費用を削減できる。</li> </ul>                                                | <ul style="list-style-type: none"> <li>信頼関係を築きにくい。</li> <li>プライバシー確保が難しいことがある。</li> <li>周囲の音・家族の同席などで集中しにくいことがある。</li> </ul> |
| オンライン会議<br>(カメラ付き) | <ul style="list-style-type: none"> <li>カメラを使用することで、参加者が顔や表情を確認できるため、対面に近いコミュニケーションが可能。</li> <li>地理的制約がない。</li> <li>対面調査および電話調査の強みの一部を併せ持つ。</li> </ul> | <ul style="list-style-type: none"> <li>PC やインターネット環境などの整備が必要。</li> <li>通信トラブルやセキュリティ、プラットフォーム利用料などの課題がある。</li> </ul>         |

複数の方法を組み合わせることも有効で、例としては、事前にインターネット上で幅広く意見を集め、詳細は対面や電話で深掘りするなどが挙げられる。また、研究参加者の多様性などを考慮することも重要である。

上記に挙げた以外にもソーシャルメディアでのコミュニケーションも可能で、アク

セスが困難な集団にもアプローチができるという利点がある。但し、自己選択バイアスや自己申告による診断や臨床情報の不確実性などの課題があるため、特に医薬品開発においては検討の際に注意が必要である。

#### 4.3.2.2. その他の配慮事項

障がいや年齢など、多様な背景を持つ参加者への配慮も重要である。例えば、文字のサイズを大きくする、スクリーンリーダー対応にする、面接時間を短くする、未成年者に対しては保護者同席の工夫をするなどが挙げられる。また、多言語対応や文化的背景への配慮が必要となる場合もある。

倫理的配慮およびプライバシー保護の観点から、インフォームド・コンセントの実施、匿名化处理、適切なデータ管理を行う必要がある。

#### 4.3.3. データ収集の主な方法

データ収集方法には、一対一面接（One-on-One Interview）やフォーカスグループ（Focus Groups）、デルファイパネル（Delphi Panel）、観察法（Observation Method）などの方法がある。最も一般的なのは、一対一面接とフォーカスグループであるため、本節ではこれらに絞って説明する。いずれの方法においても司会者・質問者には能力や経験が必要となる。事前に決めた質問ではなく、議論の中で深掘りする点ほど、司会者や質問者の能力が問われることに留意する。また、面接ガイドにヒアリング方法を丁寧に記載することが推奨される。なお、面接ガイドに関する文献や資料は参考文献を参照（20-21）。

##### 4.3.3.1. 一対一面接

一対一面接では、参加者が治療や疾患に対して重視することや、どのような妥協点を許容するかななどを深掘りできる。手法としては、構造化面接（Structured Interview）、半構造化面接（Semi-Structured Interview）や非構造化面接（Unstructured Interview）などがある。

構造化面接は、事前に定められた質問項目および順序に従い、一貫した手順で実施される面接方法。面接者による質問の変更や追加的な深掘りは基本的に行わないため、標準化されたデータ収集ができる。

半構造化面接は、事前に質問やテーマを用意しながらも、会話の流れに応じて質問を深掘りしたり内容を調整する面接方法。決まった質問のみに限定されず、一方で完全

に自由な雑談でもない中間的なスタイルで、一定の枠組みを持ちながらも、個々の回答に応じた探索的な対話をするため、相手の考え方や感じ方を体系的かつ深く理解できる。誘導的な質問を避けて「あなたにとって大事なことは何ですか」といった自由回答を促す問いかけを行い、回答に応じて追加質問をすることで価値観や選好を把握する。

非構造化面接は、事前質問をほとんど決めず、テーマのみを提示したなかで回答者の語りに応じて柔軟に質問を展開する方法である。半構造化面接よりも自由度が高く、会話の流れに沿って進めることで、回答者の考えや価値観を自然に引き出しやすい特徴がある。「治療で最も改善したいことは何ですか」「それはどのような場面で強く感じますか」「その中で特に負担が大きいのはどの点ですか」などの追加質問を行う。このように回答内容に応じて段階的に深掘りすることで、研究参加者の価値観や選好を引き出すことができる。

#### 4.3.3.2. フォーカスグループ

複数の研究参加者が参加し、司会者の進行のもとで治療の効果や副作用、日常生活で困っていることなどについてグループで議論する方法である。他の参加者の意見を聞くことで、自分自身の価値観や選好を見直したり、優先順位の違いに気づくことができる。年齢や重症度、性別などのサブグループに分けて実施することで、より多様な選好を把握できる。

一対一面接とフォーカスグループの強みと課題をそれぞれ下表に示す。

| 収集方法    | 強み                                                                                                                                                                                                                                                         | 課題                                                                                                                                          |
|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| 一対一面接   | <ul style="list-style-type: none"> <li>● 個人の経験や視点について深くヒアリングできる。</li> <li>● アンケートの内容を解釈できているかを確認できる。</li> <li>● 柔軟性がある</li> <li>● フォーカスグループよりはスケジュールを調整しやすい。</li> <li>● フォーカスグループよりはプライバシー開示に抵抗がある人に向いている。</li> <li>● 面接場所に柔軟性がある（例：施設・自宅・電話会議）</li> </ul> | <ul style="list-style-type: none"> <li>● 時間を要する（例：複数の研究参加者の面接の実施）</li> <li>● センシティブな設問に正直に回答することに抵抗を感じることがある。</li> <li>● 費用がかかる。</li> </ul> |
| フォーカスグル | <ul style="list-style-type: none"> <li>● 興味のあるテーマについて多様な</li> </ul>                                                                                                                                                                                        | <ul style="list-style-type: none"> <li>● 各参加者の考えを十分に深</li> </ul>                                                                            |

| 収集方法 | 強み                                                                                                                                                                                    | 課題                                                                                                                                                                                                                                                                             |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ープ   | <p>視点から深い意見が得られる。</p> <ul style="list-style-type: none"> <li>柔軟性がある</li> <li>一度に複数の参加者から意見収集できるため効率的。</li> <li>参加者同士の意見を聞くことで自身のアイデアが誘発されることがある。</li> <li>一対一面接と比較して費用が安い。</li> </ul> | <p>掘りするには限界がある</p> <ul style="list-style-type: none"> <li>他の参加者によって参加者の気が散ることがある。</li> <li>同調圧力により、参加者が意見表明をためらうことがある。</li> <li>一部の参加者が会話を独占するし、多様な意見が得られなくなることがある。</li> <li>集団の場がプライベートな情報の取得を妨げることがある。</li> <li>グループや公共の場で話すことに抵抗を感じることもある。</li> <li>スケジュール調整が大変。</li> </ul> |

#### 4.3.4. データ分析の準備

データは膨大な量になるため、分析や解釈がしやすいように標準化された方法で整理することが重要である。データの準備には、以下のような情報源からのデータの集約やテキスト形式への変換が必要となる。

- 面接やフォーカスグループ
- 録音・録画
- インターネット（ソーシャルメディア、チャットルームの対話など）
- メタデータ（面接日時、司会者名、回答者の所属、分析のアイデアなど）

記録方法のポイントとしては、面接ガイドや面接記録（メモ・録音・録画など）を活用し、情報の一貫性と質を担保する。録音した音声の書き起こしには、後述するソフトウェアの活用も有用である。面接ガイドには、傾聴姿勢を保つこと、発言していない人にも声をかけること、冒頭で自己紹介や雑談をして心理的安全性を高めること、必要に応じてお菓子を用意するなど、面接を円滑に進めるための工夫を記載する。

データを適切に管理するためには、データ管理計画を作成してデータの保存、必要に応じた匿名化、バックアップおよびアクセス管理を含む運用を計画的に実施することが望ましい。データ品質を確保するため、研究チームへのトレーニング、役割分担の

明確化、トラブル対応マニュアルの整備を推奨する。

4.4. 定性的研究におけるデータハンドリングと解析

定性的研究においては、インタビューなどでテキスト情報をはじめとする PED を収集した後に、単語や文節、文章を分類・コード化し、適切な方法で結果を要約・可視化することで、第三者が解釈可能な形に整理することが一般的である。コード化とは、インタビュー等で得たテキストデータから、重要な概念やテーマを見つけ出し、ラベル（コード）を付けて分類・整理する作業である。

FDA の PFDD ガイダンス 2 では、FDA は通常、規制当局への申請のために定性的データをコード化することが推奨されると述べられている<sup>12)</sup>。本節では、コード化を実施する場合を想定して、得られた定性的研究データからどのように結果を生成し、医薬品開発に資する形でまとめるか、ならびに、まとめる際の注意点について述べる。

4.4.1. コード化によるテキスト分析

定性的研究におけるデータ解析は、収集したテキストデータを整理し、記述・分類（コード化）を行ったうえで解釈し、最終的に結果を分かりやすい形で提示するという、複数の工程から構成される。代表的な工程は、以下のとおりである。

| Steps       | Description                                                                                                                                                                              |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| データのまとめと整理  | <ul style="list-style-type: none"><li>研究及びその他のデータ収集から得たメモを、有用で標準化された順序（電子的保存及びコンピュータプログラム）で整理する</li></ul>                                                                                |
| データの記述及び分類  | <ul style="list-style-type: none"><li>まとめたデータを細分化する</li><li>細分化したものを別のグループ/順序に再編成する（例:コード）</li></ul>                                                                                     |
| データの解釈      | <ul style="list-style-type: none"><li>グループ化/順序化されたデータを使用して、データのより大きな意味を特定する</li><li>データの概念を他のエビデンスに関連付ける（関連する文献、専門家の意見等）</li><li>研究目的に該当する場合、新しい重要なコンセプトが現れていないか（すなわち飽和度）を評価する</li></ul> |
| データの表現及び可視化 | <ul style="list-style-type: none"><li>データを分かりやすくパッケージ化する（例:テキスト、表、図）</li></ul>                                                                                                           |

[FDA の PFDD ガイダンス 2]

研究計画時（データ収集開始前）に研究計画書の解析計画の章または解析計画書に、

予め収集するデータ範囲、解析方法、コード化の手順（複数のコード化作業者がいる場合の合議方法を含む）等を定めておく必要がある<sup>12)</sup>。解析計画の記載例は Appendix 2 を参照のこと。

なお、定性的研究におけるデータ解析のステップは反復的であり、各工程は必ずしも一方向に連続して進むとは限らない。必要に応じて前工程に戻りながら、複数工程を並行して進めることで、データへの理解が深まり、より妥当な解釈に到達しやすくなる点に留意する。

4.4.1.1. テキストのコード化の推奨事項

コード化を適切に実施し、最終的に解釈可能な結果としてまとめるために、推奨事項を以下に示す。コード化から報告までの一般的な流れを図 8 に示す。

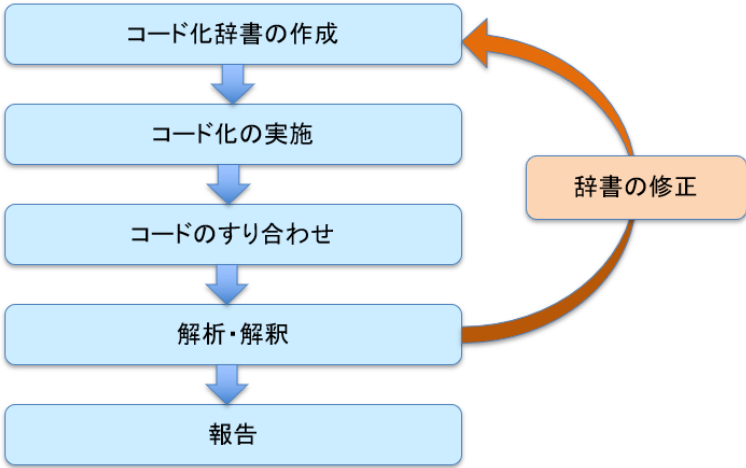


図 8 コード化から報告までの流れ

(1) コード辞書（coding dictionary）の作成

図 8 に示すとおり、コード化の起点としてコード辞書の作成が重要である。コード辞書には、少なくともコード（カテゴリまたは概念の説明）、コード間の階層構造、ならびにコード導出に関する判断の根拠や考えを記録したメモを含める<sup>12)</sup>。

| Study material    | Components                                                                                                          | Considerations                                                                                                                              |
|-------------------|---------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Coding dictionary | <ul style="list-style-type: none"><li>コード（カテゴリまたは概念の説明）</li><li>コード同士の階層構造</li><li>メモ（コードの導き方に関するアイデアや考え）</li></ul> | <ul style="list-style-type: none"><li>コードの定義、指示、考慮事項など、分類のための明確な指示の概要を示す</li><li>既存の知識から初期コードを導き出す（例:自然史、概念モデル、疾患モデル、ディスカッションガイドの</li></ul> |

| Study material | Components | Considerations                  |
|----------------|------------|---------------------------------|
|                |            | 構造)<br>• メインコードと詳細コードの階層構造を作成する |

特にコード化を複数人で実施する場合、コード辞書により、定義・指示・粒度・階層といったコード化方針を事前に明確化することで、コード化実施者間の解釈のずれを抑え、コード化の一貫性を高めることができる。

## (2) コード化アプローチ選択時の考慮事項

解析目的に応じて、適切なコード化方法や詳細度（例：行単位でのコード化、意味単位でのコード化、発言全体をまとめたテーマコード化等）を選択する必要がある。一般的に考慮すべき点を以下に示す。

| 考慮事項                                             | 補足説明                                                                     | 具体例                                                                                              |
|--------------------------------------------------|--------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| 関心のあるデータに適切なコード化方法を選択する                          | 研究の焦点に応じて、異なるコード化方法が適切となる                                                | 患者の症状の頻度や重症度に焦点を当てる場合と、症状が日常生活に与える影響に焦点を当てる場合では、異なるコード化方法を選択する                                   |
| コード化対象に関する適切な詳細度を決定する（例：ライン毎のコード化、またはテキストの一部の選択） | インタビュー記録全体を大まかなテーマでコード化するのか、一行ごと、または特定の意味を持つテキストの一部を選択して詳細にコード化するのかを決定する | 患者が「痛みで夜眠れず、翌日の仕事に支障が出る」と述べた場合、文全体を「日常生活への影響」とコード化することもできれば、「痛み」「睡眠障害」「就労への影響」と細分化してコード化することもできる |
| どのデータがコード化するのに十分重要かを判断する                         | すべての発言を同等に扱うのではなく、研究目的に照らして重要性を評価する                                      | 疾患の症状に関する研究であれば、患者が繰り返し言及する症状や、複数の患者が共通して述べる症状は重要なデータとして優先的にコード化する                               |
| データをコード化する際は、段階的に高い概念レベルへ移行するようにする               | 個別の発言を最初から高次の概念でコード化するのではなく、まず中間レベルの概念でコ                                 | 「階段が登れない」「買い物に行けない」「子供と遊べない」といった個別の発言を、まず「移動の制限」「日常活動の制限」「社会的活動の制限」と                             |

| 考慮事項                                                           | 補足説明                                                            | 具体例                                                                                      |
|----------------------------------------------------------------|-----------------------------------------------------------------|------------------------------------------------------------------------------------------|
|                                                                | ード化し、その後に必要なに応じてより高次の概念へと統合していく                                 | いった中間レベルの概念でコード化し、その後「身体機能の低下」という高次の概念へと統合する                                             |
| コード化された単語の文法的な形式を注意深く検討する (e.g. actions vs processes vs nouns) | 患者の体験を動作として捉えるのか、変化する過程として捉えるのか、状態として捉えるのかによって、データの解釈が変わる可能性がある | 「痛む(action:動作)」 「痛みの悪化(process:過程)」 「痛み(noun:名詞)」では、それぞれ異なるニュアンスを持つため、研究目的に応じて適切な形式を選択する |

### (3) コードのすり合わせ（合議）と信頼性の確認

複数のコード化担当者がある場合、担当者が独立にコード化を実施した後、コードのすり合わせ（合議）を行うことが有用である。コード辞書を用いても、コーダーごとの判断の癖や解釈の差は残り得るため、合議を通じて判断基準を明確化し、文書全体に対するコード付与の一貫性を高める。

定性的研究の透明性および信頼性を確保するために、コード化ならびにコード化結果からカテゴリ等を導出する過程を十分な詳細とともに記録し必要に応じて報告することが重要である。具体的には、コード辞書の作成者、コード化担当者、合議への参加者およびテーマ導出の手順を明確化し、分析上の重要な判断過程やコード辞書の改訂履歴を記録しておくことが望ましい。これにより、第三者が分析結果に至った過程を追跡し、その妥当性や信頼性を評価できるようになる<sup>12,23)</sup>。

すり合わせの際は、単純な一致率の要約のほか、Cohen's kappa などの定量的な指標も有用である<sup>23)</sup>。すり合わせ結果を踏まえ、必要に応じてコード辞書（コードマニュアル）の見直しや追加のコード化を行う。

#### 【コード化の具体例】[FDA PFDD GD2]

##### 例1) コードを定性的データに当てはめる

| コード | 書き起こし                                                                   |
|-----|-------------------------------------------------------------------------|
|     | インタビュアー)<br>薬を服用したときの気分はどうですか？<br><br>患者)<br>薬を飲んだ後、 <b>極度の疲労を感じます。</b> |

| コード                                              | 書き起こし                                                                   |
|--------------------------------------------------|-------------------------------------------------------------------------|
| コード：疲労感<br>コード：時間依存性の薬剤                          | それが服用する時間帯に関係しているのかわかりません。<br>いずれにしても、家の周りの雑用をこなしたり、長い散歩をしたりすることができません。 |
| コード：日常生活への影響<br>コード：身体機能への制限<br>コード：発疹<br>コード：痒み | また、上腕に沿って発疹が出ていることに気づきました。<br>それがかなりの痒みを引き起こしています。                      |

## 例2) データからコードを生成する

| 書き起こし                                                                                                                                              | コード化例                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|
| 患者)<br>極度の痛みがあったため、医師は私の薬のいくつかを再評価したいと考えました。<br>医師は、モニタリングのために病院に入院する必要があると言いました。<br>これが保険でカバーされないのではないかと心配でした。<br>結局、家族に電話をして、お見舞いに来てもらえるか確認しました。 | 痛み<br>治療レジメンの評価要請<br>入院<br><br>医療へのアクセス<br>家族のサポート |

## 4.4.2. ソフトウェアの活用

コード化を効率的かつ一貫性をもって実施するために、ソフトウェアの活用が有用である。具体的には Computer Assisted/Aided Qualitative Data Analysis Software (CAQDAS) として NVivo、MAXQDA など市販の製品が存在するほか、オープンソースのソフトウェアも存在する (図 9)。

ソフトウェアを用いることによってインタビュー音声データの文字起こし、コード化の一括置換やマージ、サブカテゴリ化などの作業を効率化できる。また、コードの集計や可視化を一連の作業として実施でき、複数のコード化作業者間での共有や、前節で述べた合議 (すり合わせ) の運用も容易になる。

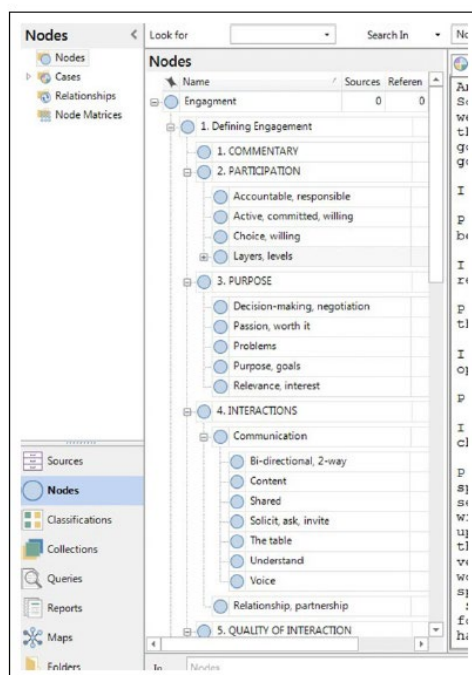


Figure 3. A screenshot of the coding framework in NVivo.

図 9 NVivo のコーディング画面例 22)

#### 4.4.3. 定性的研究の解析手法

前節までに述べたとおり、定性的研究では、テキストデータを整理し、記述・分類（コード化、カテゴリ化）したうえで解釈し、結果を提示する。コード化・カテゴリ化の後には、コードや語の出現傾向を定量的に把握して可視化するアプローチ、AI-assisted analysis（生成 AI を用いた支援的解析）のいずれか、または両者を組み合わせてデータをまとめることができる。その他、会話・ディスコース分析、ナラティブ分析、テーマ分析を用いて分析することも可能だが、本報告書では取り上げない。

本節では、コード化後に用いられる代表的な解析手法を、目的に応じて整理して示す。

データ分析（データ整理、解釈、まとめも含む）方法としては以下が代表的である：

- コード化・カテゴリ化（e.g. 理論的コード化、テーマ的コード化など）、コードを単語と見なした定量的集計、自然言語処理手法（e.g. 共起分析（4.4.3.1 も参照））等
- 会話・ディスコース分析、ナラティブ分析、テーマ分析、AI-assisted analysis（生成 AI にテキストを読み込ませまとめさせる）等

4.4.3.1. コード化結果の定量的集計と可視化（テキストマイニングを含む）

コード化・カテゴリ化の結果を、「コードを単語（または特徴量）とみなす」立場で集計・可視化することで、全体傾向や特徴的な語・概念の関連を俯瞰できる。これらの定量的集計および可視化手法は、定性的データの特徴を俯瞰的に把握し、分析や解釈の手がかりを得るうえで有用である。一方で、コードや語の出現頻度、共起関係、極性といった指標は、概念の重要性や臨床的意義そのものを直接示すものではない点に留意が必要である。これらの結果は、あくまで探索的な情報として位置づけ、研究目的や文脈、定性的な解釈と併せて用いることが重要である。

代表例として、以下に列挙する解析方法が挙げられる。

- 文章の基本情報の解析  
文書量、発言数、語彙数等の基本統計を確認し、データの規模や偏りを把握する。

| 基礎集計データ |        | 品詞の出現頻度 |        |
|---------|--------|---------|--------|
| 総行数     | 14,520 | 名詞      | 45,120 |
| 総文数     | 8,345  | 動詞      | 28,450 |
| 延べ単語数   | 94,210 | 形容詞     | 12,300 |
| 単語種別数   | 3,105  | 副詞      | 5,100  |
|         |        | 感動詞     | 3,240  |

図 10 文章の基本情報の出力例

- コードの出現頻度集計  
コード（または語）の出現頻度を集計し、主要トピックや頻出概念を把握する。

|                    | 患者A | 患者B | 患者C |
|--------------------|-----|-----|-----|
| 痛み・身体的苦痛           | 12  | 3   | 10  |
| 将来への不安             | 5   | 8   | 4   |
| 治療の副作用             | 15  | 6   | 9   |
| 医療者との<br>コミュニケーション | 2   | 11  | 3   |
| 家族の負担              | 7   | 4   | 12  |

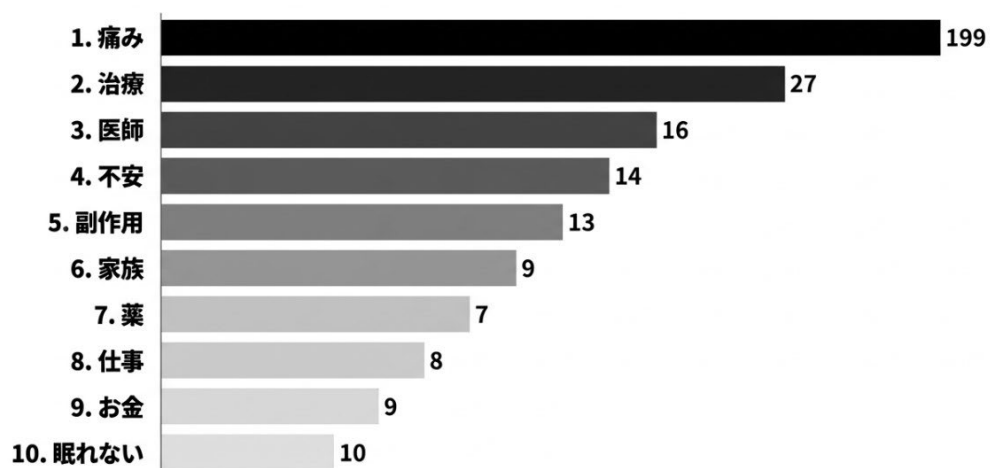


図 11 コードの出現頻度集計の出力例

- 係り受け頻度解析・共起分析  
語（あるいはコード）の同時出現や関連構造を確認し、概念間の結びつきの手がかりを得る。共起は語の同時出現を示すものであり、因果関係を意味するものではないことに留意する。

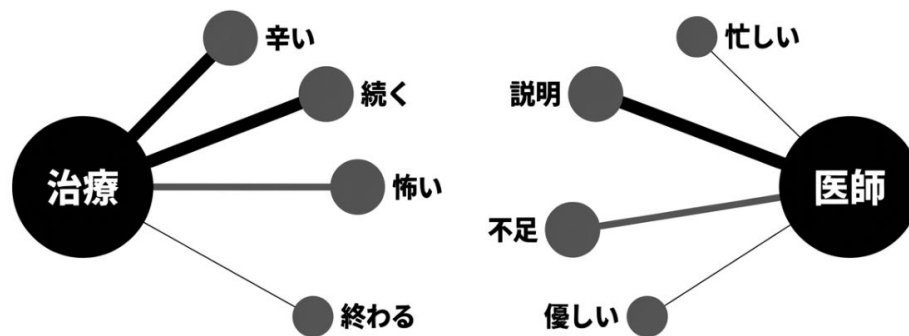


図 12 係り受け頻度解析・共起分析の出力例

- 評判分析（ポジネガ）

発言の極性（肯定／否定）の傾向を把握する。ただし、皮肉表現や否定の扱い等により誤判定が生じ得るため、辞書、ルール、人による直接の確認を含めた運用をする必要がある。

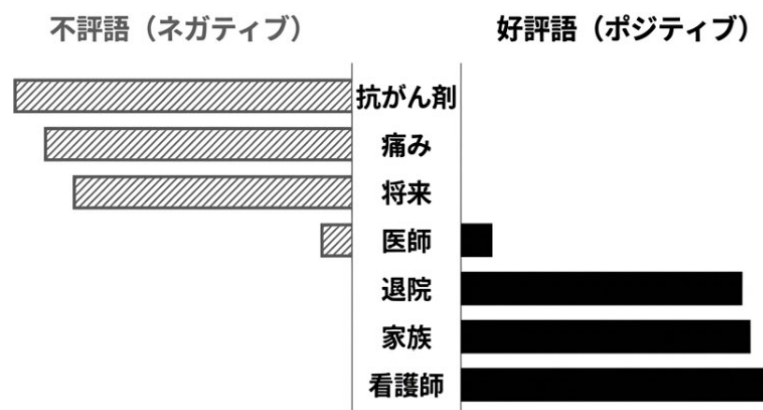


図 13 評判分析の出力例

- 対応バブル分析

対象（属性群）と語（コード）等の関係を二次元に配置し、特徴的な関連を可視化する。

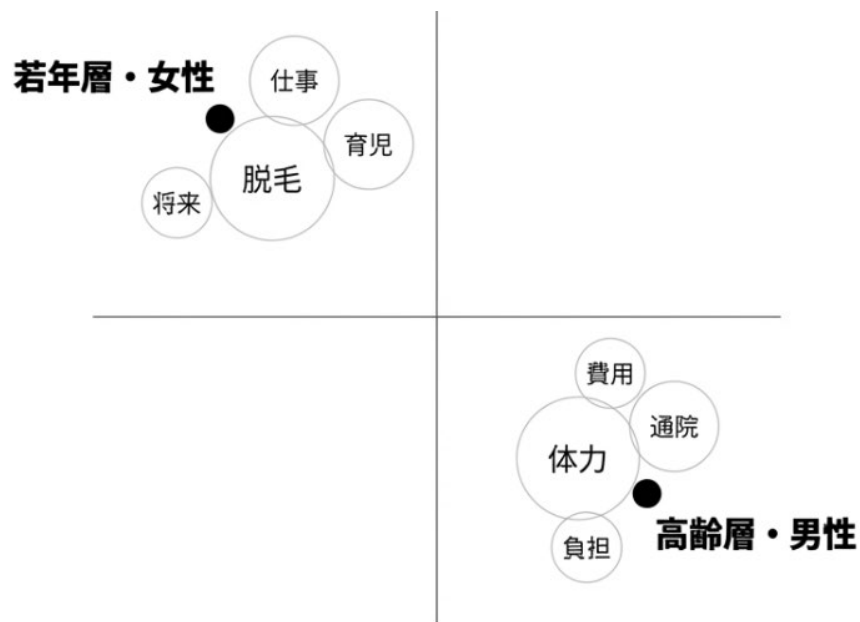


図 14 対応バブル分析の出力例

- 時系列分析  
時点（訪問、治療フェーズ等）に沿った語（コード）の推移を確認し、体験の変化や介入前後の傾向を探索する。

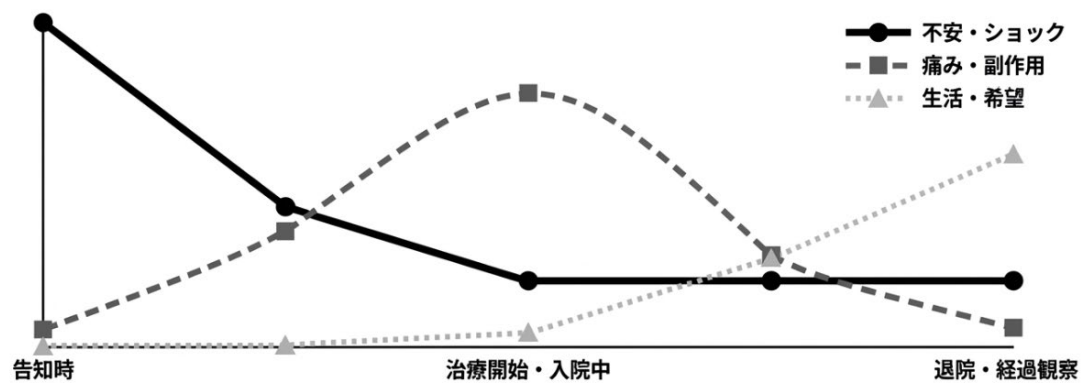


図 15 時系列分析の出力例

- ワードクラウド  
頻出語を直感的に提示する。探索やコミュニケーションには有用であるが、重要性の判断や結論の根拠として過度に依存しないことに注意する。

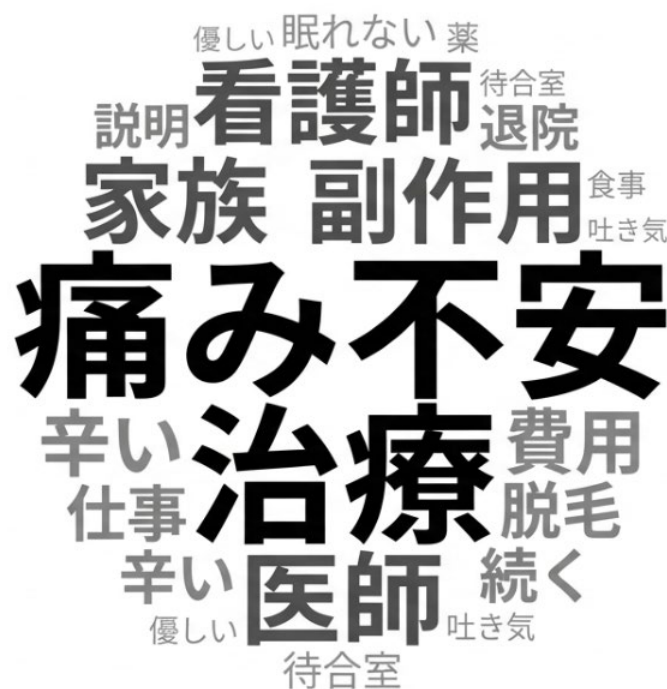


図 16 ワードクラウドの出力例

#### 4.4.3.2. AI-assisted analysis（生成 AI を用いた支援的解析）

コード化・カテゴリ化をベースとしてコードや語の出現傾向を定量的に把握して可視化する方法のほか、生成 AI を用いて文章を読み込ませ要約や論点抽出を支援する方法を組み合わせ、データの整理と理解を進めることもできる。ただし、これらは探索を加速し得る一方で、出力の再現性や根拠追跡可能性（どの記述に基づく結論か）が課題となり得るため、最終的な解釈は研究者がデータと突合して検証し、手順・プロンプト・バージョン等を含めて記録する等、透明性確保を前提として位置づける必要がある。

#### 4.5. 成果物のイメージと医薬品開発への活用

前節までに述べた方法により得られた定性的研究の分析結果は、概念モデル、テーマ一覧、特性リスト、評価項目案などの成果物として整理することで、医薬品開発上の意思決定に活用可能な形で整理することができる<sup>24)</sup>。

本節では、学術研究および製薬企業が関与した研究の具体例を通じて、定性的研究から得られたデータが、どのような分析を経て、どのような成果物が作成され、どのような意思決定に活用されているかを示す。

例1) がん患者の治療選好に関する定性的研究<sup>25)</sup>

定性的研究は、患者が治療や医療体験の中で何を重視し、どのように意思決定しているかを深く理解するために用いられる。Higuira-Gutiérrez ら (2021) は、がん患者の治療選好の構造を明らかにすることを目的として、GTA に基づく定性的研究を実施した。

本研究では、成人がん患者を対象に半構造化面接を複数回実施し、逐語録を作成したうえで段階的なコード化を行っている。分析では、初期コードの付与後、コード間の関係性を比較・統合しながら抽象度を高め、カテゴリレベルでの整理を行った。このプロセスを反復的に実施し、新たな概念が出現しないことを確認することで理論的飽和に到達している。

成果物として、患者の選好は主に、

- (1) 医療者の特性（専門性に加え、傾聴や共感といった人間的側面）、
- (2) 医師-患者関係（医師に委ねる志向と自己決定志向）、
- (3) 治療目標（延命重視か QOL 重視か）

といった複数の領域に整理され、それぞれの領域において、参加者間で共通して観察される傾向と、価値観や状況に応じて異なる側面が示された。

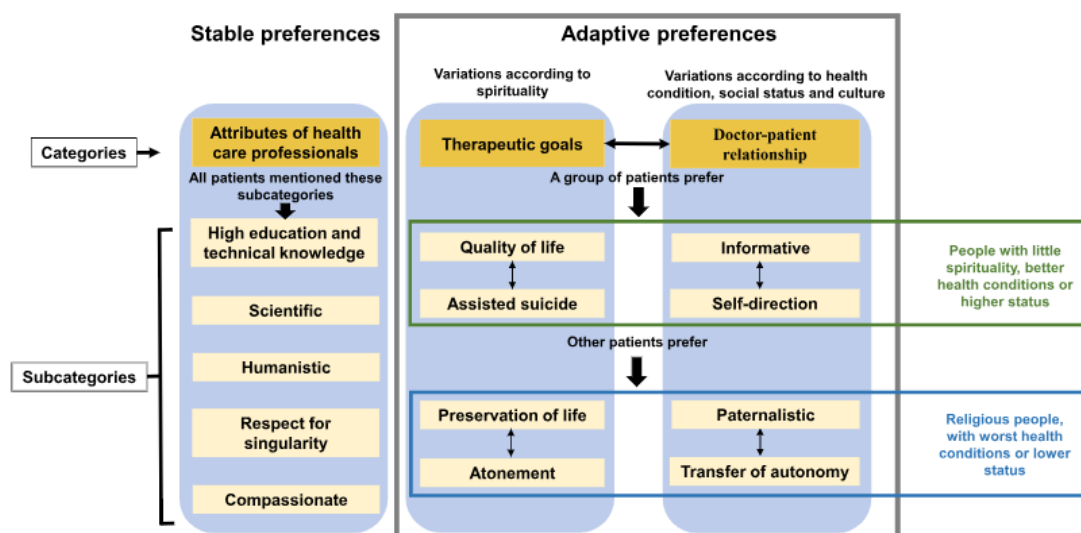


Figure 1 Significance matrix of therapeutic preferences in patients with cancer.

図 17 がん患者の治療選好を示す概念的枠組み（カテゴリ間の関係を示した図）

これらの成果は、患者選好の構造を整理した概念的枠組みとして、後続の定量的選好調査（例：DCE における特性候補の抽出）や、臨床試験における評価項目を患者視点から検討する際の基礎情報として活用することができる。

一方で、本研究は特定の地域・条件下の患者を対象とした探索的研究であり、得られた結果を他集団へ一般化するには留意が必要である。

例2) 血友病における遺伝子治療特性抽出と DCE 前段研究<sup>26)</sup>

定性的研究は、定量的選好評価（Discrete Choice Experiment：DCE）を設計する前段階としても重要な役割を果たす。血友病における遺伝子治療を対象とした本研究では、患者および関係者が治療のどの治療特性を重視しているかを明らかにする目的で、製薬企業が関与した定性的研究が実施された。

本研究では、血友病患者および介護者、ならびに臨床専門家を対象に半構造化面接を行い、治療に関連する特性について意見を収集した。分析では、フレームワーク分析\*を用い、各特性の重要度を順位付けすることで、意思決定に活用できる形で結果を整理している。

\*: 事前に定義した治療特性の枠組みに基づき発言内容を整理・分類する分析手法

成果物として、治療効果の持続性、長期的な安全性に関する不確実性、日常生活への影響など、患者および関係者が重視する主要な治療特性のリストが抽出された。

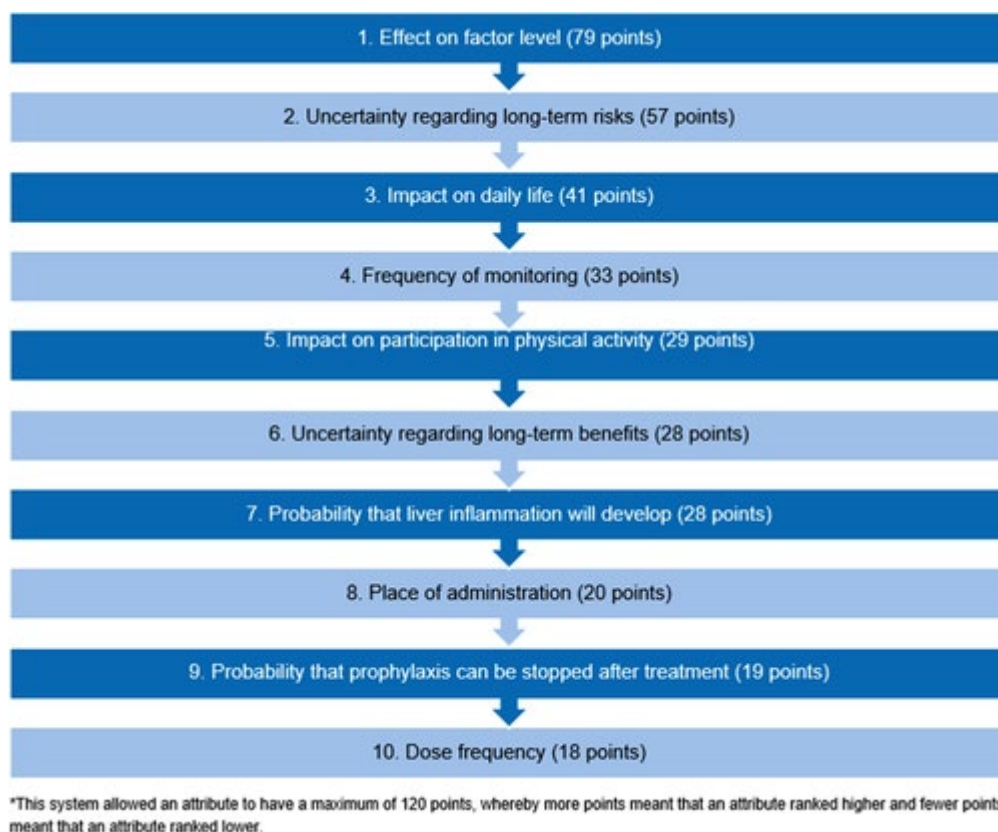


図 18 研究で抽出された遺伝子治療に関する主要な治療特性の上位 10 項目と順位（数字が順位に対応）

これらの成果は、DCE における特性および水準を設定する際の重要な根拠情報として用いられており、定量的選好評価の設計段階における意思決定を支えている。

一方で、対象者数や地域が限定されている点、また一部の特性が事前に定義されている点については、結果の解釈にあたって留意が必要である。

例3) SURMOUNT-4 試験における試験終了後面接<sup>27)</sup>

定性的研究は、臨床試験の結果を患者の視点から補足・解釈する目的でも活用されている。SURMOUNT-4 試験では、試験終了後の参加者を対象とした企業主導の面接が実施され、試験参加体験に関する定性的データが収集された。本研究では、試験参加者に対してオンラインで半構造化面接を行い、治療による変化、日常生活への影響、治療に伴う負担などについて意見を聴取した。得られた逐語録データは内容分析\*により整理され、参加者が語った体験や認識が概念として整理された。

\*: 面接で得られた発言を体系的にコード化・分類し、参加者が経験した内容や認識を概念ごとに整理する質的分析手法。本研究では、研究者による過度な解釈や抽象化を避け、参加者の発言内容に沿って概念化することで、患者経験の全体像と各概念の出現状況を整理している。

成果物として、体重変化に伴う生活上の改善、エネルギーや活動性の変化、有害事象に対する認識、治療継続に関する意向など、患者が実際に経験したベネフィットおよび負担に関する主要テーマが下表のとおり整理された。

| Domain                       | Experience              | Count<br>(N = 86) | Quote(s)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------------------------------|-------------------------|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Perceived treatment benefits |                         |                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Physiological                | Weight loss             | 86                | "I definitely had started losing weight, um, in those first 36 weeks." (S5-37-F-TZP)<br><br>"My weight came off very quickly, I mean, within, within the first, I mean, four weeks." (S1-05-M-PB)                                                                                                                                                                                                                                                                                                          |
| Physiological                | Appetite decrease       | 78                | "Almost immediately, it suppressed my appetite. I would be eating and then suddenly get full and stop, where I've never experienced that in the past." (S10-14-M-PB)<br><br>"I felt like I responded quickly when I initially started it, as far as having, um, a decrease in my appetite, um, eating smaller portions, just because I wasn't able to eat very much, just due to feeling fuller." (S2-36-F-TZP)                                                                                            |
| Physiological                | Energy level increase   | 53                | "My body goes in cycles, like, as the day gets darker, you know, there's less light, then I get tired more and don't have as much energy. But I, I noticed during—while I was taking the medicine, the entire time, I had, err, lots of energy." (S11-60-F-TZP)<br><br>"Err, within the first 36 week, yes, there were changes. Like I said, um, I were getting more energy because I'm shedding weight. I wasn't so lazy anymore." (S1-76-F-TZP)                                                          |
| Physical appearance          | Improved clothing fit   | 49                | "Well, you know, when you're continually gaining weight, you find yourself having to, you know, adjust to the size. You can sometimes find yourself buying more clothes and before you know it, you have a whole new size of clothes." (S14-93-M-PB)<br><br>"I pulled out clothes I haven't wore since before having my kids, and they actually fit me." (S3-75-F-TZP)                                                                                                                                     |
| Emotional well-being         | Increased positivity    | 47                | "[I experienced] less depression, not that I had—I, I didn't have significant depression to begin with, but just less feelings of feeling down." (S2-36-F-TZP)<br><br>"I just felt a lot more positive." (S3-64-F-PB-D)                                                                                                                                                                                                                                                                                    |
| Emotional well-being         | Increased confidence    | 41                | "Like, I lost weight, um, I gained confidence back, um, just overall a happier person, um, participating in all types of, um, situations." (S14-90-F-TZP)<br><br>"No, I mean, it was, it was great. Um, I, I had confidence, um, I, I felt very good and, and happy about what I was doing and, and the direction I was heading in." (S13-732-F-PB)                                                                                                                                                        |
| Physical functioning         | Improved walking        | 36                | "... some of the [...] impact was just with, you know, being able to have better stamina and walk without, you know, being winded or, um, you know, keep up with other people's pace." (S14-88-F-PB)<br><br>"But I think by month two, I realized that I was walking better, walking further." (S6-83-F-TZP)                                                                                                                                                                                               |
| Perceived negative impacts   |                         |                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Physiological                | Loss of muscle tone     | 3                 | "I guess the only negative would've been I probably lost some strength. Err, I—I'm pretty active physically. I play a lot of what, err, what we call here as senior softball, and I probably lost some of my batting power, err, until we got to the end of the program." (S6-69-M-PB)                                                                                                                                                                                                                     |
| Physical appearance          | Increased loose skin    | 3                 | "The only bad thing is, um, like, my arms got a little more, um, I don't know, oh, saggy, so I've been trying to work on my arms." (S13-52-F-TZP)                                                                                                                                                                                                                                                                                                                                                          |
| Emotional well-being         | Increased anxiety/worry | 2                 | "Towards the end of the 36-week period, I was starting to get a little worried about how much weight I was continuing to lose. It was, actually, kind of, funny, you know, because here all this time, I've been worried about I'm eating too much and trying to find the right diet and trying to lose weight. Now all of a sudden, I'm going, 'You know, last time I checked in I lost eight more pounds' and I'm to the point where I'm starting to worry about losing eight more pounds." (S2-26-F-PB) |

これらの結果は、臨床試験で得られた定量的アウトカムを患者経験の観点から解釈するための補助情報として活用でき、治療効果や患者負担の説明、さらには将来の試験設計や患者負担軽減策の検討に示唆を与える。

一方で、著者らは、試験終了後の面接の参加者が試験参加者全体を代表していない可能性について言及しており、結果の一般化にあたっては留意が必要である。

#### 例4) 円形脱毛症における患者報告アウトカム（PRO）開発<sup>28)</sup>

患者の経験に基づく内容妥当性を確保する観点から、定性的研究は PRO 開発の初期段階において重要な役割を果たすことが、FDA の PRO ガイダンス<sup>29)</sup>等により示されている。円形脱毛症を対象とした本研究では、患者が重視する治療アウトカムを反映した PRO を開発することを目的に、製薬企業が関与する一連の定性的研究が実施された。

本研究では、文献レビューや既存尺度の検討に加え、患者に概念抽出（concept elicitation）面接を行い、症状および生活への影響に関する主要概念を整理した（図 19）。この概念モデルは、文献レビューおよび既存アウトカムの整理を踏まえ、円形脱毛症患者への概念抽出面接から得られた発話を定性的に分析・統合することで作成された概念モデルである。患者が経験する症状および生

活への影響が構造化されて示されており、後続の PRO 項目生成の基盤となるアウトカムの全体像を表している。

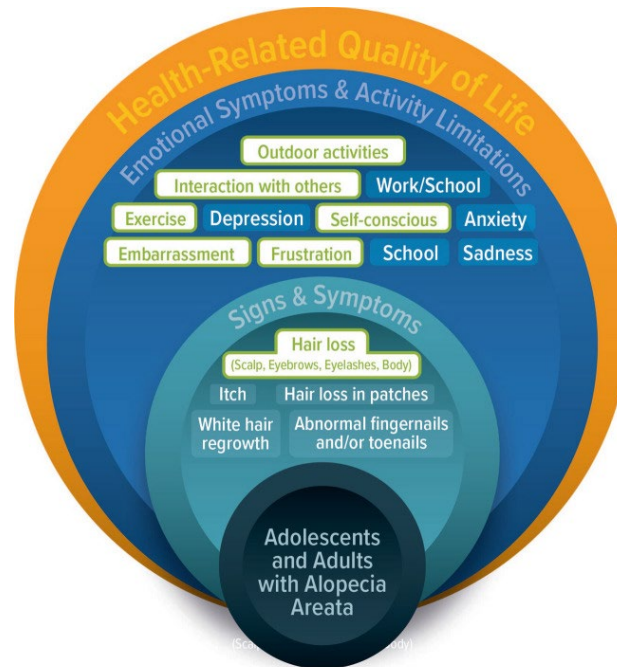


図 19 円形脱毛症の概念モデル

その後、これらの概念に基づいて項目案を作成し、認知的デブリーフィングインタビューを通じて各項目の理解可能性や解釈の一貫性を確認することで内容妥当性を高めている。ここでの認知的デブリーフィングインタビューとは、作成中の質問票（PRO 尺度など）のドラフト項目が、患者の実際の体験や優先事項を正確に反映し、適切に評価できているかを確認するために行われる面接調査のことを意味している。

最終的に、円形脱毛症患者が優先するアウトカムを反映した複数項目からなる PRO 尺度が下表のとおり作成された。

**Table 7** Alopecia areata patient priority outcomes instrument concepts

| <b>Draft concepts evaluated in cognitive debriefing interviews</b> | <b>Final concepts in pilot questionnaire</b>                   |
|--------------------------------------------------------------------|----------------------------------------------------------------|
| Item 1. Hair loss, scalp                                           | Item 1. Hair loss, scalp                                       |
| Item 2. Hair loss, eyebrows                                        | Item 2. Hair loss, eyebrows                                    |
| Item 3. Hair loss, eyelashes                                       | Item 3. Hair loss, eyelashes                                   |
| Item 4. Hair loss, body hair                                       | Item 4. Hair loss, body hair                                   |
| Item 5. Problems with fingernails or toenails                      |                                                                |
| Item 6. Itching on scalp                                           |                                                                |
| Item 7. Self-consciousness about hair loss                         | Item 5. Self-consciousness about hair loss                     |
| Item 8. Embarrassed about hair loss                                | Item 6. Embarrassed about hair loss                            |
| Item 9. Sad about hair loss                                        | Item 7. Sad about hair loss                                    |
| Item 10. Frustrated about hair loss                                | Item 8. Frustrated about hair loss                             |
| Item 11. Limited outdoor activity because of hair loss             | Item 9. Limited outdoor activity because of hair loss          |
| Item 12. Limited physical activity because of hair loss            | Item 10. Limited physical activity because of hair loss        |
| Item 13. Limited interactions with others because of hair loss     | Item 11. Limited interactions with others because of hair loss |

本研究により作成された PRO は、臨床試験におけるアウトカム測定手段として活用可能であり、患者にとって意味のある治療効果を評価するための基盤となる。

ただし、定性的研究による項目生成のみで PRO が確立するわけではなく、量的検証や妥当性評価を含む追加的な検討が必要である点に留意する。

以上の 4 つの例が示すように、

- 定性的研究は患者選好の構造理解
- 定量的選好評価や試験設計への橋渡し
- 臨床試験結果の解釈補強
- PRO をはじめとする評価指標の開発

など、医薬品開発のさまざまな場面で具体的な成果物として活用されている。研究目的に応じて適切な分析方法と成果物を設計・整理することが、定性的研究を医薬品開発に有効に活かすうえで重要である。

## Appendix. 1 定性的研究に特有の概念や考え方の解説

定性的研究を行う際、定量的研究とは異なる前提や独特の考え方を採ることがある。本章では、定性的研究で頻出する用語を取り上げ、概要と留意点を整理する。

- 理論的飽和状態<sup>12)</sup>

繰り返し調査を実施しても新たな関連情報や重要情報が現れない時点のことを理論的飽和状態（理論的飽和）と呼ぶ。定性的手法においては、データ収集および/または解析を中止するための原則としてしばしば使用される。解析計画においては、症例数設計の記載の際は症例数を事前規定できないため、該当箇所において例えば「理論的飽和に達するまで参加者募集を続ける。」などと書いておくことが望ましい。一方で、実務上は研究期間や予算の制約が存在することが多く、理論的飽和を厳密に目指した運用が常に現実的であるとは限らない。したがって、理論的飽和を目標とする場合であっても、研究資源やタイムラインを踏まえた「終了基準（どこで区切るか）」を併せて検討しておく必要がある。

### 例1) 理論的飽和評価

理論的飽和の評価は、例えば以下のように、計画されたインタビューを複数の区分に分け、各区分で新規概念の出現状況を比較する方法で整理できる。各段階において新たに出現した重要概念を整理し、追加データから新しい重要概念がほとんど得られなくなった場合には、理論的飽和に達したと判断できる可能性がある。

理論的飽和は、例えば以下の手順で評価できる<sup>12)</sup>。

1. 計画されたインタビューの最初の 25%で得られた概念を、次の 25%のインタビューで得られた概念と比較する。
2. 上記の 1 で得られた概念（計 50%分）を、さらに次の 25%で得られた概念と比較する。
3. 上記のすべてのインタビュー（計 75%分）を、最後の 25%のインタビューと比較する。

最初に計画したサンプル数で理論的飽和に達しなかった場合、または理論的飽和を完全に評価するために特定の患者グループでより多くの情報が必要な場合、さらにインタビューが必要になることがある。インタビューは通常、理論的飽和になるまで行われ、最後のインタビューセットから新しい重要なコンセプトが浮かび上がることはない。

理論的飽和に達したかどうかに関する単一の数値基準は存在せず、新規概念の出現状況に加え、その概念の重要性や研究目的との関連性を総合的に評価し

て判断することが推奨される<sup>30)</sup>。

## 例2) 理論的飽和グリッド

理論的飽和グリッドは、フォーカスグループ等の各回で「既知概念」と「新規概念」がどの程度出現したかを整理する表である。例えば、研究者が事前に 下表<sup>12)</sup>にしめすように Symptom A、B を想定していた場合でも、調査の過程で Symptom C～G が追加的に特定されることがある。ただし、理論的飽和の達成を判断する前に、参加者の人口統計学的特性等をレビューし、対象集団を十分に代表しているか（偏りが過度でないか）を評価することが重要である。

**Table 7. Saturation Table Example**

| Concept   | Group 1        | Group 2        | Group 3        | Group 4 |
|-----------|----------------|----------------|----------------|---------|
| Symptom A | 5 <sup>1</sup> | 4              | 3              | 3       |
| Symptom B | 4 <sup>1</sup> | 3              | 5              | 3       |
| Symptom C | 3 <sup>1</sup> | 4              | 4              | 3       |
| Symptom D |                | 2 <sup>1</sup> | 1              | 2       |
| Symptom E |                | 3 <sup>1</sup> | 2              |         |
| Symptom F | 4 <sup>1</sup> | 3              | 5              | 4       |
| Symptom G |                |                | 1 <sup>1</sup> |         |

N=6 patients per focus group; 24 total patients

<sup>1</sup> Highlighted cells indicate the first time a concept is mentioned, and numbers within the cells indicate the number of patients in that focus group endorsing the concept.

## ● グラウンデッド・セオリー・アプローチ<sup>31)</sup>

グラウンデッド・セオリー・アプローチ（GTA）は、グレイザーとストラウスにより体系化された定性的研究の方法論であり、データから理論を生成するためのアプローチである。先に固定的な仮説を設定するのではなく、得られたデータに基づき帰納的に仮説を構成・修正していく点に特徴がある。GTAにより、インタビュー等のフィールドワークで得られたデータから新たな理論を生成したり、複雑な概念を体系的に整理したりすることが可能となる。一方で、方法論の発展には現在も議論があり、ストラウス／コービン派、グレイザー派等の流派が存在する点には留意が必要である。

以下に、本章で紹介した実例のうち、がん患者の治療選好研究<sup>25)</sup>は GTA の典型的な流れを踏襲しているため、そのプロセスを例として一般的な GTA 研究の流れを示す。

### 1. 研究課題の設定と理論的枠組みの整理

目的：がん患者が治療や医療体験の中で何を重視し、どのように意思決定しているかを明らかにする。

枠組み：既存の理論や仮説に縛られず、患者の経験をそのまま捉えることを

重視。GTA の帰納的アプローチを採用。

2. サンプルング（理論的サンプルング）

対象：コロンビア・メデジンの成人がん患者（診断から 2 年以内）。

方法：理論的飽和に達するまで、患者を追加リクルート。患者の年齢・性別・疾患タイプなど多様性を考慮。

具体例：15 名の患者を対象に、計 40 回（24 時間 48 分）のインタビューを実施。

3. データ収集（半構造化インタビュー等）

手法：半構造化インタビューを複数回実施。患者が自由に語れるような質問設計。

内容：診断～治療の経過、初回受診、主治医との関係、肯定的・否定的経験、治療方針の選択場面など。

具体例：「治療で最も重視することは何ですか？」「医師との意思決定でどんな点を大事にしましたか？」など。

4. 逐語録作成と初期コード化

プロセス：インタビューを録音・文字起こしし、逐語録を作成。

初期コード化：患者の発言から「医療者の特性」「治療目標」「医師-患者関係」などのコードを付与。

具体例：患者が「医師がよく話を聞いてくれた」「治療の選択肢を説明してくれた」などの発言を「傾聴」「情報提供」などのコードに分類。

5. カテゴリ化と比較分析（段階的コード化）

段階的コード化：オープンコード化（初期コード化）→軸足コード化（コード間の関係性整理）→選択コード化（主要カテゴリの抽出）。

比較分析：参加者間で共通点・相違点を比較し、カテゴリを統合。

具体例：医療者の特性（専門性＋共感）、医師-患者関係（委ねる志向／自己決定志向）、治療目標（延命／QOL 重視）などのカテゴリを整理。

6. 理論的飽和の確認

プロセス：新たな概念が出現しなくなった時点でデータ収集を終了。

具体例：40 回のインタビューを通じて、主要な選好カテゴリが網羅されたことを確認。

7. 概念枠組みの構築

成果物：患者選好の全体像を示す概念モデル（図として整理）。

具体例：医療者の特性、医師-患者関係、治療目標の 3 領域が相互に関連し、患者の意思決定に影響する構造を図示。

8. 信頼性・妥当性の確保

方法：複数の研究者でコード化の一致度を確認。分析過程や判断の根拠を文

書化。

具体例：コード化の合議、カテゴリの再検討、分析手順の記録。

## 9. 結果の応用

活用：得られた概念枠組みをもとに、後続の定量的選好調査（DCE 設計等）や評価項目の検討に活用。

具体例：患者選好の構造を整理した概念モデルが、DCE の特性候補や臨床試験の評価項目選定の基礎情報となる。※他の実例でも、GTA ほど厳密ではないものの、「患者インタビュー→コード化→カテゴリ化→成果物（特性リスト・テーマ・項目）」という定性的研究の標準的な流れは共通している。

## ● トライアングレーション<sup>31)</sup>

異なった方法、研究グループ、調査地、調査時期、理論的視角などを組み合わせて得られた知見の頑健性を高めることをトライアングレーションという。デンジンが 1970 年代にこの概念を取り上げて、社会調査のためのより体系的なアプローチとして発展させた。彼は 4 タイプのトライアングレーションを区別している<sup>32)</sup>。

### ① データのトライアングレーション

異なったデータ源のこと。データ産出のための異なった方法の使用とは区別される。「データ・トライアングレーションのサブタイプ」として、デンジンは時間、空間、人物を区別する。これは現象を、異なった時点や場所において、あるいは人物を通して研究するということである。これはグレイザーとストラウスの理論的サンプリングの戦略と近い考え方である。共に、人物、グループ、調査地、調査時期を、目的をもって、体系的に組み入れていくという前提に立つからである。

### ② 研究者のトライアングレーション

ひとりの研究者の偏り（バイアス）を見つけ出したり、最小化したりするために、異なった観察者ないしインタビュアーが研究に加わるというものである。これは単なる作業分担や、助手への単純作業の委任ということではなく、異なった研究者が対象や結果に与える影響を体系的に比較するということである。

### ③ 理論トライアングレーション

「多元的な視点と仮説を念頭に、データにアプローチすること」がその出発点であり、「多様な理論的観点が、それらの有用性と力とを吟味するために並列される」<sup>32)</sup>。そうすることで、知を産出する可能性を拡大することを目指すのである。

### ④ 方法のトライアングレーション（二つのサブタイプに分けられる）

A) 方法内トライアングレーション

(例) 定性的研究以外で伝統的なのは、ある質問表の中のひとつの項目を異なった下位尺度 (subscale) を用いて測定する。

B) 方法間トライアングレーション

(例) 質問表と半構造化インタビューとの組み合わせ

詳しくは Flick らの書籍<sup>31)</sup>を参照されたい。

## Appendix. 2 定性的研究における解析計画の記載例

本章で述べたような研究を実施する場合、解析計画書を作成し、解析方針を事前にすり合わせておくことが推奨される。本節では、解析計画書に記載すべき事項の章立ての例を示す。

- 研究参加者

目標症例数とその根拠についても記載する。なお、理論的飽和に達するまで研究を続ける方針など、目標症例数が事前に明示できない場合はその旨を記載する。

- データクリーニングおよび管理手順

インタビュー書き起こしの方針やコード化の方針（何人でコードするか、どのような方針でコード化するか、得られたコード化済みデータをどのように解析するか）等を記載する。コード化マニュアルを含めても良い。

品質管理の方針についても、再現性を担保するために記載しておくといよい。

- 解析方針（進め方）

図 4-2 に示したような解析（再解析も含む）と報告のフロー、具体的な解析内容、解析に必要な仮定、欠測をはじめとするデータの取り扱い等を記載する。

- 出力レイアウト（Mock）

グラフ、表など、解析を出力結果としてどのような形式でまとめるかを記載する。

参考までに、PFDD GD1<sup>11)</sup>では (a) 収集されるデータ；(b) 研究目的に対応するために実施される分析、ならびにそれらの分析に必要な仮定；(c) データのクリーニングおよび管理手順；(d) 該当する場合のデータ変換；および (e) 研究結果の提示方法（例：グラフ、表）を解析計画に示すことを推奨している。

## 5. 定量的研究について

本章では、患者選好研究における定量的研究について、その目的や意義、研究設計の考え方、代表的な手法と留意点を体系的に解説する。定量的研究は、患者の価値観や選好を定量化し、複数の治療特性間に存在するトレードオフを明示することで、規制当局や HTA、開発戦略における意思決定を支える重要な役割を担う。

定量的患者選好研究に関しては、IMI PREFER をはじめ、ISPOR が公表している Good Research Practices 報告書など、国際的に広く参照されている指針やチェックリストが存在する<sup>33-35)</sup>。本章の内容は、これらの文献で整理された考え方や推奨事項を踏まえつつ、実務で理解・活用しやすい形でまとめたものである。

### 5.1. 定量的研究の概要

#### 5.1.1. 定量的研究の目的と意義

定量的研究は、患者が治療の各特性をどれだけ重要と感じているかを「相対的重要度」として数値化し、特性間のトレードオフ（例：効果改善と副作用増加の許容度）を明確にすることができる。

そのため、治療の効果や副作用など複数の特性の間でトレードオフが存在し、優先順位が患者の価値観によって左右されるような preference-sensitive な状況では特に有用である。

たとえば臨床試験の設計段階では、どのエンドポイントが患者にとって最も意味があるかを定量的に示すことで、試験に含めるべきアウトカムの選択を裏付けることができる。また、BR バランスが明確でない状況でも定量的研究は有用である。効果は期待されるものの長期安全性などのリスク情報が不足している場合、どのベネフィットを患者が最も重視し、どの程度の改善であれば“不確実なリスクがあっても受けたい”と考えるのかを、相対的重要度として定量化できる。

#### 5.1.2. 定量的研究の標準的な進め方

定量的研究の流れの全体像を図 21 に示す。

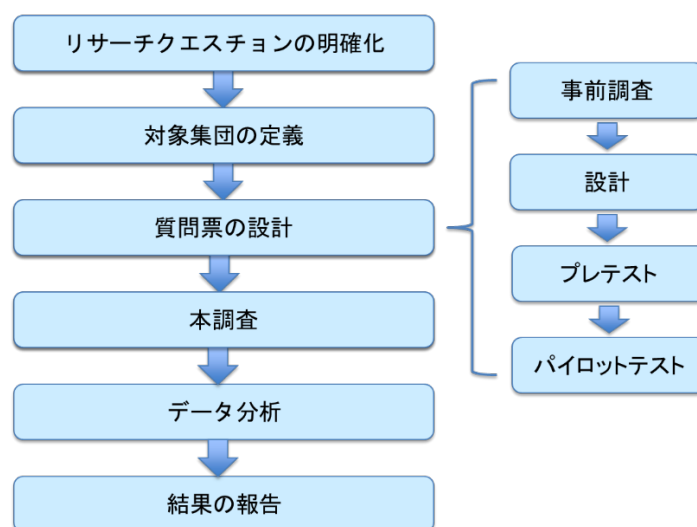


図 20 定量的研究の流れ

## 5.2. 定量的研究実施の留意事項

### 5.2.1. リサーチクエスションの明確化

定量的研究を実施するにあたり重要なことは、通常の臨床研究と同様に、最初に「この研究で何を知りたいのか？」という RQ を明確に定義することである。

定量的研究は単に患者の意見を収集するものではなく、特定の意思決定を支えるためのエビデンスを生成する研究であるため、「どの意思決定者（規制当局、HTA、製薬企業）が、どのような状況で、どのような判断を下す際に本研究結果を利用するのか」、即ち「研究目的」を最初に特定する必要がある。この過程では、既存情報では不十分な点を明らかにし、当該判断が「患者選好が影響する = preference-sensitive」かどうかを慎重に検討することが求められる。また、RQ の設定には、誰の選好を測定すべきかという対象者の明確化が不可欠である。疾患領域や治療状況、経験の有無、重症度など、意思決定に関連する患者特性を明確に定義し、それらは研究目的と直接結びついている必要がある。

さらに、定量的研究の実施に先立っては、定性的研究や文献調査により、患者が重要視する領域や意思決定で考慮されうる特性に関する仮の選好構造を把握しておくことが重要である。IMI PREFER<sup>16)</sup>でも、文献レビューやインタビュー等の基礎調査を通じて、特性候補や研究目的の妥当性を精緻化してから定量フェーズに進むというプロセスが推奨されている。

以上のプロセスを経て、研究目的、意思決定の文脈、選好の対象となる患者集団、そして本研究が提供する価値を明確化することにより、後続の研究設計（対象集団の定

義、方法選択、質問票設計）が一貫性を持って進められる。RQ の明確化は、研究全体の質と意思決定での活用可能性を左右する基盤となる工程である。

## 5.2.2. 対象集団の定義・募集方法

患者選好研究においても通常の臨床試験と同様に、選択除外基準によって研究対象集団を定義する必要がある。対象集団の定義は、研究の妥当性を左右する中心的要素であり、最初に立てた RQ との整合性を確保することが重要である。

同じ疾患の診断を受けた患者であっても、患者は以下の要因により選好（リスク許容度・利益の優先順位）が大きく異なる可能性がある：

- 病期（軽症～重症）
- 症状の種類（例：パーキンソン病の運動症状 vs 非運動症状）
- 進行スピード
- 既存治療の経験
- 生活背景や心理的要因（リスク選好、テクノロジー受容度など）

そのため、対象集団は研究目的（規制当局対応、HTA 対応、製薬企業内利用）と一致していることが理想的である。たとえば、ある医薬品の BR 評価を規制当局に提出するために研究を行う場合、その医薬品の適応患者（適応疾患・重症度・背景治療などが、臨床試験の参加基準に合致している患者）を研究の対象とすることが望ましい。

一方で、実務上は RQ と完全に一致する集団での研究が難しいケースも多い。希少疾患や特定の治療フェーズに限定される患者では対象者数が非常に限られるだけでなく、がん領域のように疾患の重症度が大きく異なる領域では、患者パネル等のオンライン調査に参加できる患者が“若年・軽症”に偏り、本来医薬品の投与対象となる“高齢・重症”患者の分布が著しく不足することが起こりうる。そのような場合には、

- 完全一致集団を揃えられなくても、分析時にサブグループを区別して扱う
- 両方の募集方法（例えば、臨床試験内からの募集と、オンライン・パネル）を併用

といった代替策が推奨される。IMI PREFER<sup>16)</sup>でも、「理想的な対象と異なる場合は、選択理由とバイアスの考慮を明記すべき」とされている。

対象者の募集方法について、IMI PREFER は 臨床試験内・臨床試験外の両方が可能であり、それぞれの長所・短所があると整理している。各募集方法の利点・留意点を下表にまとめる<sup>36)</sup>。

| 募集方法            | 利点              | 留意点                          |
|-----------------|-----------------|------------------------------|
| 臨床試験内<br>（治験参加患 | ・ 適応患者像との一致度が高い | ・ 選択バイアス（リスク許容度が高め等）がある場合がある |

| 募集方法                | 利点                                                                                                                           | 留意点                                                                                                                           |
|---------------------|------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| 者)                  | <ul style="list-style-type: none"> <li>臨床情報が豊富</li> <li>新薬の新しい機序やモダリティを経験したうえでの選好を得ることができる</li> </ul> →意思決定（規制・BR評価）に直結させやすい | <ul style="list-style-type: none"> <li>施設・地域の偏りが存在する可能性がある</li> </ul>                                                         |
| 臨床試験外：<br>医療機関      | <ul style="list-style-type: none"> <li>診断の確実性が高い（医師紹介・EMRに基づく）</li> <li>重症度や病期などの情報精度が高い</li> </ul> →適応患者像との一致度が高い           | <ul style="list-style-type: none"> <li>手続き（IRB/倫理審査、現場調整）や期間・コストが増すことが多い</li> <li>単一施設は地理・社会経済的に偏る可能性がある</li> </ul>           |
| 臨床試験外：<br>オンライン・パネル | <ul style="list-style-type: none"> <li>迅速・大規模な募集が可能</li> <li>全国的に幅広い患者にアクセス可能</li> </ul>                                     | <ul style="list-style-type: none"> <li>自己申告診断（self-report）であるため診断確度に限界があり、若年・軽症・デジタル利用者に偏りやすい</li> <li>回答率が低いこともある</li> </ul> |
| 臨床試験外：<br>患者会等      | <ul style="list-style-type: none"> <li>疾患理解の深い参加者を確保しやすい</li> </ul>                                                          | <ul style="list-style-type: none"> <li>疾患によっては団体が存在しない場合もある</li> <li>関与度・志向性が高く偏る可能性がある<br/>（一般患者像との乖離）</li> </ul>            |

このように、複数の方法があるものの、それぞれ異なる利点・欠点を持つことから、研究目的に応じて内外を適切に使い分け、限界と偏りを明示することが推奨されている。

また、必要に応じて複数の方法を併用することも考えられる。たとえば MDIC の心不全のケース<sup>36)</sup>では、医療機関（デューク大学医療システム内の紹介）126 例＋オンライン 500 例のハイブリッド募集を実施し、両群の結果を比較したうえで差が小さければ統合解析を行った。こうした設計は、単一チャネル由来の偏りを相殺し、一般化可能性の根拠を強化できる可能性がある。

適応患者像と一致した PPS を目指すなら、確定診断に基づく募集は強力な手段である。一方、速度・コスト・希少性の制約から、自己申告も現実的な選択肢となる。自己申告を用いる場合は、症状・治療歴などの補助情報や募集経路（患者会・医療機関の紹介等）を併用して診断確度を高める工夫が推奨されている。可能であれば両経路を併用し、結果の整合性を検証することも検討する。

### 5.2.3. 質問票の設計

本項では、定量的研究に用いる質問票の設計プロセス（事前調査 → 設計 → プレテスト → パイロット）の流れを記述する。質問票は、意思決定の文脈に整合し、かつ参加者が理解可能な負荷で、トレードオフを適切に誘発できるよう作られる必要がある。質問票の設計上の留意点には、手法に依らず共通する部分がある。本項では共通する考慮事項にフォーカスする。

#### 5.2.3.1. 事前調査（特性・水準の候補の同定）

定量的研究では「複数の特性（ベネフィット、リスク、負荷等）の間で患者がどのようなトレードオフを許容するか」を測定するため、

- 測りたい項目（例：症状改善、30 日死亡リスク、投与方法）＝ 特性 (Attribute)
- その程度（例：「30 日死亡リスク：0.5%、2%、5%」）＝ 水準 (Level)

を検討する必要がある。

特性や水準の候補を検討する方法として、大きく 2 種類存在する。

##### (1) 文献やガイダンスによる情報収集

既存の患者選好研究のエビデンスや専門家の知見を活用し、既存の臨床試験アウトカムや学会ガイドラインの重視ポイント、規制当局が重視しているベネフィットやリスクの各項目などから情報を収集する。

##### (2) 定性的調査

フォーカスグループ、半構造化面接、SNS 分析などにより、患者が実際に「日常生活で困ること」や「治療で重要と感じる要素」を抽出する。

RQ や疾患領域に対する背景知識の量に応じて、両方を組み合わせることもある。

### 5.2.3.2. 設計

#### 5.2.3.2.1. コンテキストの設定

コンテキストは回答者に対して「調査の前提・目的・比較の軸」を共有し、回答者全員が同じ理解レベルから選好判断に入れる状態をつくる役割を持つ。一般に、以下の要素を含めることが推奨されている。

##### (1) 研究目的（今回の結果が何に使われるのか）

- (2) 疾患や対象患者像の要約（誰が）
- (3) 現行標準治療およびアンメットニーズ（どのような臨床状況）
- (4) 今回比較する治療・医療技術の位置付け（どれとどれを比べるのか）

各要素は長文を避け、中立的・平易な表現とする。コンテキストが曖昧だと、同じ設問でも回答者が異なる解釈をして選好の推定が歪む恐れがある。

例) コンテキストの具体例<sup>37)</sup>

COVID-19 治療のための新薬開発に今回の結果を使うことを目的としています。  
あなたは COVID-19 に感染しており、薬剤で治療を受けています。  
あなたの症状は現在軽度です。症状は薬を服用してから5日で消えるでしょう。  
次の2つの経口治療薬の選択肢があったとき、あなたはどちらを選びますか？  
症状の消失やこれら2つの薬剤の症状治療の有効性は同じであると想定されています。

#### 5.2.3.2.2. Alternatives（比較する選択肢）の設定

定量的研究の質問票設計においては、患者が選択タスクで比較する治療の「選択肢 (alternative)」を検討する必要がある。各 alternative は特性とその水準の組み合わせからなる「治療プロファイル」として提示される。

治療プロファイルは、設定した特性・水準の組み合わせから体系的に作成する必要がある。IMI PREFER では、現実の意思決定を模す観点から、opt-out（治療を選ばない）や現状維持の導入可否も含めて検討することを勧めている<sup>16)</sup>。これらのオプションは、「どの選択肢も同価値で選べない」といった一般的に望ましくない、あるいは研究方法として不適切とみなされる選択肢とは異なり、回答者が明確に選好を示す手段として扱われる。しかし医療分野における多くの研究では、治療間の選好を比較することが主な目的であるため、回答者が opt-out または現状維持を選ぶと比較データが得られず、RQ に適さない可能性がある。一方で、新たな治療や製品の潜在的な需要や市場シェアを評価する場合には、「選ばれない」という情報自体が重要であるため、これらのオプションを含めることが有用、さらには必要となることもある<sup>33)</sup>。

さらに、opt-out や現状維持オプションの導入は、これらの選択肢が選ばれなかった際に得られる情報が限られるためデータの打ち切り (censoring) を引き起こし、選好構造の推定精度を低下させる可能性がある。そのため、これらのオプションの採用にあたっては、研究目的・RQ との適合性や統計的影響を十分に考慮することが求められる。

なお、治療プロファイルを構築する際には、実験計画法を用いて特性と水準の組み合わせを最適化することが基本である（詳細は 5.3.1.1 章参照）。注意点として、表明選好法では仮想の薬剤を治療プロファイルとして置くことが基本的な考え方であり、実

在の薬剤を治療プロファイルとして置くことは標準的な手法ではない。

#### ■コラム：表明選好法と顕示選好法<sup>38)</sup>

顕示選好法：人々が実際に行った行動（購買や移動手段の選択など）を観察して、その人の選好を推定する方法

表明選好法：アンケートやインタビューを通じて、ある仮想的な状況に置かれた場合にどのように行動するかを尋ね、その回答から選好を推定する方法

|      | 顕示選好                                                                                                                                                  | 表明選好                                                                                                                                                                           |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 選好情報 | <ul style="list-style-type: none"> <li>・ 実際の行動結果に基づいている。</li> <li>・ 現実の市場で観察される行動と整合的である。</li> <li>・ 得られるのは基本的に「実際に選ばれた選択肢 (Choice)」のみである。</li> </ul> | <ul style="list-style-type: none"> <li>・ 仮想的な状況下での回答にもとづく表明である。</li> <li>・ 現実の行動と一致しない可能性がある。</li> <li>・ 「順位付け (Ranking)」「評価 (Rating)」「選択 (Choice)」など多様な形式の情報が得られる。</li> </ul> |
| 選択肢  | 現実存在する選択肢のみが対象となる                                                                                                                                     | 現実存在しない選択肢や条件も設定できる                                                                                                                                                            |
| 特性   | <ul style="list-style-type: none"> <li>・ 特性値を正確に観測できない可能性がある。</li> <li>・ 特性の水準の範囲が現実観測できる値に限定される。</li> <li>・ 特性間に共線性が生じやすい。</li> </ul>                | <ul style="list-style-type: none"> <li>・ 特性を研究者が設定・提示するため、特性値を正確に把握できる。</li> <li>・ 特性の水準の範囲を任意に広げることができる。</li> <li>・ 特性間の共線性をコントロールすることができる。</li> </ul>                        |
| 回答数  | 通常、回答者 1 名につき 1 つの選択データが得られる                                                                                                                          | 回答者 1 名から得られる情報量が多い                                                                                                                                                            |

#### 5.2.3.2.3. 特性の設定

特性は、患者が治療を選ぶときに重視するポイント（利益、リスク、負担、利便性など）である。特性には投与経路、投与頻度、投与方法、治療を受ける場所（例：入院 vs. 自宅）など、健康に直接関係しない要素も含めることができる。

「事前調査」で特定した特性の候補に対して、以下のような観点で、認知負担と識別性の観点から“必要十分な数”に絞る。なお、特性に含める項目は意思決定者（規制当

局、医療提供者など）によって重要視する内容が異なるため留意が必要である。

- (1) 関連性：RQ や意思決定の文脈との関連がない特性は削除
- (2) 非重複性：意味が被る特性は統合
- (3) 操作可能性：質問票に落ちる明確な水準の表現が可能か
- (4) 可読性：患者に理解できる語彙・単位か
- (5) 選好の独立性（Preferential independent）

#### ■コラム：選好の独立性（preferentially independent）

「選好の独立性（preferentially independent）」とは、ある特性に対する選好が、他の特性に関する判断によって影響を受けないことを指す。

（例）晩ごはんの組み合わせ

主食（パン／ごはん）と飲み物（ワイン／日本酒）という2つの“特性”があるとき、多くの人は「パンにはワイン」「ごはんには日本酒」という組み合わせの好みを持つだろう。もし主食の選び方が飲み物の選び方に影響しているなら、この2つの“好み”は独立ではないと言える。ここで言う「独立」は、どちらか一方の条件（主食の種類）が変わっても、もう一方（飲み物）に対する“好みの強さ”が体系的にねじれないことを指す。つまり、特性ごとの“重み”を別々に比べられる状態が「選好の独立性」である。

#### 【よくある混同：特性の“値”の独立性とは別問題】

医薬品の評価では特性の“値”同士が関連していることがよくある（例：有効性と安全性が用量に応じて同方向に変わる、生存期間（OS）と奏効が統計的に関連する、など）。選好が独立であれば、値は独立でなく相関があっても、患者選好研究上は問題ない。

#### 5.2.3.2.4. 水準の設定

水準は各特性に対して提示する具体的な“値”であり、通常各特性に2~4水準を設定する。よい水準設定は、参加者が直感的に理解でき、かつ順位付けやトレードオフの判断を引き出すことができる範囲・粒度で設計される。設計時に検討することが推奨される項目を以下に記載する。

(1) データ型：カテゴリか連続量かを明確にする

- 1) カテゴリ（離散）：例）「投与経路：経口／皮下」「遠隔調整：あり／なし」
  - 2) 連続：例）「30日死亡リスク：0.5％／2％／5％」「回復期間：2日／7日」
- データ型は特性に応じて設定することで問題ないが、患者にとって解釈しやすい刻みで提示し、全設問で単位を統一することが推奨される（％と“100人中〇人”の併記など）。これは理解可能性と認知負担の軽減につながる。

(2) 絶対値か、現状からの改善（差分）か

絶対値の提示（例：「年間の入院回数：2 回」）は、指標の意味が直感的で、他特性との比較もしやすくなる。

一方で、現状からの改善（差分）を使う場合は、「基準（ベースライン）が何か」を必ず明示しないと誤解を招く恐れがある。教育モジュールで、“いまの状態（現状）”と“改善量”の関係を図示してから本番に入ると安全である。

(3) 相対リスクや割合ではなく、絶対値の変化で記述

相対リスク（例：「リスクが 30%減」）よりは、絶対値の変化（例：「5%→3.5% に低下＝100 人中 5 人→3.5 人」）の方が、誤解が少ないため。

また、水準は現実的、または参加者に“現実的だ”と認識される範囲で設定することが推奨される。実臨床で観察される範囲や開発中治療の合理的な予測値、公表された既存治療のデータを参考に決定することが望ましい。

さらに、解釈可能で、判断（順位付け・トレードオフ）を促すのに十分な差（幅）を持たせることも重要である。差が小さすぎる（例：0.5% vs 0.6%）と識別が難しく、逆に大きすぎると非現実的に感じられる。

例) 安全性（有害事象）に関する「水準」の設定例

有害事象に関する水準設定は様々な「軸」により設定できる。いくつかの例を以下に記載する。

- (1) 発生確率（絶対値）：(例) 30 日以内の重篤な出血：0.5%/2%/5%
- (2) 重症度：(例) 出血の重症度：軽度/中等度/重度（入院が必要）
- (3) 持続期間/機能影響（QOL 低下の程度）：(例) 視界のぼやけが続く期間：0 日/3 日/7 日

何を水準化すべきかは、RQ や比較対象との差別化ポイント、患者が重要と考える視点で決める。

- RQ との整合

例) 「標準治療より短期の重篤出血が多いか」を知りたいなら、30 日重篤出血の頻度を軸に。

「日常生活の支障がどれほど受け入れられるか」を知りたいなら、持続期間や機能低下の程度を軸に。

- 比較対象との差別化ポイント

例) 新規製剤は「30 日死亡は低い、入院合併症がやや多い」であるならば、30 日死亡と入院合併症ともに頻度で“差が見える軸”を選ぶ。公表データや開発品の予測値で水準レンジを設定する。

- 患者が重要と考える角度

定性的研究で、患者が“どの側面のつらさ”を語ったか（痛みの強さ/長さ、入院の有無、職業・家事への影響 など）を水準軸に設定する。

#### 5.2.3.2.5. 認知的負担とキャパシティへの配慮

質問票の設計では、回答者の認知的キャパシティに適合した情報量・表現・タスク難度に調整することが不可欠である。誤読・早とちり・単純化ヒューリスティックを未然に防ぎ、本来の選好を引き出すため、設問は「理解しやすく、偏らず、簡潔」でなければならない。なお、単純化ヒューリスティックとは人間が複雑な情報や不確実な状況下で意思決定を行う際、脳への負荷を減らし、素早く結論を出すために用いる「思考の近道」や「経験則」のことであり、過度に思考を簡素化することで、患者自身の真の選好と異なる回答をしてしまうことを指す。

このように、設計段階で理解可能性・負荷管理・提示方法を意識することは、IMI PREFER でも推奨されている。設計に当たっては、ISPOR Conjoint Analysis Good Research Practices (Bridges et al., 2011)<sup>33)</sup>も参考にすることができる。

以下では、IMI PREFER<sup>16)</sup>や Bridges ら (2011)<sup>33)</sup>の推奨事項から、認知的負担とキャパシティへの配慮として考慮すべき点を示す。

##### 【基本原則（設計指針）】

#### 1. 対象者中心の設計：リテラシー適合・表現最適化

対象集団の年齢、教育歴、健康リテラシーなどを踏まえて語彙・視覚化・タスク難度を調整する。

定義や説明は平易な言葉（plain language）で記述し、初回提示時に加えて調査全体を通じて参照可能に配置する。用語・単位・桁・表記（%/回数/期間など）は一貫させ、恣意的解釈を最小化する。

さらに、Bridges ら (2011) は、特性や水準の定義を明確かつ一貫して示し、回答者が過度な推測を行わないようにすることを強調している。たとえば「軽度/中等度/重度」の言葉が状況により回答者ごとに異なるイメージを喚起しうることから、追加的な説明を設けることが望ましい。

#### 2. 少数精鋭の情報提示：情報量の最適化

1画面に提示する情報量は必要最小限に絞る。DCE では特性数は4-6、各水準は2-4を目安に設定し、画面あたりの処理負荷と単純化ヒューリスティックへの依存を低減する。専門家レビュー・予備調査を通じて重複・周辺的なものを削減し、認知キャパシティに適合させることが推奨される。

#### 3. 数値・確率情報の明確化

水準はカテゴリ・連続値・確率のいずれでも設定できるが、単一値で明確に定義する。Bridges ら (2011) は、範囲表現（例：5-10%）は、回答者が主観的に補完・再符号化し曖昧さを導入するため避けることを推奨している。

また、率・確率・頻度を混在させず、桁・単位の表記を統一することも重要である。必要に応じて、%表示に加えて「100 人中○人」を併用することで理解を補助する。また、アイコンアレイやリスクグリッドなど視覚化を活用することも有用であろう。

#### 4. 教育ツールの活用

理解と動機付けを高める教育ツールを、負担を増やさない範囲で組み込むことが推奨される。例えば以下のようなものが考えられる。

- 視覚化ツール：図示・アイコンアレイ・比較図など
- シナリオ：状況をイメージしやすくする簡潔なコンテキスト提示
- 練習問題：本番と類似した選択課題で、回答者が形式・切り替えルールに慣れるための事前演習

なお、教育ツール自体が長過ぎる／複雑すぎると逆に認知負荷を上げるため、長さ・情報量・複雑性を最適化することが必須である。

#### 5. 段階的学習＋理解チェック

導入→練習→本番の三段階で、理解度クイズや注意テストを挟み、本人の理解を確認する。

特に、選択肢の理解や説明文の読解、リスク／確率概念の理解を確認することで、回答品質と結果解釈の妥当性が向上することが期待される。

#### 6. 実測に基づく反復修正

後の章で示す、プレテスト（認知インタビュー）やパイロットテストで、回答時間・誤読・迷いやすい表現・タスク放棄などを観察し、語彙・図・タスク数を繰り返し調整する。Bridges ら（2011）は、タスク数は通常 8-16 が妥当とし、ブロック化の活用も推奨している。

#### 7. 認知的脆弱性への配慮

認知障害（例：認知症、神経筋疾患など）を持つ対象では、文言・図表・タスク構造をさらに簡素化し、プレテストを徹底する。場合により DCE 自体が理解困難な可能性があるため、介護者を研究の対象集団に含め、当該集団の選好把握の補助情報を得ることも検討する。

#### 例）視覚化の例

Table 1 Attribute descriptions and levels used for the DCE

| Attribute                                      | Attribute description                                                                                            | Level 1                                                                                                                                                                     | Level 2                                                                                                                                                                       | Level 3                                                                                                                                                                            |
|------------------------------------------------|------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Method of delivery                             | The form, indication, and frequency of the treatment                                                             | <br>Oral medication when experiencing a headache (acute treatment)                         | <br>Oral medication every other day to prevent headaches from occurring                     | <br>Self-administered injection once a month to the stomach to prevent headaches from occurring |
| Reduction of pain when experiencing a headache | The reference level for 100% pain will be the maximum amount of pain the patient has experienced from a headache | <br>Pain reduced by 10% from the maximum                                                   | <br>Pain reduced by 30% from the maximum                                                    | <br>Pain reduced by 50% from the maximum                                                        |
| Impact of headache on your daily routines      | Described as a pain level that still allows for the patient to perform daily routines as normal                  | <br>On average, 1 more day per month where you can perform your daily routines (as normal) | <br>On average, 3 more days per month where you can perform your daily routines (as normal) | <br>On average, 5 more days per month where you can perform your daily routines (as normal)     |

図 21 質問票の視覚化の例<sup>39)</sup>

## 5.2.3.2.6. 質問票に仕込む「内的妥当性 (internal validity)」の検証

内的妥当性の検証は、得られた選好データが患者の真の価値観を反映しているかを確認するために、理解不足や不注意回答、設計上の欠陥によるバイアスの有無を評価するプロセスである。これらは研究計画書や解析計画書で事前規定し、除外ルールと感度分析まで含めて運用することが推奨される。実務では、安定性（繰り返し質問）／優越ペア（セット内・セット間）／推移性／特性優越／ストレートライニング（同じ位置を連続選択）／回答完了時間の長さ（極端に回答時間が短い）などが用いられる。全ての評価方法があらゆる研究に適用できるわけではなく、研究目的や定量的手法、サンプルサイズや認知的負担に応じて選択する必要がある。

下表に、質問票に仕込むことができる内的妥当性テストの概要（目的／判定の目安／例示）を示す（Johnson 2019<sup>40)</sup> より）。詳細な設問例は Appendix. 3 に集約した。

| テスト例                                  | 目的        | 典型的な判定基準                 | 例示           |
|---------------------------------------|-----------|--------------------------|--------------|
| 安定性 (Stability, repeated question)    | 回答の一貫性を確認 | 同一設問の回答一致                | Appendix. 3① |
| セット内優越ペア (Within-set dominated pairs) | 合理的選択の確認  | すべての特性で明らかに劣っている選択肢を選ばない | Appendix. 3② |
| セット間優越ペア                              | 論理的整合性    | 特性の水準が優越                 | Appendix. 3③ |

| テスト例                         | 目的             | 典型的な判定基準                             | 例示           |
|------------------------------|----------------|--------------------------------------|--------------|
| (Across-set dominated pairs) | を確認            | する場合の整合性<br>を確認                      |              |
| 推移性<br>(Transitivity)        | 経済学的整合<br>性の確認 | $X > Y$ 、 $Y > Z$ なら、 $X > Z$ となるはず。 | Appendix. 3④ |
| 特性優越 (Attribute dominance)   | 非補償的選好<br>の検出  | 特定の特性だけで<br>ほぼ全ての選択を<br>決めない         | Appendix. 3⑤ |

注意しなければならないのは、内的妥当性テストで「非合理」と判定されても、以下のような状況では合理的な説明が成り立つ可能性があることである。

- ほぼ同じ良さに見える 2 つの選択肢  
2 つの選択肢が参加者には「ほとんど同じくらい良い」と感じられる場合、どちらを選んでも筋は通っている。にもかかわらず、同じ質問をもう一度出すと答えが変わるなどの形で、安定性テストの“不一致”として判定されることがある。さらに、同じ質問をアンケート内の近い位置に置くか・離して置くかでも一致率は変わり得る。
- 特性の水準幅の設定による見え方  
医学的に現実的な比較を重視して特性の水準幅（レンジ）を狭く設定すると、その特性は「どちらを選んでもあまり差がない」と感じられやすくなる。その結果、トレードオフが起きにくくなり、差が大きい別の特性だけで意思決定されやすくなり、非補償的選好として観察される可能性がある。  
また、「この特性だけは譲れない」という強い選好がある場合には、他の特性の改善では補いきれず、結果として一つの特性に基づく選択がなされることも合理的に説明できる。
- 学び・疲れ・考え方の簡略化（ヒューリスティクス）  
アンケートが進むにつれてやり方を学ぶ（学習）、長時間で疲れる（疲労）、情報が多くて判断を簡略化する（例：特定の特性だけ見る、同じ位置を選びがち）といったことは自然に起こる。
- 「希望の価値（value of hope）」<sup>41)</sup>  
進行性で重い病気の場合では、特性値だけでは説明しきれない「何かを試みることで自体の価値」が意思決定を後押しすることがある。その結果、優越ペアでの“非合理”として表れる場合があり、後から分析すると、特性とは独立の価値として分離推定できることが報告されている。

このように、単一の妥当性テスト結果だけで「無効」と判断することは危険である。除外有無の感度分析、“非合理”と判定された参加者の特性の確認(年齢・疾患経験など)、調査票の設計(説明文・特性の水準幅・質問配置など)の再点検を行い、総合的に判断することが推奨されている。

#### 5.2.3.2.7. 患者背景項目

定量的な患者選好研究においては、選好の不均一性（preference heterogeneity）を理解し、結果の解釈やサブグループ分析、さらには意思決定での活用可能性を高めるために、患者背景に関する情報を収集することが重要である。選好の不均一性とは、個々の患者レベルで表明される選好が、集団として観察される平均的な選好からどの程度異なるかを指す概念であり、患者選好研究の本質的な特徴の一つである。不均一性を無視すると、平均値だけで意思決定が行われ、一部の患者に不利益をもたらす可能性があるため、規制当局や HTA は、特定のサブグループにおける治療価値やリスク許容度を理解することも重視している。

患者選好研究は、個人の主観的な判断や価値観に基づく選好を測定するものであるため、患者間で選好が異なることは自然である。例えば、重症度の高い患者や治療選択肢が限られている患者では、軽症の患者とは異なるリスク許容度やベネフィットへの重み付けが示される場合がある。このような患者間の選好差は、年齢、疾患重症度、併存疾患、これまでの治療経験、個人的な価値観や嗜好、家族状況などの特性によって一部説明されることがある。こうした影響を研究設計や解析の段階で考慮しない場合、得られた結果の一般化可能性が低下するおそれがある。そのため、患者背景項目を収集し、選好と背景特性との関係を検討することは、研究結果を適切に解釈するうえで不可欠である。

一般に、患者選好に影響すると報告されている主な患者背景項目を下表に示す。これらの患者背景項目は通常オンライン調査や紙ベースでの質問表にて直接収集される。どの程度の選好の不均一性を把握すべきかは、当該研究がどのような意思決定（規制判断、HTA、開発戦略など）に用いることを目的としているかによって異なる。

表 1 患者選好に影響すると報告されている主な患者背景

| 区分         | 主な項目例                                  |
|------------|----------------------------------------|
| 基本的な人口統計情報 | 年齢、性別、教育レベル、雇用状況、年収                    |
| 疾患関連情報     | 診断の有無・疾患名、疾患重症度、罹患期間、併存疾患、(特定の)治療経験の有無 |
| 家族・生活環境    | 同居家族の有無、家族歴、介護者の有無（特に小児や認知症など自         |

| 区分      | 主な項目例                                                                                                                                                    |
|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
|         | 已報告が困難な場合)                                                                                                                                               |
| 行動・生活習慣 | 喫煙状況、服薬状況、医療へのアクセス状況                                                                                                                                     |
| 心理・認知関連 | 健康リテラシー・数的リテラシー、リスク認知・リスク許容度、疾患認識 (Illness Perception)、薬に対する信念 (Beliefs about Medicines)、患者活性度 (Patient Activation)、健康管理に対する態度 (Health Locus of Control) |

#### 5.2.3.2.8. 翻訳

患者の選好に関する資料の翻訳が必要な場合、ISPOR の「患者向け資料翻訳のためのグッドプラクティスの原則(Wild 2005)<sup>42)</sup>」に従うことが推奨される。

このガイドラインでは、翻訳品質の確保と文化的整合性を担保するために、準備、複数名によるフォワード翻訳、翻訳案の統合、逆翻訳、専門家によるレビュー、認知デブリーフィング（患者による理解性確認）など、体系化された 10 の主要ステップが示されている。

初回翻訳は、公式の翻訳会社や専門性を有する翻訳者によって実施されることが望ましいが、翻訳の妥当性は専門家によるレビューだけでは十分でない可能性がある。そのため、Wild ら（2005）でも強調されているように、最終的な翻訳の理解性と文化的適合性を担保するため、対象国の患者による認知デブリーフィングを行い、翻訳文が実際の患者にとって分かりやすく、意図どおりに解釈されるかを確認することが重要である。

以上のプロセスを踏まえることで、翻訳による意味の逸脱を最小限に抑え、原文と概念的に等価な資料を作成し、複数国データの一貫性を確保することが可能となる。

#### 5.2.3.3. プレテスト

プレテストでは、質問が意図通り理解され、現実的で、長さや複雑さが適切か、水準の設定がトレードオフを確認するために十分か、などを小規模の医学専門家や対象集団となる患者（10～20 名程度）で確認する。特に、患者選好研究では、質問内容が正しく伝わるかだけでなく、患者にとって過度な認知的負荷や心理的負担が生じないかを確認することが極めて重要である。

プレテストは一般に、1 対 1 のインタビューや回答者に考えたことを口にしてもらう思考発話法 (talk-aloud exercises) を用いて実施されることが多い。思考発話法では、質問へ回答する際に何をどのように考えたかをその場で言語化してもらうことで、

- どの表現につまずいたか
- 質問の意図と回答者の解釈にずれがないか

を詳細に観察できる。これにより、文章表現や選択肢構造の改善点が明確になるだけでなく、回答に必要な認知負荷の大きさや、参加者が実際にトレードオフを理解し選択できているかを把握することができる。

例) プレテストでの確認例 (Campoamor らの文献<sup>43)</sup> を参考)

- アイストーク: 「説明文はわかりやすかったですか？」
- 特性理解: 「“重篤な副作用”はどんなイメージですか？」
- 図表理解: 「この表の見方は直感的でしたか？」
- 疲労点: 「どの時点で疲れましたか？なぜですか？」
- 改善案: 「言い換えるとしたら、どんな表現が良いですか？」

プレテストは、単なる質問票の改善にとどまらず、研究デザイン全体へのフィードバック機能を果たす。

特に医療分野の調査では、アウトカムや副作用に関する情報が、患者に予期しない心理的・感情的ストレスを与える可能性がある。このようなリスクは、プレテスト時の観察や参加者からの直接的なフィードバックを通じて早期に発見できる。

その結果として、

- 研究資料の修正
- ディストレス (distress) 発生時の対応プロトコルの準備
- 説明方法や順序の調整

など、患者保護の観点から適切な改善を行うことが可能になる。プレテスト活用の具体例は Appendix. 4 に例示した。

また、IMI PREFER<sup>16)</sup>など国際的な取り組みでも強調されているように、質問票設計の段階から患者にパートナーとして関与してもらうことは非常に重要である。

患者（または患者代表）が初期段階からレビューに参加することで、実際の生活経験に基づいた視点、医療専門家には気づきにくい理解上のハードル、認知的負荷や心理的負担に関する知見などを取り込みやすくなる。このような参加型アプローチは、質問票の妥当性と受容性を高めるだけでなく、研究そのものが患者の価値観とニーズに寄り添ったものになることを保証する。

#### ■コラム：患者をパートナーとして巻き込むことの重要性

患者選好研究では、患者を参加者として関与させるだけでなく、研究パートナーとして計画・設計・実施・結果活用の各段階に組み込むことが推奨される。これは、①研究参加者としての倫理的権利の尊重、②研究者が把握しきれない患者にとって重要な治療特

性の発見、③研究の妥当性・実行可能性・受容性の向上、④疾患経験に根ざした患者固有の視点の取り込みという、実務的かつ倫理的な理由に基づいている。

IMI PREFER<sup>16)</sup>は、患者との関わりに関する 10 の基本原則（患者中心性、明確なコミュニケーションと透明性、包括性、相互性、尊重、十分な準備、中立性、負担の均衡、医療への不介入、持続的インパクト）を示し、これらを参加者としての関与にもパートナーとしての関与にも適用すべきとする。これにより、設問や情報提供の明瞭性、調査負担の最小化、信頼醸成が系統立てて担保されることが期待される。

【どのように・いつ関与するのが良いか】

- 関与の方法：患者は研究チームの一員やアドバイザリーボードとして参画し、研究の前・中・後の意思決定に寄与する。関与の強度は、個別論点でのスポット相談から、全期間の埋め込み型相談、さらには研究を共同生産（co-production）する形まで段階的に設定できる。
- 関与のタイミング：
  - 研究目的や RQ の定義の段階で共同作業を行うことにより、曖昧さや難解な表現を早期に除去できる。
  - 設計段階では、対象集団の定義、質問文、特性や水準の定義、提示方法や質問票の長さへの患者レビューが不可欠で、質問票の理解可能性・意味妥当性が高まる。
  - プレテストでは思考発話法で設計の抜けや誤解を洗い出すことができる。

患者パートナーは、研究の対象集団の視点と経験の多様性を全体として代表している必要がある。必要とされる経験・スキル（文書レビュー、翻訳、コミュニティ連携、パイロット等）は研究段階ごとに異なるため、柔軟な役割分担と必要に応じたトレーニングを準備する。

なお、IMI PREFER では患者がパートナーとして参画することを実現するために、わかりやすい言葉遣い、役割の明確化、時間や場所の柔軟性、遠隔参加手段、アクセシビリティ、費用補償、短納期の回避、研修の提供などを手当てすることが推奨されている。PREFER のケーススタディでは、謝金や税・給付への影響案内まで配慮した例が報告されている。

#### 5.2.3.4. パイロットテスト

パイロットテストは、本調査を開始する前に、患者を対象として小規模サンプルで行われる工程のことである。これは、前の節で説明したプレテストとは異なり、実際の研究運用が適切に機能するかどうかを検証するソフトローンチのことを指す。

パイロットテストは以下を確認するために本番と同じ設計（練習・内部妥当性テス

トを含む)で実施される。

- プログラムの技術的な正確性の確認 (設問の分岐、ランダム化、表示ロジックなどが意図通り動くか)
- 手法の妥当性・信頼性の評価(手法が想定通りの回答行動を引き出せているか)
- 提示した特性の水準が十分なトレードオフ行動を生むかの確認 (選択が一方向に偏らず、比較・判断が発生しているか)

パイロットで得られた知見に基づき、本調査に向けて必要な修正を加える。また、パイロットテストの計画や、修正内容・その理由はプロトコル内に明確に文書化することが求められる。

PREFER の複数のケーススタディ<sup>16)</sup>では、パイロットテストが研究の設計の最適化に重要な役割を果たしていた。

- 心筋梗塞患者研究の例では、39 名の慢性期患者によるパイロットを通じて、既存文献および/または専門家の意見に基づく事前仮定を患者データに基づき更新し、提示した特性の水準が十分にトレードオフを誘発していることが確認された。また、一部の特性では予期しない反応が見られたため、DCE デザインが本調査向けに修正された。
- 関節リウマチ調査の例では、約 10 名を対象としたパイロットにより、背景説明文の改善、不要な図の削除、質問文の明確化、過度な長さにつながる項目の削除など、調査票の理解しやすさと適切な負荷に直結する修正が行われた。

#### ■コラム：プレテストとパイロットテストの区別<sup>43-44)</sup>

|          | プレテスト                                                                                                                                      | パイロットテスト                                                                                                                          |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| 目的       | 質問票や研究手法が意図通り理解されるか、曖昧さや誤解がないかを確認すること                                                                                                      | 本調査と同じ条件で小規模に実施し、研究全体の運用性や技術的問題を確認すること                                                                                            |
| 対象       | 医学専門家または対象集団となる患者                                                                                                                          | 本番に近い条件の小規模サンプルの対象集団となる患者                                                                                                         |
| 実施タイミング  | 質問票設計後、研究開始前                                                                                                                               | プレテスト後、研究開始前                                                                                                                      |
| 方法       | 1 対 1 のインタビューや思考発話法                                                                                                                        | 本調査と同じ方法                                                                                                                          |
| 主に確認する項目 | <ul style="list-style-type: none"> <li>・ 内容が意図通り理解されているか</li> <li>・ 質問や演習が明確で現実的か</li> <li>・ 調査の長さが適切か</li> <li>・ 特性の水準がトレードオフを</li> </ul> | <ul style="list-style-type: none"> <li>・ 調査の長さが適切か</li> <li>・ データ収集が期待通りに機能するか</li> <li>・ プログラムやエンコードの正確性 (質問票のロジックや結果</li> </ul> |

|  |                                                                                            |                                                                                                                 |
|--|--------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
|  | <p>誘発する十分な差を持つか</p> <ul style="list-style-type: none"><li>患者に過度な認知的・心理的負担がかからないか</li></ul> | <p>が正しいか)</p> <ul style="list-style-type: none"><li>選好の抽出方法の妥当性・信頼性の評価</li><li>特性の水準がトレードオフを誘発するのに十分か</li></ul> |
|--|--------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|

### 5.3. 各定量的方法の考え方・実施方法と使い分け

5.2 にも記載の通り、IMI PREFER では DCE を含む定量的選好研究において、研究目的および意思決定文脈を明確に意識した設計が重要であると強調されている。いくつかの表明選好研究の手法が提案・比較されているが、Jones (2024)<sup>45)</sup>を参考に、最初に最もよく使われている DCE (Discrete Choice Experiment, 離散選択実験法) を詳細に紹介する。

#### 5.3.1. Discrete Choice Experiment

##### 5.3.1.1. DCE の流れ

DCE は、仮想的な治療選択肢を提示し、その中から最も好ましい選択肢を選ぶという課題を繰り返すことで、患者が治療のどの特性（例：効果、リスク、利便性）をどの程度重視しているかを定量的に把握する手法である。各選択肢は複数の特性 (attribute) とその水準 (level) によって記述され、回答者の選択行動から各特性水準に対する選好の重みを推定する。DCE は医薬品の開発、規制、HTA における意思決定を支援する目的で、ヘルスケア分野の患者選好研究において最も広く用いられている定量的手法の一つである。DCE は、概ね以下の手順で実施される。

##### ① Step 1: 特性と水準の定義

対象とする治療に関連する複数の特性（例：有効性、副作用、投与方法など）を選定し、それぞれに複数の水準を設定する。特性の選定にあたっては、患者にとって意味のある要素であること、かつ特性間の選好の独立性（5.2.3.2.3 参照）が成立することを考慮する必要がある。特性の値同士が独立である必要はないが、特性に対する選好が互いに独立であることが求められる。例えば、COPD 患者を対象とした DCE<sup>46)</sup>では、特性として、「増悪」、「睡眠の質」、「息切れ」、「尿失禁」、「痰の排出困難」、「咳」が選定された。各特性における水準は下表の通りであった。

| Attribute                 |                                     | Level 1                                              | Level 2                                                                                     | Level 3                                                                                      | Level 4                      |
|---------------------------|-------------------------------------|------------------------------------------------------|---------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|------------------------------|
| Shortness of Breath (SoB) | During a typical day                | SoB experienced during strenuous activity            | SoB experienced during light activity                                                       | SoB experienced when washing and dressing                                                    | SoB experienced when at rest |
|                           | Cough                               | Cough does not interrupt any usual activities        | Cough interrupts or disturbs some usual activities                                          | Cough interrupts or disturbs most usual activities                                           |                              |
| Incontinence              | During a typical day                | COPD symptoms do not cause any urine leakage         | COPD symptoms are causing a few drops of urine leakage                                      | COPD symptoms are causing urine leakage which makes underwear wet                            |                              |
| Mucus                     | When waking up on a typical morning | It is not at all difficult to bring up mucus         | It is a little difficult to bring up mucus                                                  | It is very difficult to bring up mucus                                                       |                              |
| Sleep                     | When waking up on a typical morning | Feel rested                                          | Feel somewhat rested                                                                        | Do not feel rested at all                                                                    |                              |
| Exacerbation              | During a typical year               | Never experience any COPD exacerbations or flare-ups | Experience one or more COPD exacerbations or flare-ups that require antibiotics or steroids | Experience one or more COPD exacerbations or flare-ups that require a hospital stay or visit |                              |

## ② Step 2: 選択課題 (Choice Tasks) の設計

設定した特性と水準の組み合わせから複数の治療プロファイルを作成し、各課題において2つ以上の選択肢を提示する。回答者には、「どの治療を選ぶか」(もしくは「どれも選ばないか」という形式で回答してもらう。実務では、D-最適化デザインなどの実験計画法を用いて、限られた課題数でも効率的にパラメータ推定が行えるよう特性と水準の組み合わせを最適化する。流れとしては、治療プロファイルをD-最適化で作成し、それらのランダム組み合わせを選択課題とする方法と、最初から選択課題(選択肢間の水準の差)をD-最適化で作成する方法が存在する。加えて、内的妥当性確認のための選択課題を追加する(5.2.3.2.6 参照)。

### ■コラム：実験計画法：D-最適化 (D-optimal design)

DCE においては、特性数や水準数の組み合わせが多くなるにつれて、理論上生成可能な治療プロファイルや選択課題数が急激に増加する。そのため、全組み合わせ(総当たり)を提示することは現実的ではない。直交計画法などで質問数を減らす方法もあるが、それでも設計が大きい場合には十分に削減できないことがある。そこで限られた質問数で推定精度(情報量)を確保し、効率的に選好パラメータを推定するための実験計画法が必要となる。

D-最適化は、このような目的のために用いられる代表的な実験計画法であり、限られた選択課題数で最大限の情報を得ることを目的としている。具体的には、推定対象となるパラメータの分散を最小化(同値的に情報行列の行列式を最大化)するように、特性と水準の組み合わせを設計する。

DCE では、複数の特性水準に対応する選好パラメータを同時に推定することが求められることが多いため、IMI PREFER においても、各パラメータを同程度に重要と考え、全体として推定精度を高める観点から、D-最適基準が一般的に適しているとされている。なお、実験計画法にはD-最適基準のほかにA-最適基準やI-最適基準なども存在するが、DCE の設計ではD-最適基準が用いられることが多い。

実務においては、D-最適化デザインは専用の実験計画ソフトウェアを用いて作成されることが多く、Sawtooth Software などの商用ソフトウェアのほか、近年は Ngene、R パッケージ `idefix`、SAS プロシジャ `%ChoiceEff` などを用いて設計される場合もある。これらのツールを用いることで、事前に仮定した選好パラメータ（事前分布）を反映したデザインの作成も可能となる。

③ Step 3: 症例数設定

定量的な患者選好研究における症例数設定は、推定されるパラメータを用いた検定の観点、あるいは実務的な経験則に基づいて行われることが多い。とりわけ DCE では、特性数や水準数、選択課題数などの設計要素が推定精度に大きく影響するため、研究目的および解析計画を踏まえた症例数設計が重要となる（詳細は 5.3.1.2 参照）。同時に、患者選好研究においては実現可能性も大きな制約になるため、患者パネルや患者団体などへアクセスし、可能な数を把握することも重要である。

④ Step 4: 調査の実施

対象患者に選択課題を提示し、回答を収集する際には、回答者の認知負荷に配慮し、課題数や特性数を適切に調整する必要がある。特性や水準の数が多い場合には、一人当たりの選択課題数を抑えるために複数の質問票セットを作成し、質問票セットへの割り当てをランダム化する方法が用いられることもある。

⑤ Step 5: 統計解析

収集された回答データを用いて、各特性水準に対応する「選好の重み（preference weight）」を推定する。最も基本的な解析モデルとしては条件付きロジットモデルが用いられるが、IMI PREFER<sup>16)</sup>では、選好の不均一性を考慮することの重要性が強調されており、ランダムパラメータロジットモデルや階層ベイズモデルなどを用いた解析が推奨されている（詳細は 5.3.1.3 参照）。推定された重みを用いることで、各特性の相対的重要度、特性間のトレードオフ（どの程度の副作用を許容するか等）、特定の治療プロファイルが選択される確率（選択確率）など、意思決定に直接関連する指標を算出することが可能である。また、背景因子による部分集団解析や、モデルに基づく不均一性評価を併せて実施することで、患者集団内の選好の分布を把握することができる。

⑥ Step 6: 結果の解釈と応用

DCE の結果は、棒グラフなどを用いて特性ごとの相対的重要度やトレードオフを可視化する形で提示されることが多い。一般に、水準による選好の重みの変

動幅が大きい特性ほど、患者にとって重要であると解釈される（図 23 参照）。各特性ごとに縦方向のパネルが設けられ、各パネル内には推定された選好の重みと対応する水準がプロットされる。パネルは、各パネル内で推定された最大の選好の重みの大きさに基づいて並んでおり、すべての特性において、特性が悪化するにつれて選好が低下するという共通の傾向が認められている。各特性の相対的重要度は、各特性について、その水準に対応する推定された選好の重みの最大変化量を算出することで求められる。すなわち、図 23 に示されている順に、0.561、0.889、1.188、1.337、1.379、および 1.639 である。相対的重要度を求めるには、それぞれの値を合計値に対する百分率で表す。例えば、 $(0.561, 0.889, \dots, 1.379, 1.639)$  の合計 6.993 を用いると、 $100 \times (0.561, 0.889, \dots, 1.379, 1.639) / 6.993 = (8.02, 12.71, 16.99, 19.12, 19.72, 23.44)$  となり、「増悪」は COPD 患者にとって最も重要な特性であることが示唆された（図 24 参照）。

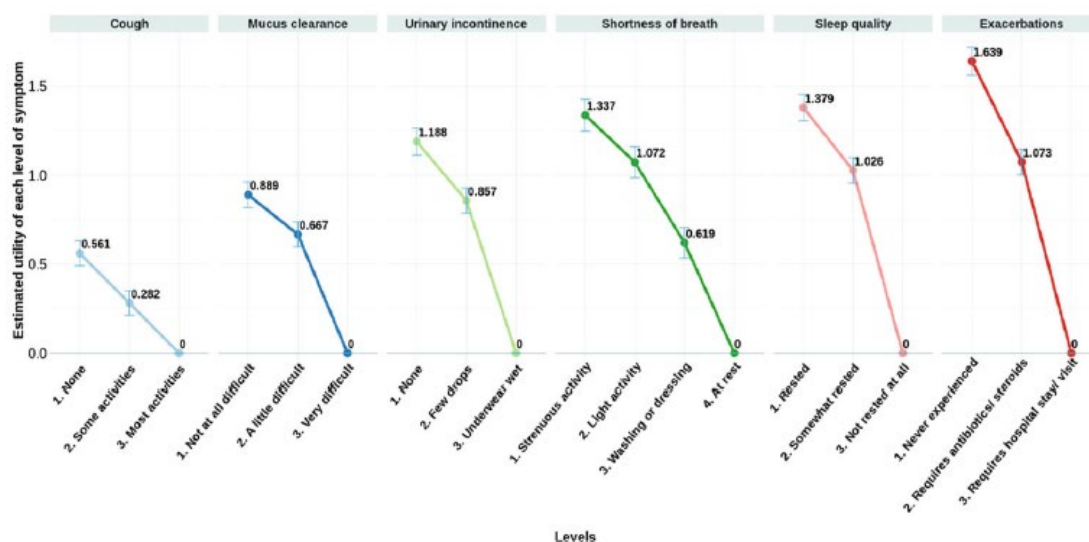


図 22 推定された選好の重みのプロットの例示

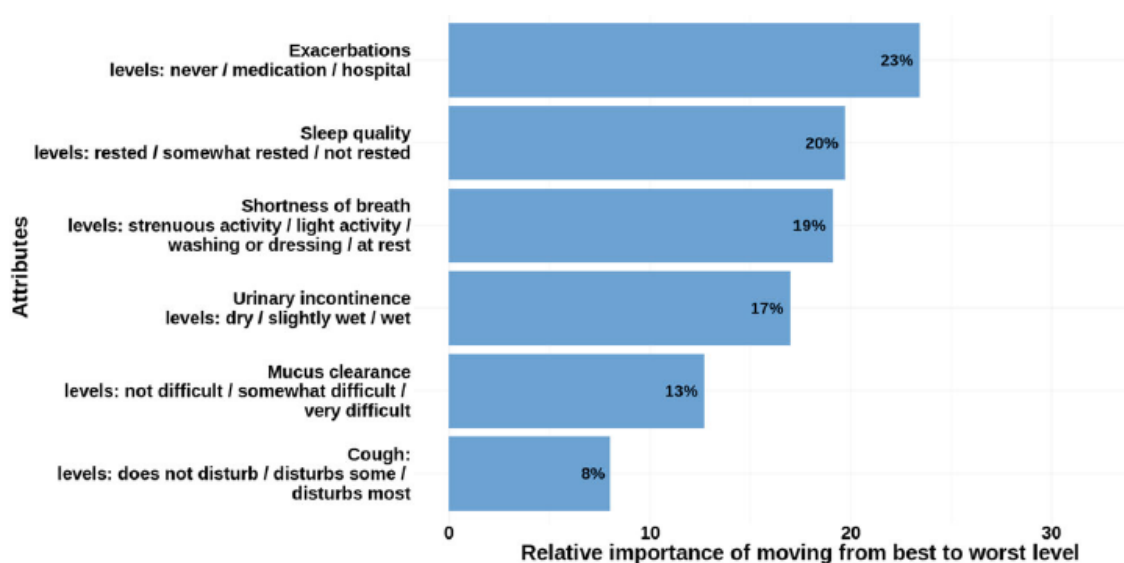


図 23 6つの特性における相対的重要度の例示

IMI PREFER では、集団平均の結果のみならず、選好の分布や部分集団間の違いについても透明性をもって報告することが重要であるとされている。これらの結果は、規制当局や HTA への提出資料として活用されるほか、医薬品のベネフィット・リスク評価、臨床試験のエンドポイント選定、製品設計、さらには意思決定支援ツールの開発などに反映され得る。

#### ■コラム：PPI を用いた MCDA による定量的 BR 評価

BR 評価では、複数のベネフィット（有効性、QOL 等）とリスク（有害事象、治療負担等）を同時に扱うため、「何をどれだけ重視するか」という価値判断が不可避である。PPI（Patient Preference Information：患者選好情報）を用いることで、この価値判断を患者視点で定量化し、意思決定の透明性を高めることができる。

MCDA（Multi Criteria Decision Analysis：多基準意思決定分析）は、①評価軸（criteria）の定義、②各評価軸に対する治療の成績を共通スケールへ変換（scoring/value function）、③DCE などを用いて PPI から各評価軸の重み（weights）を推定、④重み付きで統合（aggregation）という手順で、治療選択肢の総合的な BR バランスを構造化して示す枠組みである。これにより「どのベネフィットを得るために、どのリスクをどの程度許容するか（リスク許容度）」を明示でき、規制当局や HTA、開発戦略、Shared Decision Making 等の文脈で活用され得る。

一方で、統合スコアや順位は、選定した評価軸、重み、成績データの前提に依存するため、「単一スコアで結論を固定する」用途には適さない。二重計上（相関の高い評価軸の重複）、患者集団内の選好の不均一性（サブグループ差）、データの不確実性を意識し、感度分析や不確実性評価とあわせて、注意して解釈する必要がある。

### ⑦ Step 7: 妥当性の検証

DCE を用いた患者選好研究では、結果の信頼性を担保するために妥当性の検証が重要となる。内的妥当性としては、選択の一貫性、支配選択の確認、トランジティビティの検証などが行われる（5.2.3.2.6 参照）。外的妥当性としては、既存研究との比較や理論的整合性の検討が含まれる。一般的な分析アプローチの 1 つは、全サンプルとすべての妥当性テストに合格した回答者のサブグループの両方で分析を実施し、結果を比較することである。

DCE は、複数の特性を同時に扱い、患者がどのようなトレードオフを許容するかを定量的に評価できる点で強みを有する。一方で、設計および解析が比較的複雑であり、回答者に一定の認知的負担がかかるという限界もある。そのため、IMI PREFER では、研究目的、意思決定文脈、対象集団の特性を踏まえた上で、DCE が最適な手法であるかを検討することが求められている。

### 5.3.1.2. DCE の症例数設定

症例数に影響する主な要因としては、特性数、各特性の水準数、選択課題の数、各課題に提示される選択肢 (alternatives) の数が挙げられる。また、解析モデルとしては、条件付きロジットモデルを前提として症例数を検討することが多く、パラメータ推定の標準誤差、検出力 (power)、有意水準 ( $\alpha$ ) などが設計上の要件となる。

症例数設定の方法としては、主に以下のアプローチが用いられており、これらのアプローチに関する最近の概説は、Assele (2023)<sup>47)</sup>に記載されている。一方、DCE 以外の手法では DCE ほど確立した症例数設計法は存在せず、研究目的や実施可能性を考慮した実務的な判断が行われることが多い。

#### (1) 経験則 (Rule of Thumb) に基づく方法

実務上は、300 例程度のサンプルサイズがあれば、多くの DCE で安定した推定が可能とされることが多い。また、Orme (2010)<sup>48)</sup>で提案された経験則として、 $N=(500 \times c)/(t \times a)$  が広く用いられている。ここで、 $c$  は特性の最大水準数、 $t$  は回答者 1 人当たりの選択課題数、 $a$  は各課題の選択肢数である。例えば、最大水準数 3、課題数 9、選択肢数 2 の場合、約 84 例となる。

#### (2) 精度基準に基づく方法

このアプローチでは、各パラメータの標準誤差 (SE) を一定水準以下（例えば  $SE < 0.05$ ）に抑えることを目標とし、シミュレーションを用いて必要な症例数を決定する。パラメータの推定精度を重視する場合に有用な方法である。

### (3) 検定ベースの方法

de Bekker-Grob (2015)<sup>49)</sup>で提案されたアプローチで、事前に想定した選好パラメータ ( $\beta$ ) の値を用い、片側有意水準  $\alpha=0.05$ 、検出力  $\text{power}=0.9$  などの条件下で症例数を計算する方法である。この場合、効果サイズが小さいパラメータほど、より多くの症例数が必要となる。例えば、 $\beta=0.1$  程度の小さな効果を検出するには、数百例規模のサンプルが必要となることもある。

これらを踏まえた実務的な症例数の目安としては、単純なデザイン（特性数 3~4、課題数 8~12）では 100~150 例程度、複雑なデザイン（特性数 6 以上、課題数 15 以上）では 200~300 例以上が想定されることが多い。一方、希少疾患など十分な症例数を確保することが難しい場合には、DCE に代えて、Swing Weighting や Threshold Technique など、少数サンプルに適した代替手法の検討も選択肢となる。

症例数設計のプロセスとしては、まずパイロット研究を実施し、選好パラメータの事前推定値を取得する。その上で、情報行列（Fisher Information）を用いた精度評価を行い、必要に応じてシミュレーションを通じて症例数を最適化することが望ましい。

このように、患者選好研究の症例数設定においては、理想的な統計学的要件と研究の実現可能性（feasibility）とのバランスを考慮し、研究目的に照らして適切な症例数を設定することが重要である。

#### 5.3.1.3. DCE の解析モデル

DCE における代表的な解析モデルを紹介する。解析手法について更に理解を深めたい場合は Johnson (2025)<sup>50)</sup>が参考になる。以下に解析する際のデータ構造の例を示す。各回答者（ID）が課題（Question）ごとにどの選択肢（Alternative）を選んだかを 0 または 1 で保持する（Choice）。また各レコードには選択肢に紐づく特性ごとの水準が保持される。

| ID  | Question | Alternative | Choice | Cough           | Mucus clearance      | Urinary incontinence | ... |
|-----|----------|-------------|--------|-----------------|----------------------|----------------------|-----|
| 1   | 1        | 1           | 0      | Some activities | Very difficult       | None                 | ... |
| 1   | 1        | 2           | 1      | None            | A little difficult   | Few drops            | ... |
| 1   | 2        | 1           | 0      | Most activities | A little difficult   | Underwear wet        | ... |
| 1   | 2        | 2           | 1      | Some activities | Not at all difficult | Few drops            | ... |
| 1   | 3        | 1           | 1      | Some activities | Not at all difficult | None                 | ... |
| 1   | 3        | 2           | 0      | Most activities | A little difficult   | Underwear wet        | ... |
| 1   | 4        | 1           | 1      | None            | Very difficult       | Few drops            | ... |
| 1   | 4        | 2           | 0      | Some activities | A little difficult   | None                 | ... |
| ... | ...      | ...         | ...    | ...             | ...                  | ...                  | ... |

#### (1) 条件付きロジットモデル

DCE の最も基本的な解析モデルとして条件付きロジットモデル (Conditional Logit Model) が挙げられる。これは McFadden によって定式化され、回答者がどの選択肢を選ぶかという選択行動を、選択肢の特性 (水準) に基づいてモデル化し、特性  $k$  の重要度を表す選好パラメータ  $\beta_k$  を固定効果として推定する。つまり、各特性の重要度は個人によらず共通であるという仮定を置いており、この仮定により選好の不均一性を考慮できない点に限界がある。

各選択肢は複数の「特性」とその「水準」によって記述され、回答者は各選択肢から得られる効用 (utility) が最も高い選択肢を選ぶと仮定する。効用は、観測可能な部分 (特性水準とその重み) 及び観測不可能な誤差項から構成される。選択肢  $j$  を回答者が選択する際の効用は次式で表される。

$$U_j = \sum_{k=1}^K \beta_k x_{jk} + \varepsilon_j$$

- $x_{jk}$  : 選択肢  $j$  の特性  $k$  の値
- $\beta_k$  : 特性  $k$  に対応する共通の選好パラメータ
- $\varepsilon_j$  : 誤差項

すべての回答者で選好パラメータ ( $\beta_k$ ) は共通であると仮定しているため、個人の選好の違いは考慮されない。各選択肢の誤差項が独立同一の標準ガンベ

ル分布に従うと仮定する条件付きロジットモデルでは、推定された選好パラメータ  $\beta_k$  と各選択肢の特性（水準）を用いて、選択確率を算出できる。推定された  $\beta_k$  は特性（水準）に対する選好の重みを表し、関連する水準間の重みの差などを用いて、特性の相対的重要度やトレードオフ指標（例：最大許容リスク）を算出できる。

推定が比較的容易で、解釈も直感的なため、DCE 解析の入門的・基礎的モデルとして広く用いられている。ただし、個人間の選好の違い（選好の不均一性）をモデル化できないため、限界がある。また選好のばらつきが大きい場合、集団平均のみの推定では重要な情報を見落とす可能性がある。よって、部分集団解析などで不均一性を確認することが重要である。

## (2) ランダムパラメータロジットモデル

ランダムパラメータロジットモデル（Random Parameters Logit Model）は、条件付きロジットモデルを拡張し、選好パラメータを個人ごとの確率変数として扱うことで、回答者間の選好の不均一性を考慮可能な解析モデルである。各特性について平均的な選好の重みと個人差の大きさを同時に推定できる点が特徴であり、実際の患者選好をより現実的に反映できる。

条件付きロジットモデルでは、すべての回答者で選好パラメータ（ $\beta_k$ ）は共通と仮定していたが、ランダムパラメータロジットモデルではこの仮定を緩和し、回答者  $i$  ごとに異なる確率変数として選好パラメータ（ $\beta_{ik}$ ）を扱う。これにより、同じ特性（水準）であっても、患者ごとに異なる重みづけがされていることをモデル化できる。ランダムパラメータロジットモデルにおける効用関数は、以下のように表される。

$$U_{ij} = \sum_{k=1}^K \beta_{ik} x_{ijk} + \varepsilon_{ij}$$

- ・  $x_{ijk}$ ：回答者  $i$  に提示された選択肢  $j$  の特性  $k$  の値
- ・  $\beta_{ik}$ ：回答者  $i$  に固有の特性  $k$  に対応する選好パラメータ（例：正規分布に従うと仮定）
- ・  $\varepsilon_{ij}$ ：誤差項

ランダムパラメータロジットモデルの特徴として、個人間の選好の不均一性を明示的にモデル化することが可能である。各特性について、患者集団の平均的な選好の重みだけでなく、選好のばらつきの程度（標準偏差）を同時に推定できる。IMI PREFER でも、実際の患者選好により近いモデルとして位置づけら

れている一方、モデル構造が複雑で、推定に計算負荷がかかる、仮定する分布に結果が依存する可能性がある、条件付きロジットモデルよりもサンプルサイズや解析専門性がより求められるなどの特徴がある。

### (3) 階層ベイズモデル

階層ベイズモデル (Hierarchical Bayes Model) は、ベイズ推定の枠組みを用いて、個人レベルと集団レベルの 2 階層構造を同時に推定することで、DCE における選好の不均一性を考慮する解析手法である。個人レベルでの選好パラメータを推定可能であり、ランダムパラメータロジットモデルと並んで選好の不均一性を扱う代表的な方法である。

各回答者は、それぞれ固有の選好パラメータ ( $\beta_{ik}$ ) を持つと仮定する。ただし、個人レベルの選好パラメータは完全に独立ではなく、集団レベルの分布 (母集団分布) から生成され则认为。この個人レベル及び集団レベルの 2 つの階層構造を同時に推定するのが階層ベイズモデルである。階層ベイズモデルは、主に以下の 2 階層で構成される。

#### 1) 個人レベル (第 1 階層)

- 回答者  $i$  の特性  $k$  に対応する選好パラメータ  $\beta_{ik}$
- 個人の選択行動を説明

#### 2) 集団レベル (第 2 階層)

- 個人パラメータ  $\beta_{ik}$  が従う分布 (平均・分散)
- 集団全体の傾向とばらつきを表現

階層ベイズモデルにおける回答者  $i$  の効用関数は、次式で表される。

$$U_{ij} = \sum_{k=1}^K \beta_{ik} x_{ijk} + \varepsilon_{ij}$$

- ・  $x_{ijk}$  : 回答者  $i$  に提示された選択肢  $j$  の特性  $k$  の値
- ・  $\beta_{ik}$  : 回答者  $i$  に固有の特性  $k$  に対応する選好パラメータ (集団レベルの事後分布から推定)
- ・  $\varepsilon_{ij}$  : 誤差項

マルコフ連鎖モンテカルロ (MCMC) 法などを用いて推定する。個人レベルの事後分布を得ることができる。また個人別・集団別の相対重要度を算出することが可能である。サンプルサイズが限られている場合でも、集団情報を共有することで安定した推定が可能である。本モデルを利用する際の利点として、個人レベルの選好推定が可能であること、選好の不均一性を柔軟に表現できること、背景項目別の選好の推定も容易であることが挙げられる。またサンプルサ

イズが比較的小さい場合でも推定が安定しやすく、個別患者の選好を活用する分析（例：意思決定支援、シミュレーション）に適していると言える。

このように DCE の解析には、条件付きロジットモデル、ランダムパラメータロジットモデル、階層ベイズモデルの他に、潜在クラス分析などが用いられる。実務上は、R パッケージ（Apollo、mlogit、gmnl 等）、SAS プロシジャ（PROC MDC 等）、Python パッケージ（biogeme、pylogit 等）が利用されることが多い。

### 5.3.2. その他の手法の紹介と比較

IMI PREFER<sup>16)</sup>では、DCE を含む定量的患者選好研究において、単一の手法が常に最適であるとは限らず、研究目的、意思決定文脈、対象集団の特性に応じて、適切な手法を選択することが重要であるとされている。本節では、DCE 以外の代表的な定量的選好評価手法について概説し、それぞれの特徴と適用場面を整理する。

#### ■コラム：コンジョイント分析

コンジョイント分析は、医薬品など対象とするものの好み、そのものが持つ複数の特性によって構成されたと考え、特性ごとの好ましさを推定することで選好構造を理解しようとする方法である。医療分野では、患者が治療の有効性、安全性、投与方法など複数の特性をどのように評価しているかを定量的に把握するために用いられてきた。

IMI PREFER<sup>16)</sup>では、DCE や Best-Worst Scaling (BWS) などは、広義にはコンジョイント分析に位置づけられる代表的な表明選好手法として整理されている。これらの手法はいずれも、複数特性間のトレードオフを評価できる点に共通した特徴を有しており、医薬品のベネフィット・リスク評価や HTA において、患者視点を形式的に組み込むための枠組みとして活用されている。ただし、本書ではこの用語を用いて説明しない。

#### 5.3.2.1. Best-Worst Scaling

Best-Worst Scaling (BWS) は、提示された特性や特性水準の中から「最も重要 (Best)」および「最も重要でない (Worst)」項目を回答者に選択させることで、選好を測定する手法である。「最大差スケール」または「max-diff」とも呼ばれ、一連の選択肢が提示された場合、人は選択肢を評価したり順位付けしたりするよりも、最良と最悪という極端なものを識別する方が得意であるという心理学的前提に基づいている。DCE と比較して認知負荷が低く、回答しやすい点が特徴とされている。

BWS には、Case 1～Case 3 までであるが、ここではよく取り上げられる Case 1 と Case

2 を紹介する。詳細は Mühlbacher (2016)<sup>51)</sup>が参考となる。

(1) BWS Case 1 (特性の重要度評価)

BWS Case 1 では、特性そのものを提示し、その中から最も重要な特性と最も重要でない特性を選ばせる。選択対象となる特性（治療特性、試験エンドポイント、ベネフィット、リスクなど）のリストが作成され、各回答者にはいくつかの特性をサブセットとした選択肢セットが提示され、各選択肢セットで最も好ましい（「最良」または「最も重要」）特性と最も好ましくない（「最悪」または「最も重要でない」）特性を選択するように求められる（図 25 参照）。水準がないため、BWS Case 1 は通常、従来の DCE よりも多くの特性を含めることができ、選択肢セットで提示される特性のサブセットは、実験計画法を使用して作成される。BWS Case 1 のデータを解析するのに最も簡単な方法は、ある特性が最良と最悪として選択された回数を直接数えることである。これにより、各特性の相対的重要度や順位を把握することができる。あるいは、特性固有のダミー変数を独立変数とし、その特性が選択されたかどうかを示す指標を従属変数とする離散選択モデル（条件付きロジット、ランダムパラメータロジットなど）を用いて分析することもできる。これにより、特性の順位づけや相対的重要度を把握することが可能である。一方で、水準情報を含めていないことにより、水準を含めたトレードオフを考慮できないため、特性間の明示的なトレードオフ分析には不向きである。

| I would be bothered by this the MOST (Check only one) | Symptom                                   | I would be bothered by this the LEAST (Check only one) |
|-------------------------------------------------------|-------------------------------------------|--------------------------------------------------------|
| <input type="radio"/>                                 | Difficulty walking                        | <input type="radio"/>                                  |
| <input type="radio"/>                                 | Embarrassment                             | <input type="radio"/>                                  |
| <input type="radio"/>                                 | Difficulty choosing clothing              | <input type="radio"/>                                  |
| <input type="radio"/>                                 | Redness of skin                           | <input checked="" type="radio"/>                       |
| <input checked="" type="radio"/>                      | Difficulty with work or school activities | <input type="radio"/>                                  |

図 24 BWS Case 1 の選択肢セットの例示

(2) BWS Case 2 (特性水準の評価)

BWS Case 2 は、医療分野で広く用いられている BWS 手法の一つであり、回答者は様々な水準を持つ特性（DCE と同様）によって定義されたいくつかの特性水準のサブセットから成る選択課題を提示され、各選択課題において最良および最悪（あるいは最も重要および最も重要でないなど）の特性水準を選択する

よう求められる (Louviere 2015<sup>52</sup>)、図 26 参照)。選択肢として示す特性水準は実験計画によって決定され、DCE に類似した設計 (例えば、一部実施要因計画) が用いられる。BWS Case 1 と同様に、研究者は特性水準が最良と最悪として選択された回数を単純に数え、その特性水準が出現した回数を調整すればよい。この単純なカウントにより個人レベルの選好推定値が得られる。BWS Case 2 のデータは、離散選択モデルを用いて解析することも可能であり、研究目的に応じて異質性の評価や個人レベルの推定を行うこともできる各選択課題では複数の特性水準の中から最良および最悪の特性水準を選択するため、特に関心のある特性が多い場合、DCE よりも認知的負担が少ない可能性がある。この認知的負担の軽減は、高齢者や子供を対象とした健康状態評価研究でこの方法が広く用いられている理由の一つとして挙げられている。

Choose the best thing in this treatment by clicking the circle under “best” and choose the worst thing by clicking the circle under “worst.” You have to choose a best thing and worst thing to move on. Remember that a computer chose the combinations to make the experiment work, and some of them seem bad. Even so, please pick the best and worst thing.

| Best                  | Treatment                                              | Worst                 |
|-----------------------|--------------------------------------------------------|-----------------------|
| <input type="radio"/> | Slows the progression of weakness                      | <input type="radio"/> |
| <input type="radio"/> | 2-year gain in expected life span                      | <input type="radio"/> |
| <input type="radio"/> | 1-year of postapproval drug information available      | <input type="radio"/> |
| <input type="radio"/> | Causes loss of appetite                                | <input type="radio"/> |
| <input type="radio"/> | Increased risk of bleeding gums and increased bruising | <input type="radio"/> |
| <input type="radio"/> | Increased risk of harmless heart arrhythmia            | <input type="radio"/> |

図 25 BWS Case 2 の質問票の例示

## 5.3.2.2. Swing Weighting

Swing Weighting (SW) は、複数の特性を「最悪の状態から最良の状態へ変化させたときの重要度」を比較することで、特性の重み付けを行う手法である。SW は Multi-Criteria Decision Analysis (MCDA) における重み付け手法として用いられることが多い。なお、SW は選好研究の手法である一方、MCDA は複数のベネフィット・リスクを統合して意思決定を支援するフレームワークであり、両者は区別される (IMI PREFER Section 4.3 参照<sup>16)</sup>)。

SW は通常、2 つの段階から構成され、第 1 段階では回答者は特定の特性における水準の変化 (特性スイング) の重要度を高い順から低い順にランク付けするよう求められる。第 2 段階では、回答者は特性スイングの相対的な価値を判断するよう求められる。最も一般的な方法は、最も高いランクの特性に 100 の値を割り当て、次に最も高いランクの特性スイングと比較して、2 番目に高いランクの特性スイングの価値 (0~100) を回答者に表現してもらう。このプロセスはすべての特性に対して繰り返され、得られた重みは合計が定数 (通常は 1 または 100) になるように正規化され、特性スケールの変動の相対的な重要度が算出される (Angelis & Kanavos 2017<sup>53)</sup>)。

| Attribute                   | Worst | Best | Weight |
|-----------------------------|-------|------|--------|
| Likelihood of ESRD/dialysis | 30%   | 10%  | 100    |
| Risk of infections          | 20%   | 0%   | 37.5   |

SW では特性間の重みやトレードオフ情報が回答者から直接取得されるため、DCE のような複雑な離散選択モデルを必ずしも必要としない。そのため SW のデータ解析は比較的容易で、基本的な記述統計を用いて行うことができる。集団の選好はディリクレ回帰 (Tervonen 2019<sup>54)</sup>)、回答者の特性が選好に与える影響を評価することも可能になる。

## 5.3.2.3. Threshold Technique

Threshold Technique (TT) は、2 つの選択肢を提示し、ある特性の水準を段階的に変化させることで、回答者が 2 つの選択肢の価値がちょうど同じ程度と感じる点 (閾値) を特定する手法である。この方法は、最大許容リスク (Maximum Acceptable Risk) や最小必要ベネフィット (Minimum Required Benefit) の推定に適している。詳細は Hauber and Coulter (2019)<sup>55)</sup> が参考となる。

TT を用いた研究では、回答者は参照オプション (通常は標準治療、例えば現在の治療または無治療) と目標オプションのどちらか好ましい方を選択するよう求められる。

どちらのオプションも、特性とその水準によって定義される。目標オプションは通常、参照オプションと比較してベネフィットとリスクの両方が増加しているため、回答者はどちらかを選択する際にトレードオフを行う必要がある（図 27 参照）。

最初の質問では、通常、参照オプションと目標オプションの両方で主要特性の水準が同じになるように設定されている。最初に参照オプションが選択された場合、目標オプションの主要な特性がより良い値に設定され、質問が繰り返される。最初に目標オプションが選択された場合、目標オプションの主要な特性がより悪い値に設定され、質問が繰り返される。この手順は、回答者が参照オプションと目標オプションに対して価値がちょうど同じ程度になる閾値レベルを研究者が特定できるまで繰り返される。関心のある主要な特性が負担の尺度（例：副作用、時間、費用）である場合、推定閾値は、対象となる選択肢によって得られる増分的なベネフィットを正確に相殺する追加リスクの水準となる。主要な特性がベネフィット（例：生活の質の向上、生存期間）である場合、推定閾値は対象となる選択肢の増分的なリスクを相殺するために、対象が提供しなければならない最小限の追加ベネフィットを示す。

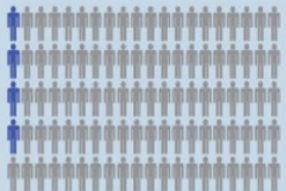
|                                                                 | Your current treatment                                                              | A new device                                                                                              |
|-----------------------------------------------------------------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| Hours of "on time" each day                                     | 10 hours of "on time"<br>6 hours of "off time"                                      | 10 hours of "on time"<br>6 hours of "off time"                                                            |
| Severity of movement symptoms                                   | 8<br>(on a scale from 0 to 10)                                                      | 4<br>(on a scale from 0 to 10)                                                                            |
| Severity of pain                                                | 7<br>(on a scale from 0 to 10)                                                      | 7<br>(on a scale from 0 to 10)                                                                            |
| Difficulty thinking clearly, getting organized, or making plans | 2<br>(on a scale from 0 to 10)                                                      | 2<br>(on a scale from 0 to 10)                                                                            |
| Number of pills you need to take                                | 6 pills each day                                                                    | 6 pills each day                                                                                          |
| Risk of getting depression or anxiety after getting the device  | None                                                                                | None                                                                                                      |
| Risk of having bleeding in the brain after getting the device   | None                                                                                | <br>4 out of 100 (4%) |
| Risk of dying within 1 year after getting the device            | None                                                                                | None                                                                                                      |
| Which option would you choose?                                  |  |                      |

図 26 Threshold Technique の質問票の例示

## 5.3.2.4. 各定量的手法の特徴の比較

これらの手法はいずれも、患者選好を定量的に把握するための有用なアプローチであるが、研究目的、求めるアウトプット、対象集団の特性、サンプルサイズ、認知負荷、解析の複雑さなどを考慮して選択されるべきである。

IMI PREFER では、複数特性を同時に扱い、トレードオフを明示的に評価し、規制当局や HTA における意思決定に活用する場合には、DCE が特に有用であるとされている。一方、サンプルサイズが限られている場合や、特性の順位付けや重要度評価を簡便に把握したい場合には、BWS や SW などの代替手法が適切となる場合もある。

このように、定量的患者選好研究においては、単一の手法に固執するのではなく、意思決定文脈に適した手法を選択・組み合わせる柔軟な姿勢が求められる。以下の表に各手法の長所及び短所をまとめる。これらの手法を比較した Veldwijk (2023)<sup>56)</sup>も参照されたい。

| 手法                                                        | 長所                                                                                                                                                                | 短所                                                                                                                                                |
|-----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>DCE (Discrete Choice Experiment)</b><br><b>離散選択実験法</b> | <ul style="list-style-type: none"> <li>- 複数の特性・水準を同時に扱える</li> <li>- トレードオフの定量化が可能</li> <li>- 選好の相対的な重要度や MAR など多様な指標が得られる</li> <li>- 実績が多く、規制当局でも使用例あり</li> </ul> | <ul style="list-style-type: none"> <li>- 研究設計が複雑(実験計画が必要)</li> <li>- 認知負荷が高く、理解困難な場合あり</li> <li>- 特性数が多いと回答困難になる</li> </ul>                       |
| <b>Swing Weighting (SW)</b>                               | <ul style="list-style-type: none"> <li>- 個人レベルの選好データを直接取得可能</li> <li>- 設計・解析が比較的簡便で、小規模サンプルでも実施しやすい</li> <li>- 面接により理解支援や定性的情報の収集が可能</li> </ul>                   | <ul style="list-style-type: none"> <li>- 面接者主導のため人的・時間的負担が大きい</li> <li>- サンプルサイズが限られ、代表性に課題が生じやすい</li> <li>- 特性間の相互作用や複雑なトレードオフを評価しにくい</li> </ul> |
| <b>Best-Worst Scaling (BWS) Case 1</b>                    | <ul style="list-style-type: none"> <li>- 設計・実施が比較的簡便で、迅速に実施しやすい</li> <li>- 順位付けや相対的重要度を把握しやすい</li> <li>- 結果が透明で、さまざまな</li> </ul>                                  | <ul style="list-style-type: none"> <li>- 水準を含めた選好や特性間のトレードオフを評価できない</li> <li>- MAR などの閾値推定には限界がある</li> </ul>                                      |

| 手法                                         | 長所                                                                                                                                                                                | 短所                                                                                                                                                                     |
|--------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                            | 関係者に理解されやすい                                                                                                                                                                       | - 複雑な研究課題には対応しにくく、補完的な手法としての利用にとどまりやすい                                                                                                                                 |
| <b>Best-Worst Scaling (BWS)<br/>Case 2</b> | <ul style="list-style-type: none"> <li>- 特性・水準の相対的重要性を把握でき、Case 1 より多くの情報を得られる</li> <li>- 単一プロファイルの評価で済むため、DCE より認知的負担が低い場合がある</li> <li>- 単純な集計から離散選択モデルまで幅広い解析が可能</li> </ul>     | <ul style="list-style-type: none"> <li>- 特性間の明示的なトレードオフを要しないため、MAR などの推定は困難</li> <li>- 「最良／最悪」の判断が一部の回答者には概念的に難しい場合がある</li> <li>- DCE ほど直接的に選択行動を評価する手法ではない</li> </ul> |
| <b>Threshold Technique (TT)</b>            | <ul style="list-style-type: none"> <li>- MAR や MRB の推定に適する</li> <li>- 設計・解析が比較的簡便で、解釈しやすい</li> <li>- 小規模サンプルでも実施しやすく、希少疾患研究に適する</li> <li>- 各回答者の閾値（MAR、MRB など）を直接推定できる</li> </ul> | <ul style="list-style-type: none"> <li>- 評価できるトレードオフが比較的単純で、適用範囲が限定される</li> <li>- 特性数が多い場合は設計が難しくなる</li> <li>- 反復的な質問が必要で、回答者にとって理解が難しい場合がある</li> </ul>                |

#### 5.4. 選好が均一でない集団・部分集団の考え方

定量的な患者選好研究においては、患者集団における選好の不均一性を前提として研究を設計・解析することが重要である。同一疾患であっても、患者ごとに治療に対して重視する要素は異なり、年齢、疾患重症度、治療経験、価値観などにより選好は大きく変化し得る。そのため、集団平均のみで患者選好を評価することは、必ずしも十分とは言えない。IMI PREFER<sup>16)</sup>においても、患者選好の不均一性を考慮せずに平均的な結果のみを用いた意思決定は、一部の患者群にとって不利益となる可能性がある点が指摘されている。規制当局や HTA 評価においては、全体集団としての平均的な選好に加え、特定の部分集団における治療価値やリスク許容度を理解することが重視されつつ

ある。

選好の不均一性は、大きく以下の 2 つに整理される。一つは、年齢、性別、疾患重症度、治療経験など、収集した患者背景情報によって説明可能な不均一性であり、これは「観測された不均一性」と捉えられる。もう一つは、価値観や心理的特性など、調査で直接測定することが困難な要因に起因する説明困難な不均一性である。後者は、「観測されない不均一性」として統計モデルの中で扱われる。このような不均一性を考慮するため、定量的選好研究では、研究目的の段階で不均一性評価を明示することが望ましい。例えば、「特定の患者背景による部分集団間で選好に差があるか」を探索的・仮説検証的に評価することが考えられる。ただし、日本においては過去の患者選好研究の蓄積が限られている疾患領域も多く、事前にどの背景因子による不均一性を想定すべきか判断が難しい場合も少なくない。

そのため、サンプルサイズ設計においても、不均一性解析に十分な症例数を理想的に確保することが難しい状況がしばしば生じる。このような場合には、研究の実現可能性 (feasibility) を踏まえ、現実的な症例数の中で実施可能な不均一性評価を行うという考え方が重要となる。

不均一性の評価方法としては、次のような解析アプローチが用いられる。まず、年齢や疾患重症度などの特性ごとに結果を分けて報告するサブグループ解析がある (例: 65 歳以上 vs 65 歳未満、重症 vs 軽症)。また、ロジットモデルに特性×水準の交互作用項を組み込むことで、特性による選好の違いを直接評価する方法もある。ランダムパラメータロジットモデルや階層ベイズモデルを用いれば、個人レベルの選好分布を推定し観測された不均一性を評価することができる。さらに、潜在クラス分析 (Latent Class Analysis) やクラスタリングでは、選好パターンに基づいて患者をクラスに分類し、観測されない不均一性を含めた評価が可能となる。潜在クラス分析では「リスク回避型」「ベネフィット重視型」といった選好タイプの把握にも用いられる。

不均一性が確認された場合には、結果の提示方法にも工夫が求められる。集団平均のみを示すのではなく、選好分布やクラス構造を併せて提示し、部分集団ごとの相対的重要度やトレードオフの違いを明示することが望ましい。規制当局や HTA 向けの資料においては、「全体集団の結果」と「主要な部分集団との比較」を併記することで、意思決定に資する情報提供が可能となる。

患者選好の不均一性は、実際の意思決定にも影響を及ぼす。例えば、特定の部分集団でリスク許容度が低いことが示された場合には、適応症設定や注意喚起などラベリングへの反映が検討され得る。また、治療選択支援ツールにおいて個別患者の選好を反映させることで、Shared Decision Making を支援することも可能である。さらに、開発戦略の観点からは、特定のサブグループに焦点を当てた追加試験やエビデンス収集を検討する契機となり得る。

このように、定量的患者選好研究においては、選好の不均一性を前提として研究を

計画し、結果を解釈・活用することが、患者中心の意思決定を実現する上で不可欠である。以下に、医薬品分野において選好の不均一性を評価した代表的な事例を示す。

例1) アルツハイマー病治療における事例<sup>57)</sup>

アルツハイマー病治療を対象とした研究では、医師と介護者という異なる意思決定者の間で治療選好に差があるかを検討するため、DCE が用いられた。本事例では、治療効果、副作用、投与方法など複数の特性を設定した DCE を実施し、介護者と神経内科医の選好を比較した。解析には階層ベイズモデルが用いられ、個人レベルおよび集団レベルの選好分布が推定された。その結果、治療特性に対する重視度やトレードオフの考え方が、医師と介護者の間で必ずしも一致しないことが示された。特に不一致が大きかったのは「利便性」に関わる治療特性であり、神経内科医は介護者に比べて投与の負担軽減を相対的に強く重視した一方で、介護者ではこれらの特性の重視度は医師ほど大きくなく、効果や安全性といった臨床的アウトカムに比べると選択への寄与が小さい傾向が示された。これは、患者本人に代わって意思決定に関与する立場の違いが、治療ベネフィットやリスクの捉え方に影響を与える可能性を示唆している。本事例は、自己報告が困難な疾患領域において、どの集団の選好を評価対象とするかを明確にした上で研究を設計・解析する重要性を示す例である。

例2) 不妊治療および小児用 EQ-5D-Y の評価における事例<sup>58)</sup>

不妊治療や小児用 EQ-5D-Y 健康状態の評価を対象とした研究では、年齢や評価者グループの違いによる選好の差が検討された。これらの事例では、DCE を用いて選好データを収集し、年齢層や評価者（患者、保護者、医療従事者など）による選好の違いという観測された不均一性を解析した。解析には、ランダムパラメータロジットモデルや潜在クラス分析が用いられ、集団全体の平均的な選好に加え、特性ごとの選好分布の違いや、異なる選好パターンを持つ集団の存在が示された。不妊治療の研究では治療経験や置かれている状況の違いによって、重視する治療特性が異なることが示された。治療成功率をより重視する集団が存在する一方、治療負担や副作用、治療期間をより重視する集団が存在した。小児用 EQ-5D-Y 評価の研究では年齢・評価主体（自分自身か、子どもを想定するか）によって、健康状態のどの側面を重要と捉えるかが異なった。これらの事例は、一つの集団平均のみでは、意思決定に必要な患者視点を十分に捉えられない場合があることを示している。

## 5.5. 国や文化の違いの考え方

患者選好は、疾患特性や個人の価値観だけでなく、国や文化の違いによっても影響を受ける。医療制度、薬剤へのアクセス、医療提供体制、社会的・文化的価値観などの違いにより、同一の特性や水準であっても、その意味づけや受け止め方は国ごとに異なる可能性がある。そのため、定量的患者選好研究を計画・解釈する際には、国や文化の違いが選好に与える影響を考慮することが重要である。

IMI PREFER<sup>16)</sup>においても、国や文化の違いは選好研究における重要な課題として位置づけられており、特に、選好データの転用可能性 (transferability) は「最優先の研究課題のひとつ」とされている。特性や水準の意味が国によって異なる場合、選好結果を単純に比較したり、他国の結果を自国に適用したりすることは慎重であるべきとされている。

### (1) 国内選好研究

国内選好研究は、当該国の医療制度、文化、患者背景に即した選好データを直接的に取得できるという利点を有する。例えば、医療アクセスの状況や治療パスが国ごとに大きく異なる場合、それらを前提とした選好の構造も異なり得る。そのため、国内で実施された選好研究は、規制当局や HTA に対して自国の状況に即したエビデンスを提供できる点で有用である。

一方で、国内選好研究には、コスト増加やタイムラインの延長といった課題も伴う。また、希少疾患などでは十分な症例数を確保することが難しい場合も少なくない。さらに、国際的な比較や整合性の観点では、国内データのみでは限界があることも指摘されている。

### (2) ブリッジング的な海外選好研究の転用

多くの実務場面では、まず海外で実施された選好研究を自国に転用できるか (ブリッジング) を検討することが多い。国ごとに治療パスや文化的背景が異なる場合、選好データの転用可能性は大きな課題となる一方、転用できれば患者コミュニティへの負担を下げるのが可能となる。特に情報に限りがある場合、参考情報として海外の患者選好情報を示すことの意義は大きいと言える。

IMI PREFER では、選好データの転用可能性が低い場合には、追加調査や文化的適応 (Cognitive Debriefing) を検討すべきとされている。特に、特性や水準の解釈が国によって異なると考えられる場合には、翻訳にとどまらない文化的適応を行い、質問票の意味の一貫性を担保することが重要である。転用可能性が十分に担保できない場合には、国内追加調査の実施も選択肢となり得る。

### (3) 国際共同選好研究

国際共同選好研究は、複数の国で共通の設計に基づいて選好データを収集するものであり、規制当局や HTA に対して国際的に一貫したエビデンスを提示できるという利点がある。また、症例数を確保しやすく、希少疾患領域では特に有効なアプローチとなる。

その一方で、国ごとの治療パスや文化的背景による選好差など、転用可能性の問題が生じる可能性がある。特性や水準の意味が国によって異なる場合、結果の平均値のみを解釈することは適切でない場合もある。そのため、国際共同選好研究では、国差を不均一性として捉え、解析計画の段階でどのように扱うかを明示することが重要である。

IMI PREFER では、結果を規制当局や HTA に提出することを想定する場合、計画段階において不均一性や国差をどのように解釈・報告するかについて、早期に当局と対話することが推奨されている。特に、グローバルデータをローカルの意思決定に用いる場合には、どの範囲まで国差を許容するかなどを事前に合意しておくことが重要である。

このように、定量的患者選好研究においては、国や文化の違いが選好に与える影響を踏まえ、国内研究、海外研究の転用、国際共同研究のいずれを選択するかを研究目的や意思決定の文脈に応じて検討する必要がある。いずれのアプローチを取る場合であっても、選好データの解釈や活用に際しては、国差・文化差を意識した慎重な検討と透明な報告が不可欠である。

## Appendix. 3 内的妥当性テストの例

## ① 安定性 (Stability, repeated question)

下図のように、同じ質問を繰り返した時に異なる選択肢を選択した場合、『非合理』と判断する。

| 特性         | 治療A                                 | 治療B                      | 質問 #2 |
|------------|-------------------------------------|--------------------------|-------|
| 痛みの軽減      | 10 %の軽減                             | 30 %の軽減                  |       |
| 副作用の発現     | 中等度                                 | 軽度                       |       |
| 月あたりのコスト   | 1,000円                              | 2,000円                   |       |
| どちらを選びますか？ | <input checked="" type="checkbox"/> | <input type="checkbox"/> |       |

| 質問 #6 |  | 特性         | 治療A                      | 治療B                                 |
|-------|--|------------|--------------------------|-------------------------------------|
|       |  | 痛みの軽減      | 10 %の軽減                  | 30 %の軽減                             |
|       |  | 副作用の発現     | 中等度                      | 軽度                                  |
|       |  | 月あたりのコスト   | 1,000円                   | 2,000円                              |
|       |  | どちらを選びますか？ | <input type="checkbox"/> | <input checked="" type="checkbox"/> |

## ② セット内優越ペア (Within-set dominated pairs)

明らかに全ての特性で治療 B が優れているときに「治療 A」が選択されるのは、『非合理』と判断する。

| 特性         | 治療A                                 | 治療B                      |
|------------|-------------------------------------|--------------------------|
| 痛みの軽減      | 10 %の軽減                             | 30 %の軽減                  |
| 副作用の発現     | 中等度                                 | 軽度                       |
| 月あたりのコスト   | 2,000円                              | 1,000円                   |
| どちらを選びますか？ | <input checked="" type="checkbox"/> | <input type="checkbox"/> |

## ③ セット間優越ペア (Across-set dominated pairs)

質問#2 で「痛みの軽減」と「副作用の発現」が劣り「費用」が優れている「治療 A」が選ばれたなら、「副作用の発現」の水準を良い方向に変更しただけの質問#6 でも同じように選択されるべきである。もし質問#6 で「治療 B」が選択される場合は、『非合理』と判断する。

| 特性         | 治療A                                 | 治療B                      | 質問 #2 |
|------------|-------------------------------------|--------------------------|-------|
| 痛みの軽減      | 10 %の軽減                             | 30 %の軽減                  |       |
| 副作用の発現     | 中等度                                 | 軽度                       |       |
| 月あたりのコスト   | 1,000円                              | 2,000円                   |       |
| どちらを選びますか？ | <input checked="" type="checkbox"/> | <input type="checkbox"/> |       |

| 質問 #6 |  | 特性         | 治療A                      | 治療B                                 |
|-------|--|------------|--------------------------|-------------------------------------|
|       |  | 痛みの軽減      | 10 %の軽減                  | 30 %の軽減                             |
|       |  | 副作用の発現     | なし                       | 軽度                                  |
|       |  | 月あたりのコスト   | 1,000円                   | 2,000円                              |
|       |  | どちらを選びますか？ | <input type="checkbox"/> | <input checked="" type="checkbox"/> |

## ④ 推移性 (Transitivity)

質問#2 で「治療 A > 治療 B」、質問#4 で「治療 A (質問#2 の治療 B と同じ設定) > 治療 B」の場合、「質問#2 の治療 A > 質問#4 の治療 B」となるのが合理的である。もし質問#6 で「治療 B」が選択される場合は、『非合理』と判断する。

なお、このような設計は効率的な実験デザインでは自然に発生しないため、別途設計が必要となる。

| 特性         | 治療A                                 | 治療B                      | 質問 #2 |
|------------|-------------------------------------|--------------------------|-------|
| 痛みの軽減      | 25 %の軽減                             | 50 %の軽減                  |       |
| 副作用の発現     | 中等度                                 | 軽度                       |       |
| 月あたりのコスト   | 1,000円                              | 3,000円                   |       |
| どちらを選びますか？ | <input checked="" type="checkbox"/> | <input type="checkbox"/> |       |

| 質問 #4 |  | 特性         | 治療A                                 | 治療B                      |
|-------|--|------------|-------------------------------------|--------------------------|
|       |  | 痛みの軽減      | 50 %の軽減                             | 5 %の軽減                   |
|       |  | 副作用の発現     | 中等度                                 | なし                       |
|       |  | 月あたりのコスト   | 3,000円                              | 2,000円                   |
|       |  | どちらを選びますか？ | <input checked="" type="checkbox"/> | <input type="checkbox"/> |

| 質問 #6 |  | 特性         | 治療A                      | 治療B                                 |
|-------|--|------------|--------------------------|-------------------------------------|
|       |  | 痛みの軽減      | 25 %の軽減                  | 5 %の軽減                              |
|       |  | 副作用の発現     | 中等度                      | なし                                  |
|       |  | 月あたりのコスト   | 1,000円                   | 2,000円                              |
|       |  | どちらを選びますか？ | <input type="checkbox"/> | <input checked="" type="checkbox"/> |

## ⑤ 特性優越 (Attribute dominance, noncompensatory preferences)

選択実験では通常「補償的選好 (compensatory preferences)」を仮定する。補償的選好とは、ある特性が悪化しても別の特性が十分改善されれば全体として

その選択肢を選ぶ可能性があるという考え方である。特性優越（Attribute dominance）はこの逆で、ある特性だけでほぼ全ての選択を決めるパターンのことをいう。

例：

ある質問票で「副作用のリスク」「治療効果」「服薬頻度」という3つの特性があるとする。補償的選好なら、患者は「副作用が少し増えるが、効果が大幅に改善するなら選ぶ」ことがある。特性優越の場合は、患者が「副作用が最も少ない選択肢」を常を選び、他の特性（効果や頻度）をほとんど考慮しない。

以下の例では、質問#2、#4、#6で「痛みの軽減」の特性だけを見て治療選択をしている。

| 特性         | 治療A                      | 治療B                                 | 質問 #2                    |
|------------|--------------------------|-------------------------------------|--------------------------|
| 痛みの軽減      | 25 %の軽減                  | 50 %の軽減                             |                          |
| 副作用の発現     | 中等度                      | 軽度                                  |                          |
| 月あたりのコスト   | 1,000円                   | 3,000円                              |                          |
| どちらを選びますか？ | <input type="checkbox"/> | <input checked="" type="checkbox"/> |                          |
| 質問 #4      | 特性                       | 治療A                                 | 治療B                      |
|            | 痛みの軽減                    | 50 %の軽減                             | 5 %の軽減                   |
|            | 副作用の発現                   | 中等度                                 | なし                       |
|            | 月あたりのコスト                 | 3,000円                              | 2,000円                   |
|            | どちらを選びますか？               | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| 質問 #6      | 特性                       | 治療A                                 | 治療B                      |
|            | 痛みの軽減                    | 25 %の軽減                             | 5 %の軽減                   |
|            | 副作用の発現                   | 中等度                                 | なし                       |
|            | 月あたりのコスト                 | 1,000円                              | 2,000円                   |
|            | どちらを選びますか？               | <input checked="" type="checkbox"/> | <input type="checkbox"/> |

## Appendix. 4 プレテストの活用例

| 筆頭著者、年                         | プレテストで確認した内容                                                         | プレテストの目的   | プレテストの方法                                    | プレテストの結果                                                                                                                           |
|--------------------------------|----------------------------------------------------------------------|------------|---------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| Abdel-All, 2019 <sup>59)</sup> | 理解（質問を意図した通りに理解する能力）、情報の検索（質問について考え、結論を導くこと）、判断および回答の選択              | DCE 調査票の改訂 | 思考発話法による認知インタビュー及びフォーカスグループ                 | DCE は ASHA（インドの地域保健ワーカー）に好意的に受け入れられ、提示された選択肢集合を理解するのに困難はなかった。例示付きの導入文は仮説的な DCE であることを説明する上で役立ち、ASHA が DCE の認知的理解を備えていることが確認された。    |
| Katz, 2018 <sup>60)</sup>      | 全体的な理解度および選択課題の遂行上の困難を評価した。加えて、単調性（支配的選択肢の検証）、安定性（同一選択肢の再現）、非連続性を評価。 | DCE 調査票の改訂 | 思考発話法による認知インタビュー                            | 「ほとんどの参加者は困難なく DCE を完了したが、一部は完全に集中しておらず、認知的負荷や疲労の兆候もみられた。少数の参加者は、2つの仮想薬剤から選ぶという概念に苦労した。」…「多くの参加者は、選択課題を良好に理解し、2つ以上の特性を考慮して選択していた。」 |
| Klojgaard, 2012 <sup>61)</sup> | 特性内容を明確かつ簡潔な方法で理解すること、特にリスク特性を重視                                     | DCE 調査票の改訂 | 患者インタビュー内で「フレーミング（提示文の認知・理解）」「複雑さおよびパイロット時の | 「6 特性は多すぎることで明確となり、回答者は最初の特性を後半では記憶していられず、レキシコグラフィック（特定属性のみで判断）や単純化のヒューリスティックを用いる回答がみら                                             |

| 筆頭著者、<br>年 | プレテストで確認<br>した内容 | プレテスト<br>の目的 | プレテスト<br>の方法 | プレテストの結果                                                                                                                                                                           |
|------------|------------------|--------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|            |                  |              | 体験」について議論    | れた。」…「デンマーク語で ‘ability’ と ‘aim’ が似ているため誤読が生じ、特性の理解を困難にし、再構成の必要性が示された。」…「リスク特性は理解されていた。」…「課題の数が多いと認知的限界に達するため、レベル数を減らすべきである。」…「テキストを読むだけで課題が理解できたが、実際に課題を遂行する際は難しく、より丁寧な説明が必要であった。」 |

## 6. まとめ

前述のとおり、患者選好研究の方法論は臨床試験の方法論とは異なるところがあれば存在する。実際に手を動かしてみると、定量的な質問票の特性や水準をガイドラインに従って設定することができたとしても、質問票の各選択肢の特性と水準をどのような組み合わせにすべきか、症例数はどのように設定すべきか、などは各規制当局のガイドラインなどでも明確には示されていない。本書ではこうした実務で直面する課題に対して実際に手を動かせるように記載してきたが、一方で方法や一般的な進め方が文書上の記載を読んで理解できても、実際の経験を積まないと実施できないものもある。例えば、定性的研究のフォーカスグループにて、患者が話しやすい雰囲気を形成したり、会話の展開に応じてトピックを掘り下げたりなど、その場での対応が求められる司会の役割などは、経験を要する事項の一例である。

経験や専門性が社内に不足していて実施が難しいような場合には、国内でも CRO も含め市場調査を請け負う会社にこれらのケイパビリティがあり、そういった会社の支援を受けることも実行するための一つの手段となる。それでも患者選好研究の実施に際し最初の内は困難を伴うことが多く、エビデンスを率先して蓄積する労力の方が大きいと感じられるかもしれない。

だが患者の考えていることが医薬品開発に反映され、患者にとって有用だと感じられる医薬品が開発される世界にしていくためには、各関係者が患者選好研究をエビデンスとして積み重ねていくことが重要である。患者選好研究が増えることで、国や年齢など様々な性質を持つ患者の選好の多様性への理解が進み、多様性に応じた医薬品の開発可能性が上がるだろう。また、多くの疾患で患者選好研究が蓄積されてくれば、新規の患者選好研究をしなくても過去の患者選好研究を用いて新薬の開発に患者の声を反映できることも増えるだろう。さらに患者選好の理解が進めば、医薬品開発のみならず、実際の医療現場での処方判断にも活用可能で、そのことがより良い医療にもつながりうる。本書がそのような患者選好研究のエビデンス蓄積のきっかけとなれば幸いである。

## 7. 参考文献

- 1) 日本製薬工業協会. Patient-Focused Drug Development への患者経験情報の活用, [Available from: [https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_202504\\_PFDD.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202504_PFDD.html)] (アクセス最終日 2026 年 7 月 25 日)
- 2) FDA. 21st Century Cures Act, [Available from: <https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf>] (アクセス最終日 2026 年 7 月 25 日)
- 3) FDA. Assessment of the Use of Patient Experience Data in Regulatory Decision-Making, [Available from: <https://www.fda.gov/media/150405/download?attachment>] (アクセス最終日 2026 年 7 月 25 日)
- 4) D Almeida, D Umuhire, R Gonzalez-Quevedo, A Antonio, JG Burgos, P Verpillat, et al (2024). Leveraging patient experience data to guide medicines development, regulation, access decisions and clinical care in the EU, *Front Med*, 11:1408636.
- 5) SPH Nouwens, SM Marceta, M Bui, DMAH van Dijk, CGM Groothuis-Oudshoorn, Jorien Veldwijk, et al (2025). The Evolving landscape of disease choice experiments in health economics: a systematic review, *Pharmacoeconomics*, 43:879-936.
- 6) M Zhang, X He, J Wu, F Xie (2023). Differences between physician and patient preferences for cancer treatments: a systematic review, *BMC Cancer*, 23:1126.
- 7) AC Muhlbacher, C Juhnke (2013). Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making?, *Appl Health Econ Health Policy*, 11:163-80.
- 8) SR Wilson, P Strub, AS Buist, SB Knowles, PW Lavori, J Lapidus, et al (2010). Shared treatment decision making improves adherence and outcomes in poorly controlled asthma, *Am J Respir Crit Care Med.*, 181:566-77.
- 9) MV Korff, W Katon, C Rutter, E Ludman, G Simon, E Lin, et al (2003). Effect on disability outcomes of a depression relapse prevention program, *Psychosom Med.*, 65:938-43.
- 10) 独立行政法人 医薬品医療機器総合機構, ICH-E22 患者選好研究に関する一般指針, [Available from: <https://www.pmda.go.jp/int-activities/int-harmony/ich/0127.html>] (アクセス最終日 2026 年 7 月 25 日)
- 11) FDA. Patient-Focused Drug Development: Collecting Comprehensive and Representative Input. [Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-collecting-comprehensive-and-representative-input>] (アクセス最終日 2026 年 7 月 25 日)

- 12) FDA. Patient-Focused Drug Development: Methods to Identify What Is Important to Patients. [Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-methods-identify-what-important-patients>] (アクセス最終日 2026 年 7 月 25 日)
- 13) FDA. Patient Preference Information - Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling. [Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-preference-information-voluntary-submission-review-premarket-approval-applications>] (アクセス最終日 2026 年 7 月 25 日)
- 14) FDA. Benefit-Risk Assessment for New Drug and Biological Products. [Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/benefit-risk-assessment-new-drug-and-biological-products>] (アクセス最終日 2026 年 7 月 25 日)
- 15) EMA. Patient experience data reflection paper. [Available from: <https://www.ema.europa.eu/en/patient-experience-data-ped-reflection-paper>] (アクセス最終日 2026 年 7 月 25 日)
- 16) IMI PREFER. PREFER Recommendations - Why, when and how to assess and use patient preferences in medical product decision-making [Available from: <https://zenodo.org/records/6592304>] (アクセス最終日 2026 年 7 月 25 日)
- 17) CIOMS Working Group IV. Benefit-Risk Balance for Marketed Drugs: Evaluating Safety Signals [Available from: <https://cioms.ch/wp-content/uploads/2017/01/benefit-risk.pdf>] (アクセス最終日 2026 年 7 月 25 日)
- 18) CIOMS Working Group XII. Benefit-Risk balance for medicinal products [<https://cioms.ch/publications/product/benefit-risk-balance-for-medicinal-products/>] (アクセス最終日 2026 年 7 月 25 日)
- 19) S Heidenreich, M Ross, B Flamion, AP Beyer (2025). A patient-centric benefit-risk assessment of daridorexant for the treatment of insomnia disorder using patient preference data collected in two phase 3 clinical trials, *Sleep Epidemiology*, 5: 100108.
- 20) L Nyanchoka, C Tudur-Smith, R Porcher, D Hren (2019). Key stakeholders' perspective and experiences with defining, identifying and displaying gaps in health research: a qualitative study protocol, *BMJ Open*, 9:e027926.
- 21) J Maher, GS Fernandez, C Robinson, KI Berger, K Cutts, M Quaife, et al(2026). Experiences and treatment perspectives of adolescents and young adults with achondroplasia and caregivers: a qualitative interview study, *Qual Life Res*, 35: 145.
- 22) LS Nowell, JM Norris, DE White, NJ Moules (2017). Thematic analysis: striving to meet

- the trustworthiness criteria, *International Journal of Qualitative Methods*, 16: 1-13.
- 23) SN Halpin (2024). Inter-coder agreement in qualitative Coding: consideration for its use, *American Journal of Qualitative Research*, 8: 23-42.
- 24) KE Lasch, P Marquis, M Vigneux, L Abetz, B Arnould, M Bayliss, et al (2010). PRO development: rigorous qualitative research as the crucial foundation, *Qual Life Res*, 19: 1087-1096.
- 25) LF Higueta-Gutierrez, DA Estrada-Mesa, JA Cardona-Arias (2021). Preferences in a group of patients with cancer: a grounded theory, *Patient Preference and Adherence*, 15: 2313-2326.
- 26) I Woollacott, G Morgan, P Chowdary, J O'Hara, B Franks, E van Overbeeke, et al (2022). *Haemophilia*, 28: 588-609.
- 27) C Carmichael, I Jouravskaya, E Collins, D Burns, JL Poon, H Kitchen, et al (2025). Patient experience of treatment with Tirzepatide for weight management: exit interviews from SURMOUNT-4, *Patient*, 18: 225-236.
- 28) R Winnette, S Martin, N Harris, LS Deal (2021). Development of the alopecia areata patient priority outcomes instrument: a qualitative study, *Dermatol Ther*, 11: 599-613.
- 29) FDA. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. [Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims>] (アクセス最終日 2026 年 7 月 25 日)
- 30) B Saunders, J Sim, T Kingstone, S Baker, J Waterfield, B Bartlam, et al (2017). Saturation in qualitative research: exploring its conceptualization and operationalization, *Qual Quant*, 52: 1893-907.
- 31) U Flick, (監訳)小田博志 (2011). 質的研究入門 – “人間の科学”のための方法論 (新版), 春秋社.
- 32) NK Denzin (2017). *The Research Act*, Routledge.
- 33) JFP Bridges, AB Hauber, D Marchall, A Lloyd, LA Prosser, DA Regier, et al (2011). Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force, *Value Health*, 14: 403-13.
- 34) FR Johnson, E Lancsar, D Marchall, V Kilambi, A Muhlbacher, DA Regier, et al (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force, *Value Health*, 16: 3-13.
- 35) B Hauber, JM Gonzalez, CGM Groothuis-Oudshoorn, T Prior, D Marshall, C Cunningham, et al (2016). Statistical methods for the analysis of discrete-choice

- experiments: a report of the ISPOR conjoint analysis good research practices task force, *Value Health*, 19: 300-15.
- 36) Medical Device Innovation Consortium. Using Patient Preference Information in the Design of Clinical Trials Framework. [Available from: [https://mdic.org/wp-content/uploads/2024/04/MDIC-2022-SPI-Clinical-Trials-Report\\_v799.pdf](https://mdic.org/wp-content/uploads/2024/04/MDIC-2022-SPI-Clinical-Trials-Report_v799.pdf)] (アクセス最終日 2026 年 7 月 25 日)
- 37) N Hosogaya, T Takazono, K Kurazono, S Hiroi, K Iwasaki, T Takeshima, et al (2023). Estimation of intangible costs for factors associated with oral antiviral drugs for COVID-19 treatment: a conjoint analysis in Japan, *Adv Ther*, 40: 3525-42.
- 38) N Sanko (2001). Guidelines for Stated Preference Experiment Design. [Available from: [https://b.kobe-u.ac.jp/sanko/pub/Sanko2001\\_1.pdf](https://b.kobe-u.ac.jp/sanko/pub/Sanko2001_1.pdf)] (アクセス最終日 2026 年 7 月 25 日)
- 39) T Takeshima, A Yuasa, YM Lloyd, LP Watanabe, K Kamei, A Terasawa, et al (2024). A discrete-choice experiment assessing the patient preferences and real-world experiences of patients with migraine in Japan, *Neurol Ther*, 13: 1661-83.
- 40) FR Johnson, JC Yang, SD Reed (2019). The internal validity of discrete choice experiment data: a testing tool for quantitative assessments, *Value Health*, 22: 157-60.
- 41) FR Johnson, RL DiSantostefano, JC Yang, SD Reed, J Streffer, B Levitan (2019). Something is better than nothing: the value of active intervention in stated preferences for treatments to delay onset of Alzheimer's disease symptoms, *Value Health*, 22: 1063-9.
- 42) D Wild, A Grove, M Martin, S Eremenco, S McElroy, A Verjee-Lorenz, et al (2005). Principles of good practice for the translation and cultural adaption process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaption, *Value Health*, 8: 94-104.
- 43) NB Campoamor, CJ Guerrini, WB Brooks, JFP Bridges, NL Crossnohere (2024). Pretesting discrete-choice experiments: a guide for researchers, *Patient*, 17: 109-20.
- 44) Department of Global Health, University of Washington. Pre-Testing and Piloting. [Available from: <https://sites.uw.edu/surveillancetoolkit/conduct/pre-testing-and-piloting/>] (アクセス最終日 2026 年 7 月 25 日)
- 45) B Jones, C Berlin, N Cook, S Dickinson, T Lyu, M Zhou (2024). Biostatistics in Biopharmaceutical Research and Development, *Springer*.
- 46) NS Cook, GJ Criner, PR Burgel, K Mycok, T Gardner, P Mellor, et al (2022). People living with moderate-to-severe COPD prefer improvement of daily symptoms over the improvement of exacerbations: a multicountry patient preference study, *ERJ Open Res*, 8: 00686-2021.
- 47) SY Assele, M Meulders, M Vandebroek (2023). Sample size selection for discrete choice

- experiments using design features, *Journal of choice modelling*, 49: 100436.
- 48) BK Orme (2010). Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research, *Research Publishers*.
- 49) EW de Bekker-Grob, B Donkers, MF Jonker, EA Stolk (2015). Sample size requirements for discrete-choice experiments in healthcare: a practical guide, *Patient*, 8: 373-84.
- 50) FR Johnson, W Adamowicz, C Groothuis-Oudshoorn (2025). What can discrete-choice experiments tell us about patient preferences? An introduction to quantitative analysis of choice data, *Patient*, 18: 425-40.
- 51) AC Muhlbacher, A Kaczynski, P Zweifel, FR Johnson (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview, *Health Econ Rev*, 6: 2.
- 52) JJ Louviere, TN Flynn, AAJ Marley (2015). Best-Worst Scaling: Theory, Methods and Applications, *Cambridge University Press*.
- 53) A Angelis, P Kanavos (2017). Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in health technology assessment and beyond: the advance value framework, *Social Science and Medicine*, 188: 137-56.
- 54) T Tervonen, F Pignatti, D Postmus (2019). From individual to population preferences: comparison of discrete choice and dirichlet models for treatment benefit-risk tradeoffs, *Med Decis Making*, 39: 879-85.
- 55) B Hauber, J Coulter (2020). Using the threshold technique to elicit patient preferences: an introduction to the method and an overview of existing empirical applications, *Appl Health Econ Health Policy*, 18: 31-46.
- 56) J Veldwijk, E de Bekker-Grob, J Juhaeri, E van Overbeeke, S Tcherny-Lessenot, CA Pinto, et al (2023). Suitability of preference methods across the medical product lifecycle: a multicriteria decision analysis, *Value Health*, 26: 579-88.
- 57) G Dranitsaris, Q Zhang, L Mu, C Weyrer, E Drysdale, P Neumann, et al (2023). Therapeutic preference for Alzheimer's disease treatments: a discrete choice experiment with caregivers and neurologists, *Alz Res Therapy*, 15: 60.
- 58) C Vass, M Boeri, S Karim, D Marshall, B Craig, K Ho, et al (2022). Accounting for preference heterogeneity in discrete-choice experiments: an ISPOR special interest group report, *Value Health*, 25: 684-94.
- 59) M Abdel-All, B Angell, S Jan, M Howell, K Howard, S Abimbola, et al (2019). What do community health workers want? Findings of a discrete choice experiment among Accredited Social Health Activists (ASHAs) in India, *BMJ Global Health*, 4: e001509.
- 60) DA Katz, KR Stewart, M Paez, MWV Weg, KM Grant, C Hamlin, et al (2018). Development of a Discrete Choice Experiment (DCE) questionnaire to understand

Veterans' preferences for tobacco treatment in primary care, *Patient*, 11: 649-63.

- 61) ME Klotjgaard, M Bech, R Sogaard (2012). Designing a stated choice experiment: the value of a qualitative process, *Journal of Choice Modeling*, 5: 1-18.

日本製薬工業協会 医薬品評価委員会 2026 年度タスクフォース 9-5  
「医薬品開発・エビデンス創出への患者経験データの利活用検討」

タスクフォースメンバー

| 氏名     | 所属               |
|--------|------------------|
| 和泉 奈穂子 | ノバルティスファーマ株式会社   |
| 岩崎 敬介  | 中外製薬株式会社         |
| 奥田 伊奈葉 | グラクソ・スミスクライン株式会社 |
| 川崎 温子  | 中外製薬株式会社         |
| 白石 亜矢子 | ヤンセンファーマ株式会社     |
| 杉谷 康雄* | 中外製薬株式会社         |
| 富金原 悟  | 小野薬品株式会社         |
| 吉田 詩織  | 帝人ファーマ株式会社       |

\*: タスクフォースリード