

---

## 業界向け指針

患者報告アウトカム (**Patient-Reported Outcome: PRO**) の測定法：  
医薬品/医療機器における適応申請のための方法

米国連邦保健福祉省

食品医薬品庁

医薬品評価研究センター (CDER)

生物製剤評価研究センター (CBER)

医療機器・放射線製品センター (CDRH)

2009年12月

臨床/医学

監訳

*International Society for Pharmacoeconomics and Outcomes Research (ISPOR)*

(国際医薬経済・アウトカム研究学会) 日本部会 ワーキンググループ

2011年5月

---

## 業界向け指針

患者報告アウトカム（**Patient-Reported Outcome: PRO**）の測定法：

医薬品/医療機器における適応申請のための方法

追加コピー請求先：

薬剤情報部通信局  
医薬品評価研究センター  
食品医薬品庁

10903 New Hampshire Ave., Bldg. 51, rm. 2201 Silver Spring, MD 20993-0002  
電話：301-796-3400；Fax：301-847-8714；電子メール：druginfo@fda.hhs.gov  
<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>

あるいは

通信、アウトリーチ、および開発局、HFM-40  
生物製剤評価研究センター  
食品医薬品庁

1401 Rockville Pike, Suite 200N, Rockville, MD 20852-1448  
電話：800-835-4709 または 301-827-1800；電子メール：ocod@fda.hhs.gov  
<http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/default.htm>

あるいは

通信、教育、放射線プログラム局  
小規模製造業者、国際、消費者支援部 HFZ-220  
医療機器・放射線製品センター  
食品医薬品庁

1350 Piccard Drive, Rockville, MD 20850-4307  
DSMICA 電子メール：dsmica@cdrh.fda.gov  
DSMICA Fax：301-443-8818

(電話) 製造支援：800-638-2041 または 301-443-6597  
(電話) 国際スタッフ：301-827-3993

<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/default.htm>

米国連邦保健福祉省

食品医薬品庁

医薬品評価研究センター（CDER）

生物製剤評価研究センター（CBER）

医療機器・放射線製品センター（CDRH）

2009年12月

臨床/医学

## 目次

I. 緒言	1
II. 背景	2
III. PRO 尺度の評価	3
A. エンドポイントモデル	3
B. PRO 尺度の選択	5
C. PRO 尺度の概念枠組み	8
1. 測定する概念	8
2. 対象集団	11
D. 内容妥当性	13
1. 項目の作成	13
2. データ収集法と尺度記入法	15
3. 想起期間	16
4. 回答方法 (回答選択肢)	16
5. 尺度のフォーマット、説明、研修	18
6. 患者の理解	18
7. 項目とドメインのスコアリング	19
8. 回答者や記入者の負担	20
E. 信頼性、その他の妥当性、変化検出能力	21
1. 信頼性	21
2. その他の妥当性	23
3. 変化検出能力	24
F. 尺度の修正	24
G. 特定集団を対象とした PRO 尺度	25
1. 小児と青年	25
2. 認知障害患者またはコミュニケーションが不可能な患者	26
3. 文化または言語によるサブグループ	26
IV. 臨床試験のデザイン	26
A. プロトコルの一般的留意点	26
1. 盲検化とランダム化	27
2. 臨床試験の品質管理	27
3. 欠測データの取り扱い	28
B. 評価の頻度	28
C. 臨床試験の期間	29
D. 複数のエンドポイントに関するデザイン上の留意点	29
E. 奏効例の定義を使用した臨床試験結果に関する解釈の計画	30
F. 電子的な PRO 尺度を用いた場合に特有の留意点	31
V. データ解析	33
A. 統計解析に関する一般的留意点	33
B. 複数のエンドポイントを用いる場合の統計解析留意点	33
C. 複合エンドポイントを用いる場合の統計解析留意点	35
D. 患者レベルの欠測データに関する統計解析留意点	36
1. 同ドメイン内の欠測項目	36
2. ドメイン全体または測定全体の欠測	36
E. 臨床試験結果の解釈	38
用語集	39
付録：FDA が審査する、PRO 尺度に関する情報	44

## 拘束力をもたない勧告

### 業界向け指針<sup>1</sup>

#### 患者報告アウトカム

(Patient-Reported Outcome: PRO) の測定法 :

### 医薬品/医療機器における適応申請のための方法

本指針は、当該トピックに関する食品医薬品庁（FDA）の現時点での見解を示すものである。本指針は何人に対しても、いかなる権利をも発生や賦与するものではなく、FDA や一般国民を拘束するものではない。該当する法律や規制の要件を満たすのであれば、これに代わるアプローチを用いても差し支えない。代替アプローチの検討を希望する場合は、FDA の本指針執行担当者まで連絡されたい。FDA のしかるべき担当者が不明の場合は、本指針のタイトルページに記載された番号まで電話連絡されたい。

## I. 緒言

本指針には、医薬品/医療機器の適応の認可 **申請** に使われる、既存の、修正した、または新規に作られた **患者報告アウトカム (Patient-Reported Outcome : PRO) 尺度** を、食品医薬品庁（FDA）がどのように審査・評価するかが記載されている<sup>2</sup>。PRO尺度（**質問票**と、質問票の使用に必要な情報・文書）は、医薬品/医療機器の臨床試験における **治療ベネフィット** または **リスク** を測定するために使用される **PROデータ** を収集するための手段である。本指針は、医薬品/医療機器の適応申請に関する評価以外を目的としたPRO尺度の使用については対象としない。また、本指針は、疾患個別の問題については対象としない。特定の疾患の臨床試験エンドポイントに関する指針は、FDAの様々なウェブサイトから入手することができる<sup>3</sup>。

<sup>1</sup> 本指針は、食品医薬品庁に属する医薬品評価研究センター（CDER）が、生物製剤評価研究センター（CBER）および医療機器・放射線製品センター（CDRH）との共同で作成した。

<sup>2</sup> 本指針で用いている「**適応表示 (label)**」とは、臨床医が患者の治療に使用する FDA からの認可を受けた医薬品/医療機器に関する情報を示す。処方薬(生物製剤を含む)表示に関する規制については、21 CFR 201.56 と 201.57 を参照すること。虚偽ではない、または誤解を招かない適応表示の必要性についてはセクション 201.56 に記載されている。医療機器表示については 21 CFR パート 801 を参照のこと。輸血用血液および血液製剤については、21 CFR 606.122 を参照のこと。

<sup>3</sup> 以下の FDA ウェブサイトを参照のこと :

<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm> (CDER)、

<http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/default.htm> (CBER) および

<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/default.htm> (CDRH)。

## 拘束力をもたない勧告

治験依頼者は、本指針で示した審査内容を明確に検討することにより、医薬品/医療機器開発プロセス中に行う FDA との協議を一層効率よく行うこと、PRO 尺度と臨床試験において収集される PRO データが十分かどうかに関する FDA の審査を効率的にすること、医薬品/医療機器認可時点における治療効果の結論を導く際に使用する患者の視点について、最適な情報を提供すること、が可能となる。PRO 尺度の開発は繰り返しのプロセスであり、PRO 尺度の開発における正しい方法は 1 つではないことを FDA は認識している。FDA の審査で指摘された課題には、様々な戦略や方法を駆使して対処することが可能である。

「用語集」では、本指針で用いられる多くの用語について定義している。用語集に記載される言葉および表現は、初出時に**太い斜体文字**で表記される。特に、治験依頼者には「**PRO 尺度の概念の枠組み**」、「**エンドポイントモデル**」、および「**内容妥当性**」について十分に理解することが勧められる。

本指針を含む FDA の指針文書は、法的に強制力のある責任を求めているわけではない。その代わりに、指針文書はあるトピックに対する当局の現時点での見解について述べており、特定の規制または法的要求事項が示されない限り、推奨事項としてののみ受け止めるべきである。当局の指針文書において「べき (*should*)」という用語が用いられている場合は、提案または推奨はされるが要求はされていないことを表している。

## II. 背景

PRO とは、患者の回答について、臨床医や他の誰の解釈も介さず、患者から直接得られる患者の健康状態に関するすべての報告である。アウトカムは、絶対的な事項（例えば、**症状の重症度**、**徴候**、疾病の状態）または、以前の測定からの変化として測定することができる。臨床試験において、PRO 尺度は、1 つ以上の**概念**（すなわち、測定された**何らかのもの**、例えば、症状や症候群、特定機能や機能群に対する効果、または、健康状態の重症度を示す症候群や機能群などの測定対象）に対する医療介入の効果測定のために使用することができる。

一般に、適切にデザインされた研究で明確に定義された信頼性の高い PRO 尺度で測定された所見は、適応申請の範囲が、尺度の測定能力と矛盾がない場合、医薬品/医療機器の適応申請に使用することができる。FDA に提出すべき根拠の量と種類は、

## 拘束力をもたない勧告

他のどのデータに基づく適応申請の場合と同じである。患者が最もよく知っている、患者の視点から測定することが最も適しているような概念の測定においては、PRO 尺度を使うことが推奨される。PRO 尺度は、医師の評価に基づく尺度と同様、意図する概念を測定していることを示す必要があり、FDA は、特定の PRO 尺度が、対象概念を測定することを示す根拠の有無を審査する。適応申請に最もよく使われる、PRO 尺度を使って測定される概念は、患者の症状、徴候、疾病の状態に直接関連した機能に関する概念である。PRO の測定結果は多くの場合、疾病（例：心不全または喘息）が、患者の視点から健康と機能にどのような影響を及ぼすか、を表わしている。

申請される効能は一般に、適応表示の「適応と方法」または「臨床研究」のセクションに記載されるが、どのセクションに記載することもできる。どの適応表示セクションにかかわらず、ここに記載される PRO 尺度の評価原則が適用される。

### III. PRO 尺度の評価

医薬品/医療機器の適応申請を目的とした PRO 尺度の評価には、以下の点に留意する必要がある：

- 臨床試験に登録される患者集団
- 臨床試験の目的とデザイン
- PRO 尺度の概念の枠組み
- PRO 尺度の **測定特性**

PRO を測定する目的は患者の経験を捉えることであり、対象患者集団からの有用性を示す根拠がない尺度を用いた場合、信頼性のある測定はできない。治験依頼者は、尺度の開発における患者の登録と特定の用途（すなわち対象集団、状態）における尺度の能力についての文書化された根拠を提出すべきである。既存の尺度は、臨床試験に登録した患者集団において、申請された概念を確実に測定できることが示されれば、適応申請に使用することができる。

#### A. エンドポイントモデル

治験依頼者は、尺度の開発と機能を、目的とする役割の視点から審査し、適切な統計手法を計画し適用できるように、臨床試験において PRO エンドポイントが果たす

## 拘束力をもたない勧告

役割（すなわち、主要エンドポイント、重要な副次エンドポイント、あるいは探索的エンドポイント）を定義すべきである。これらのアプローチは、エンドポイントモデルと呼ばれる方法で計画することが重要な意味を持つ。

図1および2に、エンドポイントモデルの例を示す。図1において、PRO 症状評価は、疾患 X の治療の適応を明らかにするための副次エンドポイントであり、主要エンドポイントは生理学的測定である。この場合、臨床試験においては、副次エンドポイントを達成する前に、まず、生理学的エンドポイントを達成する必要がある。図2においては、PRO 症状評価が疾患 Y に伴う症状の治療適応を明らかにするための主要な臨床試験エンドポイントであり、身体的能力と限界の測定は重要な副次エンドポイントとなる。PRO 尺度の適切性は、役割およびエンドポイントモデルに示される、その他の臨床試験エンドポイントとの関係に依存する。エンドポイントモデルは、臨床試験の目的を満たし、測定する概念に対応する適応申請に役立つ根拠を得るための、PRO 尺度に対する需要を具体的に示すものである。

図1. エンドポイントモデル：疾患 X の治療

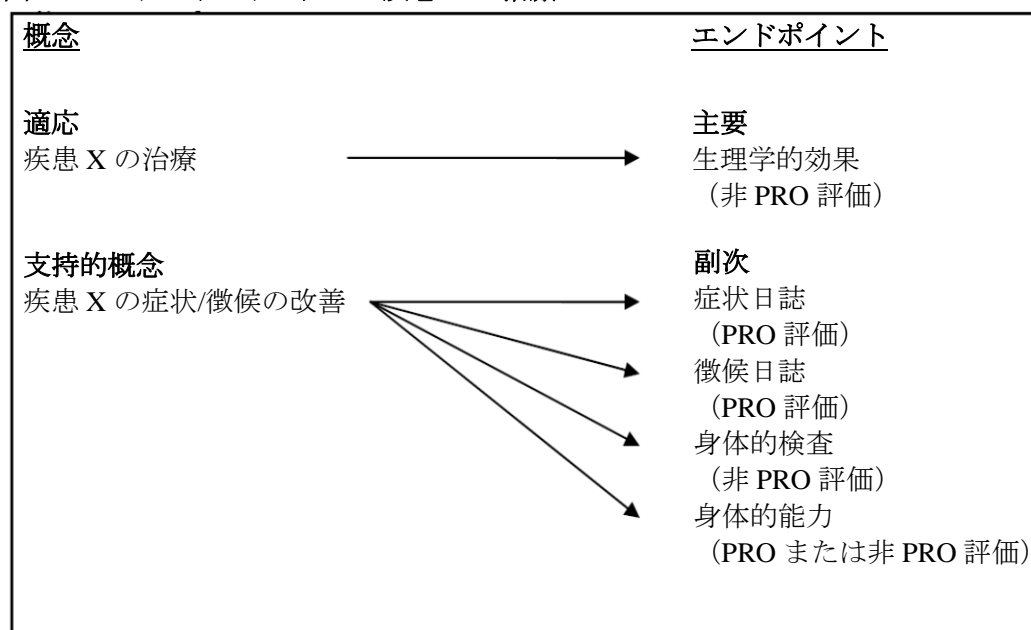
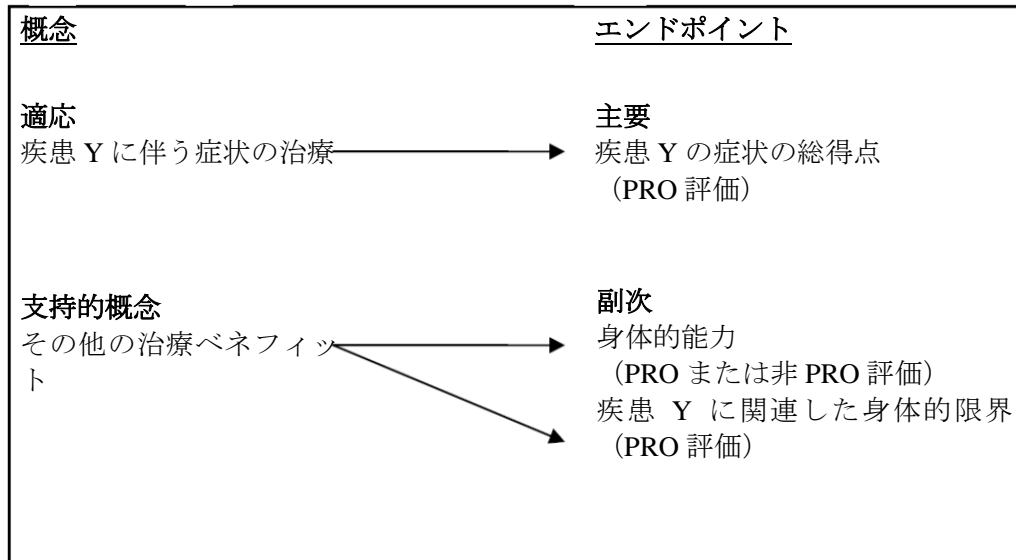


図 2. エンドポイントモデル：疾患 Y に伴う症状の治療



可能性のある適応申請を助け、予定する概念の評価のためにデザインされた特定の臨床試験についてFDAとの連絡を迅速に行うため、治験依頼者は、情報の目標（すなわち、対象とする適応申請）を規定する臨床開発プログラムの概要である**対象製品特性** (*target product profile: TPP*) を使用することができる<sup>4,5</sup>。

## B. PRO尺度の選択

適応申請に PRO 尺度を使用する予定の治験依頼者は、医薬品/医療機器開発の早期段階に、対象とする概念を評価・測定するために十分な既存の PRO 尺度があるかどうかを判断することが勧められる。ない場合、新規 PRO 尺度を開発することができる。場合によっては、既存の尺度を修正することで新規尺度を開発することができる。

PRO 尺度の適切性は、既存、修正済み、または新規開発にかかわらず、医薬品/医療機器の適応申請のための測定法として、特徴（本セクションを参照）、概念枠組み（セクション III. C を参照）、内容妥当性（セクション III. D を参照）、そして、その他の測定特性（セクション III. E を参照）が十分であるかどうか依存する。FDA は、PRO 尺度の開発および検証に関する文書を、臨床試験の結果と併せて審査し、適応申

<sup>4</sup> 業界および審査スタッフ向け指針「対象製品特性 - 戦略的開発プロセスツール」の草案を参照のこと。確定完成後、本指針は、当該トピックに関する FDA の現時点での見解を表す予定である。指針の最新版については、FDA の医薬品に関する指針のウェブページ

(<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>) を確認すること。

<sup>5</sup> TPP プロセスは医薬品および生物製剤の認可に使用されるが、希望する効能から始め、これらの効能を評価するための臨床試験を設計する概念は、医療機器についても同様である。



## 拘束力をもたない勧告

請が実証されていることを確認する。治験依頼者が、尺度の審査を迅速化するために FDA に対して提出すべき PRO 情報の種類については付録に記載する。

FDA が審査する PRO 尺度の特徴には以下のものが含まれる：

- 測定する概念
- **項目**の数
- 尺度の概念枠組み
- 使用される予定の医学的状態
- 使用される予定の患者集団
- データ回収方法
- 記入方法
- 回答選択肢
- **想起期間**
- スコアリング
- 項目または**領域 (ドメイン)**の重み付け
- フォーマット
- 回答者の負担
- 翻訳あるいは文化的適応の可能性

尺度の開発者においては、尺度と関連する開発履歴を公開し、アクセス可能とすることが推奨されている。開発履歴が入手できない場合、一般に、治験依頼者は、適切な集団からオープンエンド式の面接調査で得た患者の意見を含む、内容妥当性に関する文書（すなわち、尺度が、目的とする内容を測定できるという根拠）を申請書とともに提出すべきである。内容の妥当性の詳細については、セクション III. D 「内容妥当性」に記載する。さらに、セクション III. E の、「信頼性」、「その他の妥当性」と「変化検出能力」に詳述する、尺度の他の測定特性に関する経験的根拠も必要である。

検証的臨床試験の登録を開始する前に、尺度の測定特性を十分確立しておくことが勧められる。したがって、治験依頼者は、医薬品/医療機器の開発初期段階に尺度の開発と評価を開始し、検証的臨床試験の臨床試験プロトコルを確定する前に、新規またはその試験に特有の PRO 尺度について FDA と協議すべきである。

## 拘束力をもたない勧告

FDA からの情報提供要求には、当該医薬品/医療機器の審査担当部門が対応すべきである。FDA が早期段階に有益な意見を提供することができるように、治験依頼者は、適応の目標、PRO 尺度が仮定する概念枠組み、および予定する検証的試験の予備的エンドポイントモデルにおける PRO エンドポイントと臨床試験の他のエンドポイントとの関係を示すべきである。

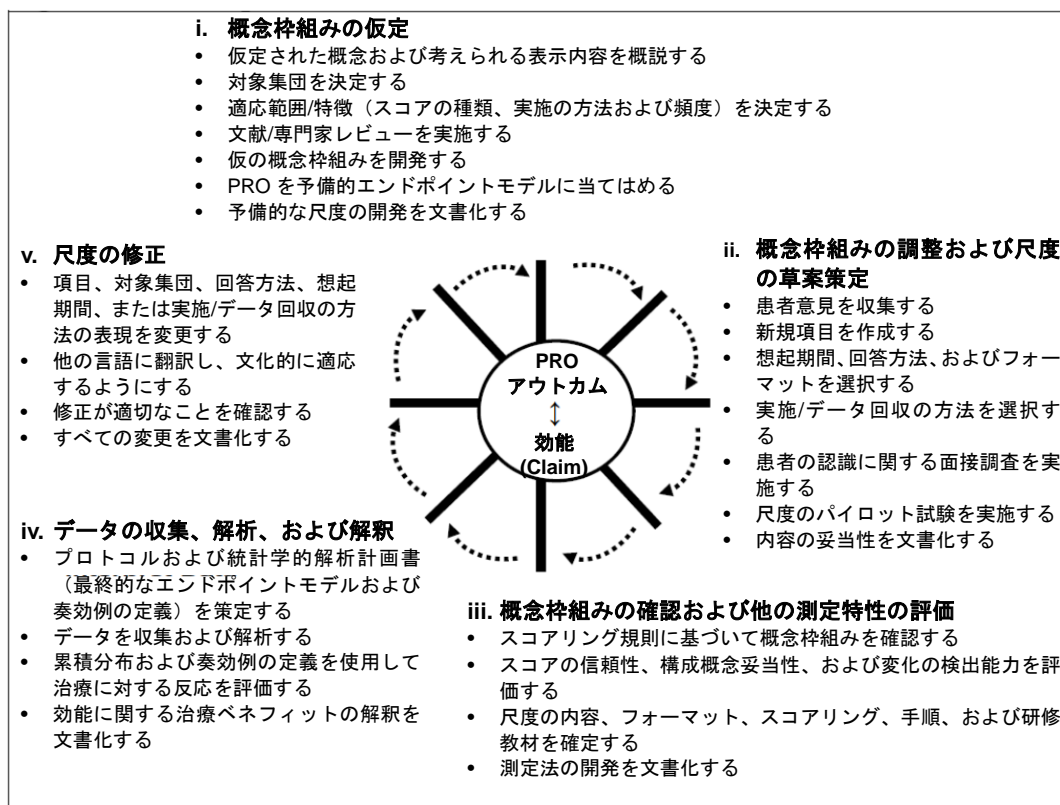
測定が複雑な、複数のドメインにわたる概念である場合、単純な概念を測定する PRO 尺度では、複雑な効能の実証には不十分な可能性がある。例えば、PRO に基づく根拠のみでは、症状の改善に関する効能しか証明できず、患者の機能あるいは心理的状態の改善に関する全般的な効能に用いることはできない。さらに、複雑な、複数のドメインについての効能は、個々の構成ドメインの概念を適切に測定しえない尺度を使用して実証することはできない。

重要な安全性に関する問題は、この問題が患者の視点から捉えることが最も適切な症状や徴候であれば、PRO 尺度を用いて測定することができる。PRO 尺度の開発における原則は、本用途においても同じである。

たとえ PRO 尺度が全般的な概念を測定するために開発されている場合であっても、全般的な概念に影響を与えるような治療の副作用か、臨床試験のデザイン時に想定していなかった治療の副作用かを区別することができないかも知れず、全般的な概念についての効能をしばしば実証できない。有害事象をもし捉えようとする場合、治療の有効性の評価とは別に、PRO 尺度は、治療の有害な結果を測定することを目指すべきである。FDA が規制する医薬品/医療機器を評価する他の臨床試験と同様、PRO 尺度で検出されたすべての有害事象は臨床試験報告書に含めるべきである。

臨床試験で用いられる PRO 尺度の開発において使用される繰返しプロセスの概要を図3に示す。開発プロセス文書のFDAによる審査については、セクション III. C「PRO 尺度の概念の枠組み」からセクション III. G「特定集団を対象とした PRO 尺度」において詳述する。

図 3. PRO 尺度の開発：繰返しプロセス



### C. PRO尺度の概念枠組み

提案された尺度を適応申請に用いることの適切性は、PRO 尺度の概念枠組みによって決定される。概念枠組みは、項目、ドメイン（下位概念）、および測定される概念同士の関係と PRO 尺度によって得られるスコアの詳細を表す図において、尺度で測定した概念として明示的に定義される。

#### 1. 測定する概念

PRO 尺度の審査において考慮する基本の 1 つに、尺度の最終的な概念枠組みを支えるための項目作成プロセスの適切性があげられる。場合によっては、治療される病態から測定すべき質問が明らかでないことがある。例えば、疼痛に対する治療の効果を評価する場合、対象集団の患者に、単一項目からなる PRO 尺度を使用して疼痛の重症度を質問する。一般に、測定すべき質問が明らかでない場合は、尺度の開発者は、最初に、文献レビューおよび専門家の意見に基づいて、測定するドメインと項目を起草し、対象とする概念の測定を支える概念枠組みを仮定することができる。その後、患者面接調査、フォーカスグループ、および質的な認識に関する面接調査により、項目に含

## 拘束力をもたない勧告

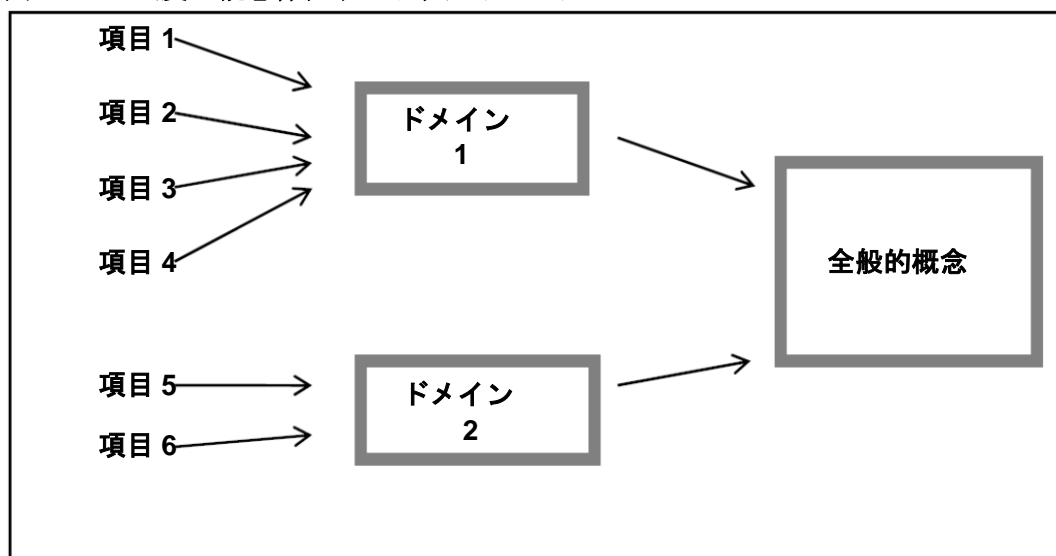
まれる概念の理解および完全性が保証される。(セクション III.D.1「項目の作成」を参照)

PRO 尺度の概念枠組みは、尺度の開発を通じて、治験依頼者が項目の分類およびスコアを支える経験的根拠を収集することで、進化し確認される。臨床試験で使用する際には、PRO 尺度の概念枠組みは、実際に項目とドメインに認められた関係に基づいて再度確認すべきである。

尺度の開発プロセスの文書では、項目とドメインの同定方法を明確にすべきである。尺度スコアから導き出される結論が妥当であることを保証するため、ドメインや合計スコアで測定される概念を正確に表現するための言葉づかいは、患者の意見を参考にすべきである。

一般的概念の測定法では、項目同士の関係、個々の項目と各ドメインの関係、ドメイン同士や各ドメインと PRO 尺度の概念枠組みに基づく一般的な概念との関係が、審査される。図 4 のダイアグラムは、PRO 尺度の概念枠組みに関する一般的な例が示されている。ドメイン 1、ドメイン 2、および一般的概念は、それぞれ関連するが個別の概念である。本ダイアグラムにおいて、項目はいくつか集まってドメインを構成する。最終的な枠組みは、測定特性の検証によって確認される。

図 4. PRO 尺度の概念枠組みを示すダイアグラム



もし単一項目が、信頼性が高く、かつ測りたい概念（例えば、疼痛強度）の測定に

## 拘束力をもたない勧告

妥当な尺度そのものである場合、PRO 尺度の概念枠組みはわかりやすい。測りたい概念が全般的なものの場合（例えば、身体的機能）、独立した単一項目ではそのドメインの全般的な概念を捉えることができないため、単一項目からなる PRO 尺度は、治療効果に関する有益な情報を提供しない。このため、複数の項目またはドメインを含む全般的概念に関する単一項目の質問群が、全般的概念に関する適応申請に十分な根拠を提供することは稀である。例えば、特定の症状群や徴候群で定義される機能障害における臨床試験では、全般的な単一項目の質問からなる PRO 尺度は、通常、適応申請のためのエンドポイントとして不十分であり、特定の症状や徴候に対する効果に関する情報を提供しない。こうした方法ではなく、適切な症状と徴候のそれぞれに対して、治療効果を適切に測定すべきである。

全般的概念における適切なドメインと項目をすべて同定することは困難であるため、全般的概念を測定するための PRO 尺度の概念枠組みは複雑なものとなる。該当する PRO 尺度が、その効能に含まれる全般的概念の重要かつ関連するドメインを測定するために開発されている場合、複数のドメインからなる PRO 尺度は全般的概念に関する適応申請に使用することができる。しかし、複数のドメインからなる PRO 尺度は複雑な性質を持っているので、誤った方向に導かないような結果の解釈と報告の方法について大きな問題となることが多い。例えば、全般的概念（例えば、特定の状態に伴う症状）のスコアの改善が1つの反応性の高い項目（例えば、疼痛強度の改善）に起因し、その他の重要な項目（例えば、その他の症状）では反応が認められない場合、その全般的概念（例えば、状態に伴う症状の改善）についての全般的な効能を証明することはできない。しかし、反応性の高い単一の項目またはドメインが、当該項目やドメインに特異的な効能を証明することができるかもしれない。

尺度の最終版については、作成項目の完全なリストの記録、項目の削除や修正理由を含む開発履歴に基づき、検証を行う。項目は表 1 に示す。PRO 尺度の最終的な概念枠組み（例えば、仮定される項目、ドメイン、および測定される概念間の関係）が適切な試験対象集団において確認され、予定される臨床試験のエンドポイントモデルと一貫性を有するかどうかは、提供される経験的な根拠に基づいて判断される。

## 拘束力をもたない勧告

表 1. PRO 尺度の開発において項目を変更する一般的な理由

項目の特性	変更または削除の理由
明瞭性または関連性	<ul style="list-style-type: none"> <li>対象集団の大部分から関連性がないと報告された</li> <li>許容できないほど大量の欠測データを生じる</li> <li>PRO 尺度の記入にあたり、患者から多くの質問や意味の明確化の依頼が寄せられる</li> <li>患者が PRO 尺度の概念枠組みと一貫性がない状況で項目を解釈し回答する</li> </ul>
回答の範囲	<ul style="list-style-type: none"> <li>床（回答スケールの最悪）または天井（回答スケールで最善）に回答する患者の割合が高い</li> <li>患者がどの回答選択肢にも該当しないと述べている</li> <li>項目に対する回答が極めて偏っている</li> </ul>
変動性	<ul style="list-style-type: none"> <li>すべての患者が同一の回答をする（すなわち、変動なし）</li> <li>大部分の患者が特定の回答選択肢のみを選択する</li> <li>既知の重要な違いが存在するにもかかわらず、患者間差が検出されない</li> </ul>
再現性	<ul style="list-style-type: none"> <li>2 時点の測定間において変動する論理的理由がない場合に、経時的に不安定なスコアがみられる</li> </ul>
項目間相関	<ul style="list-style-type: none"> <li>ある項目が同様の概念を持つ他の項目と強く相関している（重複）</li> </ul>
変化検出能力	<ul style="list-style-type: none"> <li>項目の感度が低い（すなわち、ある概念において、既に一定の変化があるとわかっているのに変化が見られない）</li> </ul>
項目の弁別	<ul style="list-style-type: none"> <li>項目が測定対象の概念と異なる概念と強く相関している</li> <li>項目が集団の既知の特徴（すなわち、重症度レベル、病態の分類、またはその他の既知の特徴）に対応する変動性を示さない</li> </ul>
重複	<ul style="list-style-type: none"> <li>同等以上の測定特性を有する他の項目で収集された情報と重複する</li> </ul>
想起期間	<ul style="list-style-type: none"> <li>集団、疾病の状態、または尺度の適用が、想起期間の適切さに影響を及ぼす可能性がある</li> </ul>

## 2. 対象集団

PRO 尺度がその対象集団に適していることを確認するため、図 3 のプロセスと表 2 の測定特性を利用して、尺度の開発プロセスで検証された患者集団と臨床試験に登録された集団を比較する。PRO 尺度に関する FDA との協議と、FDA による審査用に治験依頼者が提出すべき情報の種類については、付録を参照のこと。

小児や認知障害を有する患者、または重篤な患者に特有な留意点については、セクション III. G 「特定集団を対象とした PRO 尺度」に記載する。

## 拘束力をもたない勧告

表 2. 臨床試験で使用される PRO 尺度の審査において考慮される測定特性

測定特性	種類	評価対象	FDA による審査の留意点
信頼性	Test-retest 信頼性または面接調査担当者内信頼性（面接調査担当者が実施する PRO の場合のみ）	対象とする概念に変化が想定されない場合におけるスコアの経時的な安定性	<ul style="list-style-type: none"> <li>級内相関係数</li> <li>評価期間</li> </ul>
	内的整合性	<ul style="list-style-type: none"> <li>尺度を構成する項目が同一概念を測定する程度</li> <li>スコアに寄与する項目の相関関係</li> <li>内的整合性</li> </ul>	<ul style="list-style-type: none"> <li>概要スコアのクロンバックの<math>\alpha</math>係数</li> <li>項目-全体の相関関係</li> </ul>
	面接調査担当者間信頼性（面接調査担当者が実施する PRO の場合のみ）	PRO が複数の面接調査担当者によって実施される場合の回答間の一致	<ul style="list-style-type: none"> <li>級間相関係数</li> </ul>
妥当性	内容妥当性	意図する測定概念、集団、および使用において尺度の項目およびドメインが適切かつ十分であることを示す質的研究からの理論的根拠。他の測定特性を検証することによって、内容妥当性の問題を置き換えたり、修正できたりするものではない。	<ul style="list-style-type: none"> <li>すべての項目の導出</li> <li>質的面接調査の予定</li> <li>面接調査またはフォーカスグループの逐語記録</li> <li>逐語記録から導出された項目</li> <li>内容の開発に用いられた患者の構成</li> <li>患者の理解を評価するための認識に関する面接調査の逐語記録</li> </ul>
	構成概念妥当性	尺度の項目、ドメイン、概念の構造は、関連した概念の測定や、類似あるいは多様な患者群から得られたスコアと、理論的な関連性を有するべきであるという事前の仮説との一致性	<ul style="list-style-type: none"> <li>事前の仮説を検証する相関関係の強度（弁別妥当性および収束妥当性）</li> <li>PRO 尺度が事前に仮定された群間を区別する能力の程度（既知群妥当性）</li> </ul>
変化検出能力		PRO 尺度が測定概念に関して変化した個人または患者群（臨床試験と同様）における経時的なスコアの変化を同定できることを示す根拠	<ul style="list-style-type: none"> <li>個人内の経時的変化</li> <li>効果サイズ(effect size)の統計</li> </ul>

## D. 内容妥当性

内容妥当性とは、その尺度が測定しようとする概念を実際に測定できる程度である。内容妥当性は、測定しようとする概念や集団、使用状況において、その尺度に含まれる項目とドメインが適切かつ十分であるか否かを問題としており、質的研究から得られた根拠によって証明される。内容妥当性は、研究対象の集団、状態、および治療に特異的である。PRO 尺度においては、項目、ドメイン、および全般的スコアが、患者にとって重要な事柄を反映し、評価対象の概念に関連した患者の関心事が十分含まれていることが大切である。項目の作成過程において患者の意見を記録し文書化することは、認識に関する面接調査による患者理解の評価と同様に、内容妥当性を証明する理論的根拠になりうる。適応申請においては、使用された尺度が、対象となる概念の測定に適切であるかどうかを検証する。従って、内容妥当性は、それ以外の種類の妥当性（例えば、**構成概念妥当性**）や信頼性（例えば、スコアの一貫性）とともに重視される。尺度における種々の特性の中でも、内容妥当性の確立は優先事項である。

既存の尺度の有用性評価あるいは新規 PRO 尺度の開発の際、治験依頼者は以下に記載する内容に留意して開発プロセスおよび尺度の特性を文書化し、内容妥当性を示すことが勧められる。

### 1. 項目の作成

測定したい概念を反映し、評価に貢献できる項目を開発するためには、項目の作成過程において対象患者集団からの意見を参考にすべきである。対象患者集団からの意見は、項目の表現の作成、項目が概念範囲を完全にカバーしているかどうかの評価、明瞭性および可読性に関する初期の評価に有用である。PRO 尺度の項目は、文献レビュー、フォーカスグループからの逐語記録、あるいは、患者、臨床医、家族、研究者、その他の情報源からの面接調査により作成することができる。FDA は、PRO 尺度の開発プロセスにおいて、適切な個人や情報源が使用されているかどうか、そして収集された情報がどのように項目作成に反映されたかについて審査を行う。また、患者に対するオープンエンド式の面接調査によって、患者が問題としている概念を十分に理解できているかどうかについても審査を行う。

項目の作成では、一般に、測定対象となる患者群において幅広い情報を取り入れるべきで



## 拘束力をもたない勧告

ある。これは、想定している臨床試験のデザインに基づき、重症度ならびに年齢、性別、人種集団、および言語など、患者集団の特徴の差異を反映させるためである。

患者の意見を十分取り入れた結果の文書化なしでは、PRO 尺度の内容妥当性を証明することは困難である。FDA は、尺度の項目が患者の重要な概念の全側面を網羅しているか、および飽和に達しているかを確認するため、文書の審査を実施する。飽和とは、新たに加えるべき関連情報や重要な情報がない場合、または、追加データを収集しても、対象となる概念に関する患者の認識や、質問票の項目への理解が高まる可能性が少ないと判断される状態である。

内容妥当性を証明するために FDA に提出する文書には、理論的アプローチを含めたすべての項目作成法、研究対象集団、項目の情報源、項目の選択、編集、および削減、認識に関する面接調査の概要または逐語記録、パイロット試験、重要性評価、および項目評価のための定量的手法が記載されている必要がある。一般的な項目の変更理由を表 1 に記載する。

項目が臨床試験で対象とするすべての言語集団において作成されているわけではない場合、言語集団ごとに認識に関する面接調査を行い、内容の適切性を検討すべきである。項目の変更または削除、これらの変更理由の文書化、においては、**項目追跡マトリックス**が役立つかもしれない。

既存の尺度で、尺度の開発に患者が関与していない場合、尺度の内容妥当性を前提とすることはできない。患者の意見を明確にするオープンエンド式の面接調査またはフォーカスグループを実施すれば、新規尺度の開発で実施されると同様の新たな質的研究によって、既存の尺度の内容妥当性に関する文書を提供することができる。そのような既存の尺度の質的検証は、尺度の審査に際して、その内容に懸念が示された場合に特に重要である。例えば、臨床試験の対象集団において、対象集団に一般的に認められる症状が、当該集団の重要な症状を捉えるための測定対象に含まれていない場合、尺度の内容妥当性が疑われることになる。FDA は、内容妥当性を確立するための面接調査またはフォーカスグループの対象とすべき患者数や規模については、具体的な数字を推奨することはできない。サンプルサイズは、逐語記録の解析から得られた情報の完全性によって決まる。一般に、患者数は面接調査の質ほど重要ではなく、予定している臨床試験における対象集団の特徴の多様性を考慮したサンプル収集がなされていることが重要である。

## 拘束力をもたない勧告

患者に対して仮定の回答を求める項目では、患者が実際の状態ではなく、望ましい状態に基づいて回答する可能性があるため、臨床試験では推奨されない。例えば、日常活動を行う能力に関する概念の評価では、患者が実際には行えない作業について、行える、と回答することを避けるため、患者に対して日常活動を行えていると思うかと質問するのではなく、具体的な活動について可能かどうか（可能な場合にはどの程度大変か）を質問する方がより適切である。

複数項目の尺度を使用する場合、臨床試験に参加する患者の大部分にとって、すべての項目が関連が深いことが大切である。前パラグラフを例にとれば、臨床試験に参加する患者の大部分が行わない活動を測定項目に含めることは、大変不都合なことである。これは、測定が無効となるようなバイアスを生じさせたり、実際には治療が効果的であっても治療の効果が認められない結果を生んだりしかねない。そのような場合、否定の回答（あるいは、「ほとんど」、から、「まったく活動しない」の選択肢）は役に立たない。回答の選択肢に該当なしがある場合、スコアリングにおいて問題となる。スキップパターンは、調査実施上の問題となる可能性がある。

### 2. データ収集法と尺度記入法

治験依頼者は、データ収集法および、面接調査担当者に対する説明、自記式調査の説明、自記式調査の管理方法、を含めた尺度の記入に関するすべての手順とプロトコルについて検討しなければならない。FDA は、症例報告書や電子的な PRO 尺度のスクリーンショットとともに、データ収集法または尺度の記入方法に関するデータ品質管理手順を審査する。記入方法としては、自記式、面接調査、または両者の組合せがある。データ収集法としては、紙ベースの方法、コンピュータの助けを借りた方法、および電話を用いた評価法がある。1つの臨床試験で複数のデータ収集法または記入方法が使われる場合、方法に関わらず治療の影響が不変であることを確認するため、収集されたデータの比較性を審査する。患者日誌、または監督されないデータ入力方法が使われる場合、患者が臨床試験のプロトコルに従い記入していること、例えば、調査票を提出する来院の直前に記入がまとめて行われたりしていないことを確認するために、どのような手順が講じられているかについて、臨床試験のプロトコルを FDA は審査する。

### 3. 想起期間

治験依頼者は、PRO 尺度の想起（思い出し）期間の理論的根拠および適切性についても評価すべきである。このためには、求められる情報を適切に思い出す患者の能力について考慮することが重要である。最適な想起期間は、尺度の目的と用途、測定する概念の変動性、期間、頻度および強度、疾病や状態の特徴、および検証する治療によって選択される。PRO に基づく適応申請の評価では、尺度の想起期間が患者に確実に理解されるための手段について、FDA は臨床試験のプロトコルを審査する。多くの場合、血圧測定試験において、2 週間ごとの測定がどのように行われているかと同様、本当に知りたいことは、短期間（例えば 2 週間）の統合された影響ではなく、一定間隔（例えば、第 2、4、および 6 週目）で測定した影響である。その場合、患者にごく最近の状態を尋ねるのがいいかもしれない。また、想起の識別に起因する問題はノイズの原因となり、治療効果を曖昧にする可能性があることに注意すること。

PRO 尺度で、長期間にわたり思い出す必要がある場合や、現在の状態を過去の状態と比較する場合、あるいは一定期間の回答の平均をとる場合には、患者の記憶に依存するものであるため、内容妥当性が損なわれる可能性が高くなる。回答は、想起する時点の患者の状態に影響される可能性が高い。これらの理由から、通常望ましいのは、想起期間の短い項目や、患者に対して現在または最近の状態を尋ねる項目である。一定期間の経験を詳細に想起する必要がある場合、遡及的な報告データの妥当性および信頼性を高めるため、尺度に適切な方法や技術を使用することが推奨される（例えば、患者に対して想起期間全体を通じて最悪（または最高）の経験に基づく回答を求める、またはデータ収集に日誌を使用する）。

### 4. 回答方法（回答選択肢）

各項目の回答方法（回答選択肢）が、目的および用途に合致しているかどうかを考慮することも重要である。表 3 は、PRO 尺度で一般的に使用される項目の回答方法（回答選択肢）の一部を示している。

## 拘束力をもたない勧告

表 3. 回答方法（回答選択肢）の種類

種類	説明
ビジュアルアナログスケール（VAS）	一定の長さ（通常は 100 mm）のスケール両端にその位置の手掛かりを説明する用語があり、両端以外の位置には何も記載しない。患者には、直線上で自分が認識する状態に相当する箇所に印を付けるように指示する。印の位置がスコアとして測定される。
アンカー付またはカテゴリ型 VAS	直線上の両端以外の位置に 1 つ以上の目盛りが加えられた VAS であり、患者が尺度の両端間にある位置（半分など）を識別しやすいよう、各目盛りに参考となる用語が割り当てられている。
リッカートスケール	個別の用語または説明を順に示し、患者には自分の状態または経験を最もよく表す回答を選択するよう求める。
評定尺度	数字で表した一連のカテゴリであり、患者には自分の状態または経験を最もよく表すカテゴリを選択するよう求める。評価尺度の両端には一定の用語が付けられているが、カテゴリには言葉のラベルの代わりに番号が付与されている。
事象のリアルタイム記録	特定の事象が生じた場合に、患者日誌や他の報告システム（対話式音声自動応答装置）に含めることができるイベントログを用いて記録する。
絵入りスケール	その他の種類の回答選択肢に適用される一連の絵。絵入り尺度は小児科の質問票で多く用いられているが、認知障害を有する患者および話すことや書くことができない患者にも用いられている。
チェックリスト	チェックリストは、「はい」、「いいえ」、「わからない」など、限られた一連の選択肢から選択するという単純なものである。項目の内容が当てはまる場合は、患者が空欄にマークを入れるように求めるチェックリストもある。チェックリストの場合、すべて答えられていて重複がないか否かについて確かめる。

一般に項目の回答方法（選択肢）が適切であると見なされるのは、以下の場合である：

- 回答で使用した表現が明瞭で適切である（例えば、あるスケールに対して「正常」という手掛かりとなる言葉のラベルを付与する場合、患者が一般人における「正常」を理解していること）。
- 項目の回答方法が対象集団に対して適切である。例えば、視覚障害を有する患者にとっては VAS の記入が困難な場合がある。
- 回答方法が明確に区別されている（例えば、疼痛を表す回答として強度および重症度の両方が含まれていると、患者はこれらを区別できない可能性がある）。
- 項目の記入方法と、項目に対応する回答の選択に関して、患者への説明が十分である。
- 回答の選択肢の数が経験的に適切である（例えば、質的研究、尺度の初回検証、または既存の文献がある場合）。
- 項目の回答が、適切な順序で、同じ間隔で設定されている。
- 項目の回答で、考えられる天井効果や床効果が回避されている（例えば、悪化または改善を把握するため、最悪または最高を選択する患者の数が少なくなるように、より多くの回答選択肢を設定する必要がある）。
- 回答の方向にバイアスが生じない（例えば、軽度を表す選択肢が 2 つあり、中等度お

## 拘束力をもたない勧告

よび**重度**を表す選択肢が1つずつしかない場合、患者の回答が軽度に偏る可能性がある)。

### 5. 尺度のフォーマット、説明、研修

**PRO** 尺度の使用から得られる結果は、臨床試験中に、患者に対する説明や、面接調査担当者や、データ回収を監督する担当者への研修により左右される。治験依頼者は、記入された質問票の審査手順と、欠測データの回避または回答の明確化に使用される手順を含め、開発者が提供する文献と使用説明書に含まれる **PRO** 尺度に関するすべての説明書や手順を検討すべきである。

臨床試験で使用される **PRO** 尺度のフォーマットが、尺度の開発プロセスで使用されたフォーマットと一貫性を有することが重要である。フォーマットとは、**PRO** データの収集に使用される質問票、日誌、または面接調査用スクリプトの正確な外観を指す。フォーマットは、実施方法およびデータ収集法によって異なる。項目の順序と番号の振り方、回答選択肢の表記方法（単一またはグリッドフォーマット）、項目のグループ化、質問のスキップパターン、および面接調査担当者または患者に対するすべての説明を含め、臨床試験で使用されるフォーマットは審査の対象となる。

**PRO** 尺度の開発プロセスにおいて、開発者が提供する使用説明書に管理者負担、患者負担、欠測データ、および低品質データを最小限に抑えながら臨床試験に尺度を組み込む方法を記載することが推奨される。使用説明書では、治験担当医師および面接調査担当者を対象に、**PRO** 記入に関する重要な原則が説明されているべきである。

### 6. 患者の理解

尺度についての初回と2回目以降の草案の策定時、治験依頼者には、患者が **PRO** 尺度に含まれる項目および説明を患者が理解しているかどうかに関するパイロット試験を実施し、すべての項目および手順を検証することが推奨される。この検証では、**PRO** 尺度の概念枠組みで示されている概念が確認されていること、回答選択肢や想起期間が適切に理解されていること、尺度の可読性が対象集団にとって十分であることを文書化すべきである。これらの手順に対する **FDA** の評価では、患者との認識に関する面接調査で使用するスクリプト、

## 拘束力をもたない勧告

この面接での逐語記録、使用された可読性検査（該当する場合）、**使用可能性検査**プロセスの説明（該当する場合）、認識に関する面接調査の解析、および認識に関する面接調査またはパイロット検査の結果に基づき削除または修正された項目、回答スケール、または患者指示を含めた認識に関する面接調査報告書の審査が行われる可能性が高い。患者の認識に関する面接調査の研究で得られた根拠（すなわち、この面接の予定、逐語記録、および単一項目によって導出されたすべての概念のリスト）は、概念が適切に捉えられている時点を明らかにするために使用することができる。認識に関する面接調査を繰り返し実施することは、内容妥当性を確認する上で役立つ。

### 7. 項目とドメインのスコアリング

各項目に関しては、一般に、項目の測定に最も適した尺度（例えば、名目、序数、間隔、または比率尺度）を使用して、各回答カテゴリーに数値スコアを割り当てるべきである。FDA は回答の選択肢が適切な間隔で設定されていることを確かめるため、項目に対する回答の分布を審査する。

スコアリングアルゴリズムは、複数の項目から1つのスコアを算出する。合計スコアが適切であることを示す根拠を審査する。項目の回答に独立性がある場合、各項目に等しく重み付けしたスコアが適切である。2つの項目が相互に依存する場合、収集される情報は独立性を有する2つの項目よりも少なく、等しく重み付けした2つの項目として扱うと重み付けは過剰になる。回答選択肢の数または選択肢に割り当てられた値が、項目によって異なる場合にも、過剰な重み付けの懸念がある。ドメインのスコアを合わせて1つの全般的なスコアを作成し複雑さを加えた場合も、同じ過剰な重み付けが懸念される。治験依頼者は、項目を合計して1つのスコアを算出する場合、またはドメインのスコアを合計して1つの全般的なスコアを算出する方法を選択する場合には、質的研究または明確な統計的手法を使用して、選択した方法の正当性を説明すべきである。

項目やドメインの重み付けに、経験に基づいた患者の選好評価が使用されている場合、選好の重みの決定に使用されたサンプルの構成とプロセスを審査する。選好の重みは、資源配分（例えば、予め定められたコミュニティの重みを使用することがある費用効果分析のように）に用いるためにしばしば開発され、臨床試験において治療ベネフィットを実証するために、同一の重みを使用したくなる。しかし、予定の臨床試験集団に対する選好の重みの関連

## 拘束力をもたない勧告

性について知られていて、十分かつ適切である場合を除いて、使用することは勧められない。

複数のドメインを合わせた合計スコアは、合計スコアが複雑ではあるが単一の概念を表すことを示す根拠で裏付けるべきである。セクション III. C 「PRO 尺度の概念枠組み」において説明したとおり、尺度の最終的な概念枠組みにより、各スコアで示された概念が説明されている。スコアが、ある効能を証明することを目的に使用される場合、測定する概念は、目的とする効能の言語に適合する必要がある。一般に、ドメインまたは尺度のタイトルにより効能を表現することは推奨されない。なぜなら、多くの場合これらが測定される概念を表さないためである。

### 8. 回答者や記入者の負担

患者に過度の身体的、心理的、あるいは認知的過労を与えると、一般に PRO データの質と完答率は低下する。回答者に負担を与える原因は以下のものである：

- 質問票や面接調査の長さ
- フォーマット
- 読みづらい小さなフォント
- 項目ごとの新しい説明
- 患者に対して完全な回答を求めること
- PRO の記入時におけるプライバシー（例えば、性的能力あるいは薬物乱用歴などの私的な情報を含む質問票に記入する際にプライベートな場所が提供されない）
- 質問票に回答する時間や面接調査の時間の不足
- 対象集団に高すぎる教養が求められること
- 患者が回答したくない質問
- 面接調査担当者が特定の回答を欲したり期待したりしていると患者が感じる
- 回答に身体的な援助が必要（例えば、ページをめくること、ペンを握ること、電話やコンピュータのキーボードの使用における支援）

臨床試験において尺度の使用に当たり許容可能な回答者の負担は、プロトコルに規定される PRO 評価の頻度と時期、および、患者の認知機能、疾病の重症度、あるいは治療の毒性によって異なる。例えば、質問票で、前の質問の回答によって 1 つ以上の質問をスキップす

## 拘束力をもたない勧告

る指示があると、回答者が理解できず誤って回答したり、評価方法が難しすぎると感じたりする可能性がある。治験依頼者は、欠測データ率と回答拒否率を、回答者にかかる過度の負担、あるいは項目や回答方法が不適當である兆候と捉えるべきである。

### E. 信頼性、その他の妥当性、変化検出能力

尺度の内容妥当性が確立された後、FDA は PRO 尺度の審査において、信頼性、構成概念妥当性、および変化検出能力などの追加的な測定特性を検討する。文書化された PRO 尺度の概念枠組み、確定したスコアリングアルゴリズム、記入手順、質問票フォーマット、に特有の測定特性を、臨床試験の目的、デザイン、登録される集団、および統計解析計画書 (Statistical Analysis Plan: SAP) の視点から審査する。また、PRO 尺度の開発または検証に使用されたサンプル集団およびその病態が、予定している臨床試験と隔たりがないかについても審査する。

さらに、PRO 尺度の特定の測定特性を評価するための研究は、事前の仮説を検証するようデザインされるべきである。例えば、ある研究において、構成概念妥当性を確立するため、同一概念に関する新規 PRO 尺度と既存の尺度を、同一面接調査中に、または間隔を置かずに調査比較する場合、あらかじめ設定された相関レベルを検証するよう研究をデザインすべきであり、仮説を考慮して結果を検証すべきである。

#### 1. 信頼性

臨床試験では経時的変化を測定するので、臨床試験に使用する PRO 尺度の適切性は、信頼性や、一貫性と再現性のある真の治療効果を示す力があるか、に依存する。

FDA は、再現性 (例えば、test-retest 信頼性) が示すために行われた研究に関する文書を審査する。Test-retest で最も情報量が多くなるのは、安定した患者で、記憶の影響を最小限に抑えるのに十分な間隔をあけた検証を行った場合である。間隔をどのようにするかは、評価対象がおかれた状態や経験の変動性と、実際の変化を反映する状態や集団に生じうる経時的変化に依存するのであり、安定した患者の変動性に依存するのではない。Test-retest 信頼性は、異なる試験プロトコルにおけるさまざまな期間について検証することが可能であるとともに、同じ試験プロトコルで来院間隔が異なる場合であっても検証することが可能である。評価対象が寛解や再発といった状態



## 拘束力をもたない勧告

にある場合、または発作的な疾患の場合は、test-retest の信頼性を確立することが困難あるいは不可能なことがある。

異なる質問に対する回答間の一致性を判断するための内的整合性の信頼性検証（例えば、クロンバックの $\alpha$ 係数）は、test-retest が証明できていない場合、臨床試験の目的に合った信頼性を証明するには十分ではないことがある。しかし、他の検査においても完璧を求めることができないと同様、一般に、信頼性に欠陥がある場合、 $\beta$ （第2種）過誤を増加させる傾向があり、信頼性が低い尺度では偽陽性の結果が得られる可能性は低い。

PRO 尺度が面接調査担当者によって実施される場合、面接調査担当者間の再現性が審査される。面接調査担当者間の再現性は、尺度の使用法の標準化と、この標準に関する面接調査担当者の研修に依存する。

## 2. その他の妥当性

FDA は、内容妥当性（セクション III. D 「内容妥当性」において記載）に加え、構成概念妥当性、そして該当する場合は**基準関連妥当性**の根拠も審査する。

構成概念妥当性は、項目、ドメイン、概念間の関係が、他の尺度、患者や患者群の特徴との間に存在すべき理論的な関係に関する事前の仮説に一致することを示す根拠により決定される。

文書化された尺度を使用して収集された結果とその他の尺度を使用して収集された結果の関係が、これらの関係に関する既存の仮説（すなわち、弁別妥当性と収束妥当性）と一貫しているかどうかを確認するため、尺度の構成概念妥当性を審査する。尺度が臨床的に異なる群を区別することができることを示す根拠（すなわち、既知群妥当性）についても審査する。

前述のとおり、全般的概念に関する単一項目の質問は、効能を証明する上で役に立たない。しかし、これらは、同じ概念を測定する複数項目から構成される尺度に関する構成概念妥当性を評価する際や、全般的概念を測定する重要な項目やドメインが欠けていないことの確認に役立つ可能性がある。例えば、単一項目からなる全般的な質問を使用した結果が、同じ全般的な概念を測定する複数項目から構成される質問票を使用した結果と相関しない場合、質問票が全般的な概念のすべての重要なドメインを捉えていないことを証明しているかもしれない。

基準関連妥当性とは、PRO 尺度のスコアが、同一概念を扱う既知の**絶対的標準尺度**と関連の程度である。まれに、評価対象の PRO 概念に、基準尺度が使用される場合、尺度の基準関連妥当性を審査する（例えば、新しい睡眠尺度を睡眠ポリグラフの臨床測定と比較する場合）。しかし、PRO の多くでは、測定される概念の性質上、基準測定法が存在しないため、基準妥当性を検証することはできない。これは、患者だけが知りうる症状の尺度のすべてについて言えることである。基準尺度を使用する場合、治験依頼者は、当該基準の理論的根拠を提供すべきである。基準尺度の感度、特異度、予測値とともに、PRO 尺度が基準尺度と関連しているかどうかを FDA は審査する。

### 3. 変化検出能力

FDA は、対象とする概念に関する患者の状態の変化を示す他の同様の尺度における変化と、PRO スコアにおける変化を比較するデータを使用し、尺度が変化を検出する能力を審査する。変化検出能力の審査には、尺度が、測定概念のゲインとロス、そして臨床試験集団において想定される範囲内のすべての時点における変化に等しい感度を有することを示す根拠の審査が含まれる。

概念に関する患者経験が変化することが予測される場合、当該概念を測定する PRO 尺度の値が変化するはずである。患者の概念に関する経験が変化したことを示す明確な根拠があるが、PRO スコアが変化しない場合、変化検出能力全体が不十分であるか、PRO 尺度の妥当性を疑うべきである。PRO スコアが、対象とする概念に特有ではない変化に影響されることがわかっている場合、その PRO 尺度の妥当性が疑われる。

尺度が変化を検出する能力は、治療の有効性を評価するためのサンプルサイズに影響する。PRO 尺度の変化を検出する能力が、重要な患者サブグループ（例えば、性別、人種、年齢）によって変化する程度は、臨床試験の結果に影響を及ぼす可能性がある。重要なサブグループ間で、変化検出能力に違いがあることがわかっている場合、結果を評価する際に、これらの文献から得た差異を考慮することができる。一般に、変化を検出する能力がない場合、治療効果がない、という帰無仮説を導いてしまう傾向がある。

### F. 尺度の修正

尺度の開発と検査の適切性は、目的とする用途に特異的である。すなわち、尺度が開発された対象集団、状態、および尺度の背景に特有のものである。PRO 尺度を修正した場合、治験依頼者は一般に、新規尺度の適切性を確認するための根拠を提供するべきである。ただし、用途またはフォーマットに小さな変更を行うたびに、最終的な測定特性を文書化するための広範な研究を必ず実施する必要があることを意味するものではない。加えられた修正の種類によっては、追加的な質的研究で十分なことがある。同じ質問群に対してであっても、患者の回答が変わる可能性がある変更の例は以下のとおりである：

- 尺度のフォーマットの紙から電子媒体への変更

## 拘束力をもたない勧告

- 来院時における PRO 尺度の実施時期または手順の変更
- 異なる設定、集団、または状態に対する適用の変更
- 項目の順序、項目の表現、回答選択肢、想起期間の変更、または質問票の一部の削除
- 説明書の変更または PRO 尺度中の説明の位置の変更

異なるフォーマットの質問票間（例えば、回答尺度を垂直から水平に変更）における回答の分布の比較では、小規模な非ランダム化試験で十分な可能性がある。PRO 尺度を、有意に異なる患者集団（例えば、異なる疾患または年齢群）に使用する場合、その新たな集団における内容妥当性を確認するため、質的研究を行うことが勧められる。新たな集団における測定特性を確認するために小規模なランダム化試験を行うことは、不適切な尺度の使用リスクを最小化する可能性がある。

### G. 特定集団を対象としたPRO尺度

前述のとおり、臨床試験において複数のバージョンの尺度が使用される場合、内容妥当性と他の測定特性が互いに同等であることを示す文書を作成すべきである。小児と青年、認知障害を有する患者、または重篤な疾患のため意思疎通ができない患者、および多様な文化または言語群全体における PRO 概念の測定では、前述の問題に加えて新たな問題がある。これらについては以下に記載する。

#### 1. 小児と青年

一般に、小児用 PRO 尺度の開発プロセスに関連した審査の問題は、成人の場合と同様である。小児および青年に対する PRO 使用で、審査に追加情報が必要となる問題には、その年齢層特有の語彙、言語の理解、測定される健康に関する概念の理解、および想起期間が含まれる。比較的幅の狭い年齢層を対象として尺度を開発する場合に大切なのは、発達の違いを説明する場合に加えて、質問が理解できて全年齢層と比較可能な、信頼性と妥当性がある回答を提供することができる最低年齢を設定することである。このような集団では、**代理者報告アウトカム**測定は推奨されない（すなわち、患者ではない第三者が患者本人に代わって回答する）。自ら回答することができない患者（例えば、乳幼児患者）では、観察可能な事象または行動のみが含まれる観察者報告が推奨される。例えば、観察者は、乳児の疼痛強度を正確に報告することはできないが、疼痛に起因すると考えられる乳児の行動を報告でき

## 拘束力をもたない勧告

る。

### 2. 認知障害患者またはコミュニケーションが不可能な患者

この集団では代理者報告アウトカム測定は推奨されない。自ら回答することのできない患者（例えば、認知障害を有する患者）では、観察可能な事象または行動のみの観察者報告が推奨される。

### 3. 文化または言語によるサブグループ

開発プログラムの多くは複数の国を対象としているため、臨床試験において、PRO 尺度が複数の文化または言語に対して適用されることは多い。尺度が複数の文化または言語に合わせて並行して開発されているか、あるいは開発された尺度が新しい文化または言語に適合していたり翻訳されていたりするかどうかにかかわらず、治験依頼者には、内容妥当性および他の測定特性が、臨床試験で使用されるすべてのバージョンで同等であることを示す根拠を提供することが推奨される。臨床試験において、その対象集団に使う尺度を翻訳したり文化的に適合させたりするプロセスを FDA は審査する。

## IV. 臨床試験のデザイン

他のエンドポイントの測定に適用される臨床試験の設計原則が、PRO にも同じように適用される。したがって、本セクションでは、それらの原則の包括的な概要ではなく、主に、PRO エンドポイントに固有の問題を中心に概説する。

### A. プロトコルの一般的留意点

PRO 測定の目標が適応申請であれば、PRO 概念測定法は、具体的な臨床試験の目的または仮説として規定されるべきである。プロトコルの中の症例報告書には、使用予定である特定の PRO 尺度の正確なフォーマットとバージョンを含めることが重要である。尺度の電子版を使用する場合、プロトコルには、スクリーンショットやそれに相当する他の絵などを用いて示すことができる。新薬申請（NDA）／生物製剤認可申請（BLA）／医療機器市販前承認（PMA）、または NDA／BLA／PMA 追加申請の審査では、予定された PRO 尺度や解析

## 拘束力をもたない勧告

と、実際に使用された尺度や解析を比較する。

### 1. 盲検化とランダム化

患者と治験担当医師が割り当てられた治療を知っている、オープンラベル臨床試験においては、PRO 尺度に基づく適応申請を満足させることはまれである。実薬治療群に含まれていることを知っている患者はベネフィットを過大評価する可能性があり、実薬治療を受けていないことを知っている患者は、実際に経験した改善を過少に報告する可能性がある。同じ理由から、患者の認識に影響することを防止するため、来院時に実施する PRO 尺度は、他の臨床評価や診察の前に実施すべきである。

盲検化臨床試験において、患者は、試験全体を通じて治療の割り当てについて盲検化されるべきである。治療に、例えば有害事象などの明らかな作用がある場合、その臨床試験は意図せず非盲検化のリスクが生じるかもしれない。これらの状況では、治験依頼者は、患者に過去ではなく現在の状態を質問する方法を使用するか、あるいは同一概念に関して多くの項目を含む尺度を使用するなどの方法で、潜在的な非盲検化の影響を最小限に抑える PRO 尺度の記入法を使用することができる。

不注意による非盲検化が疑われる場合、FDA による PRO エンドポイントの評価において問題となりうる。したがって、臨床試験に PRO 尺度が含まれる場合、FDA は治験依頼者に、試験中または試験終了時に、患者がどの試験群に自分が割り当てられたと思うか、について質問する項目を含めることを推奨する。

意図的な非盲検化の影響については、臨床試験の結果の解釈において考慮することが重要である。一部の医療機器や患者に認知されやすい治療レジメンの評価など、盲検化が不可能な場合があったり、妥当な対照群がなかったりする（したがって、ランダム化されない）場合がある。そういう状況では、PRO 尺度が患者ベネフィットの評価に役立つと考えられる場合、治験依頼者には、適切な審査部門と協議することが推奨される。

### 2. 臨床試験の品質管理

臨床試験の質は、試験実施における不整合性を最小化するための手順をプロトコルで規定

## 拘束力をもたない勧告

することにより、デザインの段階で最適化することができる。FDA は、PRO と他の臨床評価を実施する順序を標準化することを推奨する。プロトコルに含めることができる標準化された説明および手順のその他の例は以下のとおりである：

- 自記式 PRO 尺度の使用に関する患者への研修と説明
- PRO 尺度を面接調査として実施する場合における面接調査担当者の研修と面接調査フォーマット
- 患者の監督、来院時または来院時外における質問票記入の時期と順序、完全に回答しているかどうかに関する調査票審査のプロセスと規則、そして、どのようにして、いつ、データが記入され、保管され、臨床試験実施施設と送受信が行われたか、などに関する治験担当医師への説明
- 臨床試験データを使用した尺度の測定特性の確認に関する計画

### 3. 欠測データの取り扱い

患者が来院しなかったり、質問票に記入しなかったり、予定される試験完了前に臨床試験から脱落することは時々ある。結果として生じた欠測データはバイアスの元となり、試験群と対照群の効果の比較性を阻害する可能性がある。なぜなら、最初にランダム化された集団のサブセットのみが寄与すると患者の群間比較性が失われるからである。欠測データは、すべての臨床試験において、成功や解釈に関わる大きな課題である。臨床試験プロトコルには、解析における欠測データの取り扱いについて規定すべきである。

プロトコルでは、患者が予定どおりに報告しなかったり、治療や臨床試験から脱落したすべての治療関連理由を収集するバックアップ計画を用意したり、試験完了前の脱落を最小限に抑える努力を行うことにより、臨床試験からの有益な情報をさらに増やすことができる。治療を中止した場合も、患者は、臨床試験の対象として PRO データを提供し続けるべきである。また、プロトコルは、治療の早期中止を回避できない場合のため、治療の中止前か直後に、PRO 測定を行うプロセスを確立すべきである。

## B. 評価の頻度

PRO 評価の頻度は、その研究で何を見たいか、尺度の回答方法が求める想起期間の長さ、

## 拘束力をもたない勧告

示された尺度の測定特性、疾病や状態の自然史、治療の性質、および予定されたデータ解析計画に対応すべきである。疾病、状態、または臨床試験デザインによっては、2回以上のベースライン評価と治療中に複数回の PRO 評価が必要となることがある。

### C. 臨床試験の期間

PRO 評価期間は、PRO に関して何を見たいかによって決まる。臨床試験の期間が、提案された効能を証明するためや、調べる疾病や状態の永続的な効果を評価するために十分であるかどうかを検討することは大事である。一般に、PRO 評価のフォローアップ期間は、有効性を測定する他の評価の場合と同じにすべきである。しかし、PRO 関連の目的に適した臨床試験の期間は、その他のエンドポイントのための期間とは異なってもよい。

### D. 複数のエンドポイントに関するデザイン上の留意点

エンドポイントモデル（セクション III.A 「エンドポイントモデル」の図1と図2を参照）に示されるエンドポイントの体系は、試験で決められている目的と、各測定が独立して、および測定間相互の関係における臨床的有意性と重要性によって決まる。事前に定められた主要および重要な副次エンドポイントの体系に含まれないエンドポイントは、すべて探索的エンドポイントと見なされる。適応申請を目的としない経済評価のためのエンドポイントは、そのことを明示すべきであり、探索的と見なされる。PRO 測定は、臨床試験の主要エンドポイント測定、その他の PRO 測定法と併せて多重主要エンドポイント測定、その他の臨床エンドポイントまたは医師が評価する測定、あるいは、体系的な順序に基づいて解析する副次エンドポイント測定となり得る。臨床試験プロトコルにおいては、臨床試験の結論が肯定的になるための条件の明確化を含め、エンドポイント測定法、結果の統計学的解析法、および解釈の基準を定義することが重要である。これは、データの非盲検化後にこれらの基準や条件を決定した場合、結果の信頼性が低くなるためである。治験依頼者は、臨床試験デザインやデータ解析に関して PRO エンドポイントを臨床試験の主要目的から分けて考えることを避けるべきである。また、治験依頼者は、提案する適応に含めるために、PRO エンドポイントの結果の良いところのみを取ったり、または事後に意図的な選択をしたりすることは避けるべきである。



## E. 奏効例の定義を使用した臨床試験結果に関する解釈の計画

臨床試験の主要エンドポイントが治療に対する個々の反応または対象全体の反応のどちらに基づくかにかかわらず、通常、多くの場合には事前の**奏効例の定義**を使用して個々の反応を表示することが役立つ（すなわち、治療ベネフィットとして解釈されるべき予め定められた期間における個々の患者のPROスコア変化）。奏効例の定義は経験的に決定され、対象集団または臨床試験の他のデザイン特性によって変わりうる。したがって、臨床試験ごとに尺度の奏効例の定義を評価する。

奏効例の定義のための経験的根拠は、アンカーに基づく方法により得られる。アンカーに基づく方法では、PRO尺度の対象概念と、アンカーで測定された概念の関係を探る。これが有用であるためには、PRO測定自体よりも容易に解釈できるアンカーを選択すべきである。例えば、失禁の不快感を評価するPRO尺度の奏効例の定義としては、失禁日誌で収集する失禁エピソードの数が用いられる。PRO尺度における奏効例として定義するためのアンカーとして、失禁エピソードの50%減少が提案されている。臨床試験の早期段階においてこのアンカーアプローチで確認すると、検証的試験において用いる奏効例の定義の基礎が得られる。

他のアンカーに基づく奏効例を定義するためのアプローチでは、異なる期間または臨床試験終了後における変化の患者評価を使用している。これらの数値評価は、*悪化から同じ*、と*改善*の範囲におよぶ。自身の状態を*同じ*、と*改善*、または*悪化*と評価した患者のPROスコアの違いは、治療に対する奏効例の定義に使用することができる。治療の割り当てについて患者が盲検化されていない場合、患者が評価した変化は、アンカーとしての有用度が低くなる。

奏効例を定義するための他のアプローチとしては、分布に基づく方法がある。例えば、尺度における意味のある変化を定義するために、患者間の標準偏差や測定の標準誤差などを使用する。分布に基づく方法では、これらの変化を大中小に分類することができ、多くの場合、奏効例の定義に対する信頼性を向上するために、アンカーに基づく推定値と組み合わせることができる。特定のスコアの変化について臨床的有意性を判定するための分布に基づく方法は、補助的なものと見なすべきであり、奏効例の定義を決定する際における唯一の基礎とすることは適切ではない。

## 拘束力をもたない勧告

それ以外に、奏効例の基準を選択することを避ける方法として、治療および対照群の回答の分布全体を示すことができる。個々の回答に意味があるかどうかは判断次第だが、この問題は、生存以外のほぼすべてのエンドポイントに存在するものである。そのような累積分布表示では、X 軸にベースラインからの変化の比率、そして Y 軸に当該変化を経験する患者の比率を示した連続プロットが描画される。奏効例のカテゴリ的定義のためには、この表示方法の方が好ましい可能性がある。様々な奏効例の定義を、反応曲線の累積分布に沿って同定することができる。

臨床試験の SAP のための解釈留意点に関する指針は、セクション V. E 「臨床試験結果の解釈」に記載されている。

### F. 電子的なPRO尺度を用いた場合に特有の留意点

PRO尺度を用いる場合、治験依頼者は、FDAの規制要件に、治験依頼者と治験担当医師による記録保持、データ管理、および閲覧が適合しているかに関して保証する必要がある<sup>6</sup>。このような責任は、臨床試験データの記録に用いる方法とは関係がなく、電子的なPROデータを含めたすべての種類のPROデータに適用される。治験依頼者は、試験を適切に実施するために必要なすべての情報を治験担当医師に提供する責任、試験をモニタリングする責任、試験が計画に従って実施されることを保証する責任、およびFDAによる試験関連の記録と報告書の閲覧、コピー、および検証を許可する責任を有する。

治験担当医師に対する記録の維持に関する主要な要求事項には、適正かつ正確な症例履歴（症例報告書および裏付けデータを含む）の作成および維持、記録の保管、FDAによる記録の閲覧、コピー、および検証に関する規定（すなわち、原データの検証）が含まれる。紙のPRO尺度を使用する場合、通常、被験者が症例履歴の一部としてオリジナルまたは承認されたコピーを保管する治験担当医師に記録を返却するため、治験担当医師は、原文書の管理、閲覧および維持に関する責任を容易に果たすことができる。しかし、電子的なPRO尺度の使用では、原データが治験担当医師ではなく、治験依頼者または開発業務受託機関によって直接管理される場合、問題を引き起こす可能性がある。治験担当医師は、FDA 査察

<sup>6</sup> 医薬品および生物製剤を開発する治験担当医師および治験依頼者に対する主要な記録の保持に関する要求事項については、21 CFR 312.50、312.58、312.62、および312.68を参照のこと。医療機器については、21 CFR 812.140および812.145を参照のこと。

## 拘束力をもたない勧告

のために電子的な元文書となる記録を管理し、閲覧を可能にすることができれば、治験担当医師の責任を果たしていると思なされる。電子的な PRO の原データの維持方法および治験担当医師が規制関連の要求事項を満たす方法については、臨床試験プロトコルまたは他の文書に規定すべきである。

さらに、FDAは、コンピュータシステムを用いた臨床データの構築、修正、管理、保管、収集、またはFDAへの提出について対処するため、また、21 CFRパート 11 の要求事項および適用を明確にするための指針をすでに提供している<sup>7,8</sup>。電子的なPROデータ（個人情報端末または電話による対話式音声自動応答装置を介して回収されたデータを含む）は、症例履歴の一部であるため、当該指針に示されるデータ標準を満たすべきである。治験依頼者は、ネットワーク攻撃およびソフトウェアの更新中に、どのようにデータの完全性を保証するのか、適切なシステムおよびセキュリティコントロール、ならびにサイバーセキュリティ計画とシステム管理計画を策定すべきである。

また、治験依頼者は以下についても回避すべきである：

- PRO データ収集機器から送信されたデータに対するすべての変更を文書化する電子的な監査証跡を使用せずに、PRO データ収集機器から治験依頼者、治験担当医師、または第三者に対して直接 PRO データを送信すること。
- 治験依頼者のみが原本を管理すること。
- 治験担当医師が、電子的な PRO データの精度を維持および確認することができないこと。治験担当医師が維持管理するデータには、患者の電子的機器からの電子的な PRO データに対する出力後のすべての変更を記録する監査証跡を含めるべきである。
- バックアップがなく、データベースが1つしか存在しないこと（すなわち、試験中におけるデータ破壊またはデータ喪失のリスクがあるが、データ復元またはデータ検証を行う方法がない）。
- 治験担当医師（および／または治験担当医師に指定された施設職員）ではなくても、誰でも原データを修正することができること。

<sup>7</sup> 業界向け指針「臨床試験において使用されるコンピュータシステム」を参照のこと  
(<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>)。

<sup>8</sup> 業界向け指針パート 11「電子記録；電子署名 - 適用範囲および適用」を参照のこと  
(<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>)。

## 拘束力をもたない勧告

- 有害事象データの紛失。
- 予定よりも早い、あるいは予定外の非盲検データの閲覧。
- 治験実施施設への立ち入り検査において FDA 査察官がデータの点検、確認、およびコピーをすることができないこと。
- 記録の変更が容易で、安全ではないシステム。
- 患者を担当する治験担当医師への適時的なデータの送信を保証することなく、重要な安全性に関する情報を治験依頼者、臨床研究機関、および／または第三者へ直接 PRO データを通信すること。

## V. データ解析

臨床試験のエンドポイント測定に PRO 尺度を用いることにより、臨床試験データの解析において新たな課題が生じる。課題の中で最も重要なものを次のセクションで記載する。

### A. 統計解析に関する一般的留意点

PROのエンドポイントを解析する際に考慮すべきことは、医薬品/医療機器の開発において用いられる他のエンドポイントと同様である<sup>9</sup>。すべてのプロトコルでは、主要なデータ解析について統計セクションに記載し、解析の詳細についてはSAPに示すべきである。臨床試験データが適応を取得するために適切であるかどうかは、事前に定められた解析手法に基づいて判断される。通常、キーオープン後に実施される計画外の、あるいは事後的な統計解析は探索的であるとみなされるため、適応申請の根拠とすることはできない。

### B. 複数のエンドポイントを用いる場合の統計解析留意点

臨床試験における PRO は、その他のエンドポイントと同様に、主要エンドポイントまたは副次エンドポイントであり得る。主要エンドポイントは、試験対象となる治療から得られる主要なベネフィットを厳密に判断するためのエンドポイントである。主要エンドポイントは、臨床試験のサンプルサイズを決定するために使用され、統計的に検証される。臨床試験におけるエンドポイントの選択には、期待される治療効果、期待されるイベント数、その他の要因が複雑に関係するため、臨床的な意味はあるとしても、必ずしも最も重要なエンドポ

<sup>9</sup> 業界向け ICH 指針「E9 臨床試験のための統計学的原則」を参照のこと  
(<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>)。

## 拘束力をもたない勧告

イントが選ばれるとは限らない。

臨床的には複数のエンドポイントに関心もたれることもある。複数のエンドポイントを解析するには、エンドポイントのどれか一つで効果があれば有効性の根拠とみなされる。そのため、第1種の過誤の発生率として知られている偽陽性率を上昇させる可能性がある。この上昇は、前向きに多重性の調整を計画することによって制御することができる。一般的には、主要エンドポイントで仮説が証明された後にのみ、副次エンドポイントを解析する。この場合は、主要でないエンドポイントを定義された順序で逐次的に解析することで、通常の有義水準( $\alpha=0.05$ )を用いて解析を行うことができる。解析は仮説の証明が失敗した時点で終了する。臨床試験プロトコルには、すべての主要エンドポイントおよび副次エンドポイントが明確に記載されていることが重要である。SAPでは、主要な解析がどのように行われるか以下のような内容が詳細に記載されるべきである。エンドポイントは連続変数(平均スコア)、二値変数(成功または失敗)、またはカテゴリ変数で解析されるかどうか、主要エンドポイントか副次エンドポイントか、全体で第1種の過誤の発生率を制御するための多重性の調整方法、および計画した統計解析手法、などである。治験依頼者は、臨床試験のSAPを、審査のためにFDAに提出すべきである。

臨床試験では、特定の有意水準(例えば、 $\alpha=0.05$ )のもとで、2つ以上の主要エンドポイントが同時に統計学的に有意と認められた場合に、臨床的に意義のある治療ベネフィットが示されたと判断する場合がある。例えば、臨床試験において2つのエンドポイントが定められており、この両方のエンドポイントで治療が対照よりも優れていると示される必要がある。そのような場合は、第1種の過誤の最大発生率( $\alpha$ )が実際には(訳注:個々の検定における有意水準よりも)低下するため、多重性の調整を必要としない。一方で、第2種の過誤は増加する。したがって、このような状況では、試験のサイズを慎重に設定することが推奨される。

多重性を調整するには、臨床試験の目的、最も重要なエンドポイント、治療のベネフィットを宣言するための条件などによって、どの調整方法を選択すべきかが異なる。そのため、一つの最善な統計的手法が存在するわけではない。より効率的な統計解析のために有益と考えられる手法として、上述のように検定順序または階層を事前に規定する方法がある。これらの方法は、従来の非階層的な方法、例えばボンフェローニ法、ステップダウンまたはステップアップ法、および **prospective alpha allocation** などよりも保守的ではない。これらの従来法

は、検定順序を事前に決めることができない場合に使用すべきである。

複数のドメインを含む PRO において、いくつかのサブドメインの結果に基づき適応申請を行う場合は、どのドメインをターゲットにするか、またどのように多重性を調整するか事前の解析計画で明確に規定していなければならない。いくつかのドメインのみを臨床試験のエンドポイントとして使用するには、当該ドメインを他のドメインから独立して用いることができるよう PRO 尺度が適切に開発・検証されていることが前提条件となる。

### C. 複合エンドポイントを用いる場合の統計解析留意点

複数のドメインからなる PRO 尺度では、個々のドメインのスコアから全般的スコアを算出することにより、複合エンドポイントを構成する。複合エンドポイントには、いくつかの利点（例えば、多重性の問題を軽減できるなど）があるが、検証的な臨床試験において治療のベネフィットを示し、適応申請を行うために用いる場合は、様々な問題や困難さの原因となる。

複合エンドポイントを解釈するための規則は、臨床試験において、その測定法が十分な使用経験を有しているか否かによる。したがって、検証的な臨床試験プロトコル作成時に複合エンドポイントを開発すべきかどうかは、以下のような考慮すべき点や経験的な根拠に応じて決まる。すなわち、構成要素が患者にとって同等に重要であること、重要性の有無にかかわらず、構成要素が同等の頻度で発生する可能性が高いこと、および、構成要素が同程度の治療効果を有する可能性が高いこと、である。したがって、構成要素間に大きなばらつきがあること予測される場合には、検証的な臨床試験において複合エンドポイントを使用しないことが推奨される。

複合エンドポイントの個々の構成要素を適応申請のために用いると、多重性の問題が生じる。一般に、構成要素が独立したエンドポイントとしてプロトコルに事前に規定され、また現行の指針<sup>10</sup>において提案されているように、すべての明記された構成要素が適応申請時に報告されない限り、複合エンドポイントの個々の構成要素を用いて適応申請を行うことは適切ではない。複合エンドポイントの構成要素は、何によって良好な結果が導かれているのかを明示するため、適応表示に示される。複合エンドポイントの構成要素の検定には、逐次的

<sup>10</sup> 業界向け指針「ヒト用処方薬および生物製剤の効能表示のための臨床試験セクション - 内容およびフォーマット」を参照のこと (<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>)。

なアプローチを使用することができる。複合エンドポイントが統計的に有意であった場合のみ、構成要素を検定する。

### D. 患者レベルの欠測データに関する統計解析留意点

欠測データが多すぎると、臨床試験の結論を導き出せないことがある。セクション IV「臨床試験のデザイン」に記載されているとおり、欠測を回避するための手法をあらかじめ臨床試験プロトコルに規定しておくことが推奨される。また、被験者が臨床試験から早期に脱落した際に、データを取得するための手順についても事前に規定しておくことが勧められる。脱落時に測定を行う場合、この情報は SAP に記載されている規則にしたがって取り扱うことができる。終末期の疾病の臨床試験では、死亡により生じる欠測データの取り扱いを事前に計画しておくことが重要である。欠測データの発生は治療または基礎疾患に起因している可能性があり、このことは群間差の解析や治療効果に関する結論にバイアスを生じさせる。

たとえ最善の計画を立てたととしても、臨床試験の終了時にデータが欠測している可能性は否定できない。SAP では、治療ベネフィットの評価を行う際にどのような統計手法を用いて欠測データを扱うか事前に計画しておくべきである。

#### 1. 同一ドメイン内の欠測項目

ある来院時の測定において、同一ドメイン内の一部の項目が欠測となる可能性がある。この種の欠測データを扱う 1 つの方法は、ドメインが適切に測定されたとみなせるためには何項目までなら欠測が許容されるか規則を定めておくことである。欠測データの取り扱いに関する規則は PRO 尺度ごとに定めるべきであり、通常、尺度開発のプロセスの中で決定される。欠測データの取り扱いに関するすべての規則は、SAP に記載されるべきである。例えば、SAP には、ドメイン全体が欠測とされないために必要な項目数を事前に決めておくことができる。

#### 2. ドメイン全体または測定全体の欠測

予定された試験完了前に被験者が早期脱落した場合にも欠測データは生じる。これらの欠測を取り扱うための様々な統計的戦略についても検討する。どのような場合にも推奨される最善の方法は存在しない。欠測データを扱うすべての戦略は不完全である。それは、強弱は

## 拘束力をもたない勧告

あってもデータが欠測する理由に何らかの仮定をおかなければならず、その仮定はデータ自体からは検証することができないためである。欠測データの補完を行う場合は、患者集団、疾患の進行、および回答者の負担を考慮すべきである。PRO のエンドポイントおよび関連するエンドポイントについてどのような補完方法を用いるかについては、プロトコルと SAP に記載すべきである。さらに、欠測データを含んだエンドポイントの統計的推定が頑健であるかを評価するために、プロトコルと SAP において感度分析を計画すべきである。治験依頼者は、補完方法が異なる 2 つ以上の感度分析をプロトコルにおいて計画することを FDA は推奨している。



## E. 臨床試験の結果の解釈

PRO においては、臨床的には意味がないかもしれない（すなわち、治療ベネフィットがない）小さな変化でもしばしば統計的な有意差が検出されることがある。そのため、単に統計的な有意差のみに基づいて適応申請を行うことは避けるべきである。

治療ベネフィットを実証するためには、治療群間における反応の累積分布関数(Cumulative Distribution Function: CDF)を検討して、治療効果の特徴づけることや被験者の部分集団で異なる反応を示している可能性を調べることは有益であると考えられる。CDF を解釈するために、治験依頼者は CDF 曲線上の各反応レベルにおける奏効例を定義することができる（セクション IV.E 「奏効例の定義を使用した臨床試験の解釈の計画」を参照）。

PRO に関するエンドポイントの解釈を行う際の留意点は、医薬品/医療機器の治療ベネフィット評価に用いられるすべての他のエンドポイントと同様である。

## 用語集

**変化検出能力 (Ability to detect change)** - 測定対象の概念が個人または集団内において変化した場合に、PRO 尺度が経時的な変化を同定できるという根拠。

**効能範囲 (Claim)** - 治療ベネフィットの申し立て。これは、FDA に認可された添付文書の任意のセクション、あるいは処方薬・医療機器の広告や販促用資材に表記することができる。(訳注：Labeling claims は「適応申請」とした。Claim や labeling が単独で出現した場合は、文脈に応じて訳し分けている)

**認識に関する面接調査 (Cognitive interviewing)** - 患者が尺度の開発者が意図するとおりに概念および項目を理解しているかどうかを判断するための質的研究のツール。認識に関する面接調査では、患者が質問をどのように解釈しているか理解を深めるため、実地での面接調査にフォローアップ質問票を組み込む。尺度の質問に回答する際に、回答者に考えを声に出してもらい、思考プロセスを説明してもらうことが多い。

**概念 (Concept)** - 測定の目標 (すなわち、PRO 尺度の測定対象)。臨床試験において、PRO 尺度は 1 つ以上の概念に対する介入効果を測定するために使用できる。PRO によって測定される概念は、健康状態や治療に関連する患者の機能や感じ方を表している。

**PRO 尺度の概念枠組み (Conceptual framework of a PRO instrument)** - 質問票や PRO 尺度の項目と測定される概念の関係を示した説明またはダイアグラム。尺度の開発を通じて、項目のグループ化やスコアを裏付けるための経験的根拠が収集されることで、PRO 尺度の概念枠組みは発展していく。最終的な概念枠組みと臨床試験の目的・デザイン・解析計画が一致しているかが審査される。

**構成概念妥当性 (Construct validity)** - 他の測定法や患者 (群) の特徴などから、あらかじめ推測される理論的な関係に関する仮説と、項目・ドメイン・概念間の関係が一致することを示す根拠。

**内容妥当性 (Content validity)** - 尺度が測ろうとする概念を測定できていることを示す質的研究から得られる根拠。目的とする測定概念・集団・使用において尺度の項目やドメインが

## 拘束力をもたない勧告

適切かつ十分であることを含む。他の測定特性を検証しても、内容妥当性の代用となったりそれで調整できたりするものではない。

**基準関連妥当性 (Criterion validity)** - 同一概念を測定する既知の標準的な PRO 尺度のスコアと関連する程度。多くの PRO では絶対的標準がないため、基準関連妥当性を測定することができない。

**領域 (ドメイン) (Domain)** - 複数の領域からなる概念を測定する尺度において、スコアによって表される下位概念。例えば、心理機能は、感情および認知機能のドメインに分割される、より大きな概念である。

**エンドポイント (Endpoint)** - 治療の効果を評価するために治療群間で統計学的に比較される測定項目。臨床試験の目的・デザイン・データ解析に対応している。例えば、ある治療が、症状 Z の強さを軽減するかどうか検証するとする。この場合、ベースラインから時間 T における、症状 Z の強さの概念を表すスコアの変化量がエンドポイントである。

**エンドポイントモデル (Endpoint model)** - PRO および非 PRO の全エンドポイント間にある関係の体系図。臨床試験の目的・デザイン・データ解析計画に対応している。

**健康関連 QOL (HRQL: Health-related quality of life)** - HRQL は複数のドメインから構成される概念であり、生活・生命の身体的・心理的・社会的側面において、疾病や治療の影響に関する患者の一般的な認識を表す。HRQL の統計的な改善、臨床的に意味のある改善、を治療ベネフィットとして申し立てるためには以下の条件を満たしていなければならない。(1) 対象とする疾患およびその治療の結果として、臨床試験の対象集団における感じ方または機能が変化した場合、その解釈に重要な全 HRQL ドメインが測定されていること、(2) 全般的な改善が明らかであること、(3) すべてのドメインにおいて悪化が一切認められていないこと。

**尺度 (Instrument)** - データを捉える手段（すなわち、質問票）だけでなく、尺度を使用するために必要なすべての情報と文書を含む。一般に、実施または回答に関する明確に定義された方法や指示、データ収集の標準フォーマット、および十分に文書化された、結果のスコアリング・解析・解釈方法を含む。

## 拘束力をもたない勧告

**項目 (Item)** - 患者が特定の概念に焦点をあてることによって評価される、個々の質問・記述・タスク（および、標準化された回答選択肢）。

**項目追跡マトリックス (Item tracking matrix)** - 尺度で使用される項目（例えば、追加、削除、修正、および変更の理由）の開発の記録。

**測定特性 (Measurement properties)** - PRO 尺度の使用に関連したすべての属性。内容妥当性、構成概念妥当性、信頼性、および変化検出能力を含む。これらの属性は、測定法の用途に固有なものであり、必ずしも尺度が使用されるすべての測定状況・目的・対象集団・設定において関連があるとは限らない。

**患者報告アウトカム (PRO: Patient-reported outcome)** - 臨床医や他の誰からも修正や解釈されない、患者の健康状態について、患者から直接得られる測定結果。PRO は、患者自記式調査または面接調査によって測定することができる。面接調査では、面接調査担当者は患者の回答のみを記録する。

**代理者報告アウトカム (Proxy-reported outcome)** - 患者以外の第三者が、患者の立場に立って報告する測定結果。代理者報告アウトカムは、PRO とは異なる。また、代理者報告は、観察者（例えば臨床医や介護者）が、観察結果を報告するだけでなく、観察結果に基づき解釈したり、意見を述べたりする観察者報告とも異なる。特に患者しか認識できない症状については、代理者報告アウトカム測定を使用することは推奨されない。

**生活・生命の質 (QOL: Quality of life)** - 生活や生命のすべての側面が、全般的な幸福感に及ぼす影響を評価した一般的な概念。この用語は、生活・生命における非健康関連側面の評価も含み、また一般的に患者の考えを意味すると認識されているため、医薬品/医療機器の効能とみなすにはあまりに一般的であり、定義しにくい概念である。

**質問票 (Questionnaire)** - 研究目的で回答を得るために、回答者に対して提示する一連の質問や項目。質問票には日誌およびイベントログが含まれる。

**想起期間 (Recall period)** - PRO 項目や質問に回答する際、その期間の健康状態等を考慮し

## 拘束力をもたない勧告

て回答することが求められる期間。想起には、現時点の想起（リアルタイム）、あるいは様々な長さのレトロスペクティブな想起がある。

**信頼性 (Reliability)** - 真の治療効果について、一貫性や再現性のある推定ができる、PRO 尺度の能力。

**奏効例の定義 (Responder definition)** - 対象となる集団において、事前に定められた期間中に患者が経験する、意味のある治療ベネフィットを表す、尺度のスコア変化量。

**飽和 (Saturation)** - 患者に面接を実施しても、新しい意味や重要な情報が見つかることなく、追加データを収集しても対象とする概念および質問票の項目を患者がどう認識しているか、理解がそれ以上に深まることがなくなった時点。

**スケール (Scale)** - 値やスコアが項目から導き出される、数値や言語アンカー（手掛かり）のシステム。例えば、VAS、リッカートスケール、評定尺度などがある。

**スコア (Score)** - 質問票の項目に対して患者の回答から得られた数値。事前に定められ、検証されたスコアリングアルゴリズムに基づいてスコアは算出され、その後に臨床試験結果における統計的解析に使用される。スコアは、個々の項目・ドメイン・概念についても算出でき、また、項目・ドメイン・概念の要約量としても計算できる。

**徴候 (Sign)** - 疾患・健康状態・治療に関連する効果の客観的根拠。徴候は通常、臨床医によって観察や解釈がなされるが、患者が発見したり報告したりすることもある。

**症状 (Symptom)** - 疾患・健康状態・治療に関連する効果の主観的根拠。患者だけが気づき、知ることができる。

**対象製品特性 (TPP: Target product profile)** - 特定の種類の根拠（例えば、臨床試験や他の原データ）を対象とする適応申請や概念と結びつけ、適応を取得するための臨床開発プログラムの概要。

**治療ベネフィット (Treatment benefit)** - 治療が患者の生存・感じ方・機能に及ぼす効果。

## 拘束力をもたない勧告

治療ベネフィットは、有効性または安全性上の利点によって示される。例えば、ある症状の場合は、改善や発症までの期間の延長として、治療関連毒性の場合は、その軽減または発症までの期間の延長として、治療効果は測定できる。患者の生存・感じ方・機能に対する治療効果を、直接的にとらえないものは、治療ベネフィットの代理測定である。

**使用可能性検査 (Usability testing)** - 回答者が尺度を使用するだけでなく、指示を理解し、記憶し、正確に遵守する能力について記載された文書に関する、正式な評価。

## 付録：FDAが審査する、PRO尺度に関する情報

以下のトピックは、審査のためにFDAに提出するPRO文書に記載すべき範囲を示す。各セクションで提供すべき背景情報の量は、使用するPRO尺度によって異なる。特定のPRO尺度においては一部のセクションが他のセクションよりも重要でない可能性や、医薬品/医療機器開発の初期段階においては、不十分な内容しか記載できない可能性もある。以下の各領域において必要とされる根拠の種類の詳細については、本指針の内容を参照すること。

PRO情報が電子的に提供される場合は、eCTDのセクション5.3.5.3に記載すべきである<sup>11</sup>。

### I. 尺度（計画された尺度のコピーが提出されない場合は審査を開始することができない）：

- A. 審査対象となる臨床試験（プロトコル）において計画または使用するものと同一バージョンの尺度とすべての使用説明書。必要に応じて、スクリーンショットまたは面接調査担当者用スクリプトを含める。
- B. 旧バージョンの尺度（該当する場合）
- C. 使用説明書：尺度の使用説明書は、付録Aとして提出する。提出された尺度の使用説明書を参照できる。
  - 1. 実施時期、方法（例えば、質問紙か、電子的か）、および調査様式（例えば、自記式か、医師による調査か、または面接調査担当者が実施するか）
  - 2. スコアリングアルゴリズム
  - 3. 質問票を実施するための研修方法および資材
    - a. 患者研修概要を記載し、すべての資材のコピーを付録A1に含める
    - b. 治験担当医師の研修概要を記載し、すべての資材のコピーを付録A2に含める
    - c. その他の研修概要を記載し、すべての資材のコピーを付録A3に含める

### II. 対象となる適応申請または対象製品の特性（TPP）<sup>12</sup>

<sup>11</sup> ICH 業界向け指針「M2 eCTD：eCTD仕様」を参照のこと：

(<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>)。

<sup>12</sup> 業界および審査スタッフ向け指針「対象製品特性 - 戦略的開発プロセスツール」の草案を参照のこと。確定後、本指針は、当該トピックに関するFDAの現時点での見解を表すことになる。指針の最新版については、FDAの医薬品に関

## 拘束力をもたない勧告

PRO であるかどうかによらず、どのエンドポイントについても、対象となる適応申請をすべて明確に記述すべきである：

- 疾患または状態および病期、重症度、あるいは分類（該当する場合）
- 対象集団（例えば、年齢、性別、他の人口統計学的情報）
- データ解析計画

### III. エンドポイントモデル

- A. PRO であるか否かを問わず、すべてのエンドポイント間の関係（既知のものおよび仮説的なもの）。これらのエンドポイントには、PROに加え、生理学的／臨床検査／身体的なもの、介護者または臨床医によって報告されるものを含めることができる。
- B. PRO であるか否かを問わず、予定されたデータ解析に対応する、適応申請のためのエンドポイントの体系。

### IV. PRO 尺度の概念枠組み

項目とドメイン、および各ドメインと合計スコアの関係を示す仮定された（提案された）、あるいは最終的な PRO 尺度の概念枠組みのダイアグラム。PRO 尺度の概念枠組みが、臨床試験プロトコルに記載されかつ適応申請を計画している臨床試験のエンドポイントに対応することを保証しなければならない。

### V. 内容妥当性の文書化

尺度が臨床的に最も重要なすべての概念および項目を捉えていること、および項目が完全で、関連性があり、かつ、患者にとって理解可能なことを示す根拠が示されなければならない。この根拠は、既存の尺度および新規に作成された尺度の両方に適用され、また、予定される臨床試験の集団および適応に固有のものである。文書化には以下が含まれる：

- A. 文献レビューおよび専門家の意見の文書化

---

する指針のウェブページ (<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>) を確認すること。



## 拘束力をもたない勧告

- B. 質的研究のプロトコル、面接調査ガイド、および以下の結果の概要：
1. フォーカスグループテスト（記録を付録 C1 に含める）
  2. オープンエンド式の患者面接調査（記録を付録 C2 に含める）
  3. 認識に関する面接調査（逐語記録を付録 C3 に含める）
- C. 項目の起源と導出。どのようにして項目が作成、修正、および確定されたかイベントの時系列提示。
- 患者において検証されたバージョンの項目追跡マトリックス。飽和している根拠によりそのまま保持された項目や削除された項目を示す。概要を記載し、すべての資料を付録 B に含める。
- D. 以下の内容妥当性を裏付ける質的研究の概要：
1. 項目の内容
  2. 回答方法（選択肢）
  3. 想起期間
  4. スコアリング
- E. 項目プールをどのように作成し、減らし、確定するかの方法を示した質的研究の概要。試験の種類（すなわち、フォーカスグループ、患者面接、または認識に関する面接調査）および対象集団の特徴を明記する。完全な記録およびデータセットを付録 C に含める。

## VI. その他の測定特性の評価

対象集団および適応において内容妥当性が確立された場合、尺度の信頼性が高く、有効であり、かつ、変化を検出することができることを示す根拠。測定特性の評価には、臨床試験で使用する尺度と同一バージョンを使用すべきである。

- A. 尺度テストのプロトコル
- B. 適応申請のために用いられる各ドメインまたは要約スコアのテスト結果の概要：
1. 信頼性（内的整合性；test-retest）
  2. 構成概念妥当性（収束妥当性、弁別妥当性、既知集団妥当性）
  3. 変化検出能力

VII. スコアの解釈

- A. 臨床試験結果の臨床的意義を解釈する際に使用する論理および方法の概要
- B. 奏効例の定義（すなわち、臨床試験集団に特有な意味のある患者内変化の定義）

VIII. 翻訳および文化的適応

- A. 試験の対象集団のために、尺度を翻訳し、文化的に適応させる際に使用したプロセス
- B. 患者テスト、言語または文化特有の懸念、および修正版の作成理由の説明。
- C. 翻訳ないし修正版のコピー
- D. オリジナルおよび修正された尺度の内容妥当性、および他の測定特性が同等であることを示す根拠

IX. データ収集法

- A. 臨床試験において使用する予定のデータ収集法（例えば、電子、紙）を開発するために用いられたプロセス  
**PRO** エンドポイントの評価に電子的なデータ収集法を使用する場合、電子的な原文書の維持、送信、および保管の手順が規制関連の要求事項に適合することを示す根拠
- B. 内容妥当性および他の測定特性がすべてのデータ収集法と同等であることを示す根拠
- C. 追加的なデータ収集法の使用説明書

X. 修正

オリジナルの尺度の変更（例えば、項目の表現、回答選択肢、想起期間、新規集団または新規適応における使用）

- A. 尺度を修正するための理論的根拠およびプロセス
- B. オリジナルおよび新規尺度のコピー
- C. オリジナルの尺度および修正された尺度の内容妥当性および他の測定特性が同等であることを示す根拠（新規適応および新規集団における使用を含む）。

XI. 臨床試験デザインおよびデータ解析に関連した PRO に特有な計画

## 拘束力をもたない勧告

A. 臨床試験プロトコル。プロトコルの中で以下について保証しなければならない：

- 各 PRO エンドポイントが明確な試験目的として規定され、多重性に関する懸念に対処できている
- 適切に盲検化された臨床試験である
- 以下の研修手順が詳述されている：
  - 患者
  - 面接担当者
  - 治験担当医師
- 尺度の記入方法が尺度の使用説明書と一致している
- PRO 尺度のスコアリングの計画が、尺度の開発中に使用されたものと一貫性を有する
- 臨床試験からの脱落前または脱落直後の PRO エンドポイント評価が手順に含まれている
- PRO 評価の頻度および調査時点が、患者集団、試験デザインおよび目的、および実証された PRO の測定特性の点から適切である
- 臨床試験の期間が PRO の目的を裏付けるために十分である
- 欠測データの取り扱いに関する計画が含まれている
- 治療群間で累積分布関数を比較するための計画が含まれている
- 電子的に収集する PRO を含め、データの収集、保管、および取り扱い、そして、送信手順が規定されている。

B. 解析計画書 (SAP)。SAP に以下が含まれていることを保証する：

- 多重性調整の計画
- 尺度および患者レベルにおける欠測データの取り扱いに関する計画
- グループ間差の表現方法の説明（例えば、累積分布関数）

## XII. 重要な参考文献

関連するすべての公表および未公表文献を記載および添付する

付録 A - 使用説明書

A1：患者研修

## 拘束力をもたない勧告

A2：治験担当医師研修

A3：その他の研修

付録 B - 項目追跡マトリックス

付録 C - 記録

C1：フォーカスグループ

C2：オープンエンド式の患者面接調査

C3：認識に関する面接調査