

患者報告結局在药物临床研究中
应用的指导原则
(试行)

患者報告アウトカムの薬物臨床研究における
使用に関するガイドライン
(試行版)

2021年12月

目 录

一、序言.....	1
二、患者報告アウトカムの定義.....	1
三、患者報告アウトカム測定用尺度の開発、翻訳、改善.....	1
(一) 患者報告アウトカム測定用尺度の開発.....	2
(二) 患者報告アウトカム測定用尺度の翻訳及び（または）異文化間調整.....	4
(三) 患者報告アウトカム測定用尺度の改善.....	4
四、患者報告アウトカム測定用尺度の選択と評価.....	4
五、臨床研究における患者報告アウトカム使用の考慮事項.....	5
(一) 推定目標フレームワーク.....	5
(二) 患者報告アウトカムを臨床研究のエンドポイントとする場合.....	5
(三) 研究プランと研究報告書における関連尺度の説明.....	6
(四) 質問票における有効回答.....	6
(五) 欠落データ.....	6
(六) 多重性問題.....	6
(七) アウトカムの解釈.....	7
(八) PROまたはePROの品質制御.....	7
(九) リアルワールド研究におけるPROまたはePRO使用.....	7
六、電子化患者報告アウトカム.....	8
(一) ePRO測定.....	8
(二) ePRO使用に関する一般的な考慮事項.....	8
七、監督管理機関との意思疎通.....	9
別添1：用語集.....	12
別添2：用語対照表（日本語と英語）.....	14

患者報告アウトカムの薬物臨床研究における使用に関するガイドライン（試行版）

一、序言

臨床転帰は薬物治療の利益とリスクを評価する核心的な根拠であり、いかに正確に、完全に、信ぴょう性が高い方法で臨床転帰をモニタリングするかも極めて重要である。患者報告アウトカム（patient-reported outcome, PRO）は臨床転帰の一つの形で、薬物登録関連臨床研究でますます広く使用されている。また、患者中心型薬物開発（patient-focused drug development, PFDD）の理念と実践が持続的に発展するのに伴い、薬物の全ライフサイクルで患者の体験、見解、ニーズなどに関するデータを獲得し、それを有効的に薬物の研究、開発と評価で活用することも日増しに重視されている。臨床転帰評価（clinical outcome assessments, COA）、特にそのうちの患者報告アウトカムは患者の体験を反映でき、患者を中心とする薬物開発の重要な部分である。

本ガイドラインの主旨は、PROの定義と薬物登録関連臨床研究における適用範囲、PRO測定、特に尺度開発と使用の一般的原則、PROデータ収集の品質制御、データ分析と解釈で注意すべき事項、及び規制当局との意思疎通などを明らかにし、申請者（治験依頼者）が薬物登録関連臨床研究においてPROデータを合理的に使用するための指導的な意見を提供することにある。

本ガイドラインは、PROをエンドポイントとし、医薬品登録を支えるための臨床研究（臨床試験及びリアルワールド・データを活用する研究を含む）に適用される。

二、患者報告アウトカムの定義

患者報告アウトカムとは、患者から直接報告され、他人に修正または解釈されていない患者自身の疾患とその治療体験に関する評価の結果である。

PROは患者が自らアウトカムを報告することを強調するが、患者が自分で評価する能力を持たず、または失った場合、保護者または保護者が指定した代表者がPROを行う必要があるため、代理人による偏差を十分に評価しなければならない。

尺度は、PRO測定で比較的に多く使用されるツールで、主に痛み、生活の質などの主観的測定に使用されるが、既存尺度は、特別症状（吐き気など）や症候群などの主観的測定関連問題をすべて解決できるわけではない。PROデータ収集には紙版記録と電子版記録という2つの手段がある。電子化手段を用いてPRO記録を行うことは電子化患者報告アウトカム（electronic patient-reported outcome, ePRO）と呼ばれる。

三、患者報告アウトカム測定用尺度の開発、翻訳、改善

臨床研究において、PROで測定用尺度を使用すると決められた場合、もし当該研究に相応しい測定用尺度がなければ、研究目的に基づいて測定用尺度を開発する必要がある。もし当該研究に相応しいと公認された中国語測定用尺度があれば、著作権を取得した上で使用することができる。もし当該研究に相応しいと公認された外国語測定用尺度があれば、開発をへて中国語版を作成して使用することになる。既存測定用尺度が当該研究に十分に適合しない場合、改善した上で使用することになる。どのように成熟度の高い既存尺度から、実施予定の研究トに適合する尺度を選択するかについては、その科学性と操作可能性を考慮しなければならない。

(一) 患者報告アウトカム測定用尺度の開発

PRO測定用尺度の開発は患者の視点を反映すべきで、当該尺度の臨床的価値に重点を置く必要がある（治療効果評価の目標、臨床的意義の説明可能性と治療方法決定を指導できるかどうかを含む）。尺度開発過程は図2に示したとおりである。尺度開発は通常、有効性評価に用いられるが、重要な安全性イベントに関する開発も可能で、その原理と過程が同様である。

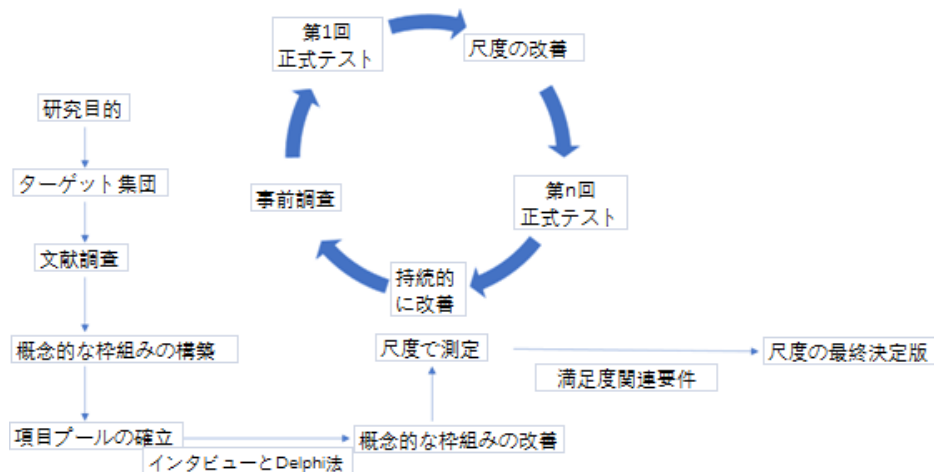


図2 尺度開発プロセス説明図

1、概念的な枠組みの構築

尺度の構造は一重、二重と三重があり、臨床研究では一重と二重構造がよく使われる。一重構造尺度には、単一項目（視覚シミュレーションで痛みを示す項目など）と複数項目（簡素化した口腔乾燥度尺度）の尺度がある。以下、二次構造尺度を例にして説明する。

二重構造尺度の第一層は領域で、第二層は詳細項目である。尺度の概念的枠組みは一般的に研究開発者による文献調査、専門家の知識と経験、患者インタビュー及びその他必要な調査研究に基づいてつくられる。領域の数と命名は研究内容への理解によって決められ、各領域の下の項目の数と内容はその所属する領域の内包と重要度を表す（例えば、各項目が同様な重要性を持つ場合、当該領域の項目数はその重要性を示す）。

2、項目プールの作成

尺度の基礎は項目であり、項目で具体的に設問できるが、領域は概念的なものにすぎない。項目設計のために、できるだけ内容豊富な項目プールを構築する必要がある。項目の出所はすべての活用可能なルートである。文献、患者または専門家インタビュー、関連領域の尺度開発プラットフォーム、研究開発報告書、開発者による設計などを含む。

項目設計は尺度開発の核心的内容の一つである。項目プールが十分に豊富で成熟している場合、ほとんどの項目は項目プールから取得できるが、開発者によって設計される項目もある。設問はなるべく閉鎖的な質問の形をとり、あいまいな言葉、両義的または誘導的な質問、二重否定的な述べ方、ネガティブな述べ方と患者が答えたくない質問を避けるべきである。同時に、応答の天井や床効果、1つの項目で同時に2つ以上の質問をすることなどを避けるべきである。また読解しやすいために、なるべくよく使われる言葉を使用し、学歴の低い患者も読めるように心掛けて項目を設定する（例えば、小学校卒業で読めるように）。

3、項目の表示方法

項目の表示方法は二分類法、ランク分け法（Likertなど）、連続法（視覚シミュレーションなど）、図形などである。そのうち、5レベルのLikert尺度が最もよく使われる。具体的にはいくつのレベルのLikert尺度を使用するかは、測定性能が最も優れていることを基準に決められる。

4、インタビュー

研究開発者が尺度の概念的枠組みをつくった後、まず患者インタビュー、専門家インタビュー及び（または）専門家調査を行い、専門家の意見とアドバイスに基づいて概念的枠組みを調整する必要がある。患者インタビューは患者報告アウトカム（PRO）の尺度の有効性をより保障し、患者のニーズと意見を表すのに役立つ。専門家調査の主な目的は、尺度の概念的枠組みの合理性、項目表現の正確さ、応答可能性と領域、項目の権限付与を検証することである。領域と項目の権限付与は、尺度開発で最も重要な部分である。専門家調査は通常、一回だけ終わらず、専門家の意見（特に項目の権限付与に関する意見）がほぼ一致するまで何回も行われる。

5、事前調査と正式調査

専門家の意見をまとめて最初の概念的な枠組みを改善した後、尺度の初回テスト版を作成し、ターゲット集団でテストを行い、テストの結果に基づいて改善し、正式なテスト版を仕上げる。正式なテスト版を使用してターゲット集団で調査を実施する際に、そのサンプル量は事前調査のパラメーターによって推定する。正式なテスト版の改善もテストの結果によって調整する。また、テストの実施回数は尺度性能への満足度によって決められる。

6、概念的枠組みの検証

事前調査と正式調査は、概念的な枠組みを検証するプロセスである。概念的な枠組みの適用性に関する評価は主にその尺度性能（信頼性と妥当性を含む）に基づいて行われる。

（1）信頼性：信頼性とは、類似する条件で得られた測定結果の一致性を意味し、測定器具の信頼性を評価するために使用される。PRO尺度でよく使われる信頼性指標は再測定信頼性、内部一致性信頼性と被験者内部信頼性である。再測定信頼性は尺度の再現可能性を評価するために使用され、初測定と再測定の相関係数が低すぎないように注意しなければならない。内部一致性信頼性は尺度内部の統合性を評価するために用いられ、よくCronbach's係数で評価を行う（通常は0.7以上）。被験者内部信頼性は通常、被験者グループ内相関係数（intraclass correlation coefficient, ICC）で評価される。ある文献によると、ICC一致性は<0.4が悪い、0.4-0.75がまずまず、>0.75が良いと分けることができる。

（2）妥当性：妥当性とは、測定したい内容がどの程度測定できたかということで、測定器具の有効性を評価するためのものである。良い尺度は信頼でき、効果的なものでなければならない。信頼性が高いことは、妥当性も高いということではない（例えば、重度うつ病症状尺度は、重度うつ病の測定で、比較的の高い信頼性と妥当性を有するが、躁病の測定で、信頼性が高いが、妥当性が低いかもしれない）。但し、信頼性が低ければ、妥当性も必ず高くない。

尺度の妥当性を評価する方法は多く存在するが、3C法がよく使われ、すなわち内容妥当性（content validity）、基準妥当性（criterion validity）と構成概念妥当性（construct validity）である。内容妥当性は主に専門家の知識と経験及び患者の主観的な判断基準に基づく領域と項目の内容が合理的であるかどうか、測定したいことを正確に測定しているかどうかを示すものである。基準妥当性は、開発された尺度といわゆる「ゴールド・スタンダード」との相関度を示すものである。ゴールド・スタンダードは通常存在せず、存在すれば、尺度開発の意味がなくなるため、実際に使用されることが少ない。構成概念妥当性は、よく探索的及び検証的因子分析によって、データ生成の構造と概念的な枠組みとの整合性を評価することである。

上記の3C概念に加えて、妥当性のもう一つの重要な指標は、変化を検出する能力であり、反応度とも呼ばれ、すなわち、患者転帰の変化（介入前後の変化、異なる介入に対する反応など）を敏感に示す能力である。

7、尺度説明書の作成

尺度の正しい使用を確保するために、尺度の使用説明書を作成する必要がある。尺度説明書には、ターゲット集団、ガイダンスを含む完全な尺度構造の説明、領域と項目のスコアリング規則、尺度性能、有効回答に関する規定、欠落データの処理、想起期間（関連する場合のみ）などが含まれるが、それらに限られるわけではない。

（二）患者報告アウトカム測定用尺度の翻訳及び（または）異文化間調整

臨床研究でPRO測定に用いられる原尺度が外国語である場合、通常は中国語に翻訳してから使用することになる。原尺度のある項目またはいくつかの項目が文化の違いで患者に理解されず、もしくは効果的に協力してもらえなかった場合、異文化間調整の問題にもかかわる。尺度の翻訳と（または）異文化間調整が適切に行われたかどうかを判断するには、翻訳と（または）異文化間調整後の尺度と原尺度の性能が近いかどうかを基準としなければならない。尺度の翻訳と（または）異文化間調整は、以下の手順で行うことになる。

- 1、準備段階。尺度開発に関するすべての資料を調べ、学際的な翻訳チーム（例えば、中国語を英語へ翻訳する専門家、英語を中国語へ翻訳する専門家、医学などの専門家を含むチーム）を設立し、尺度開発者と意思疎通を行うルートを確立し、尺度の最新バージョンを使用する許可を取得するほか、コミュニケーションを通じて尺度の意味をよりよく理解し、翻訳をより正確にする。
- 2、外国語から中国語への翻訳。2人以上の翻訳者が別々に尺度の原文を中国語に翻訳し、それぞれの翻訳原稿を統合して中国語訳の初稿を仕上げる。
- 3、チェックするための翻訳。母国語が原文の外国語で、中国語も上手な翻訳者が中国語訳の初稿を原文の外国語に翻訳する。その訳文と原文を比較し、違いが大きい場合、中国語訳と原文の違いが許容できる程度になるまで、中国語訳を修正し、初版中国語訳を完成する。
- 4、尺度の異文化間調整。尺度の一部の項目が現地の文化に合わない場合、それらの項目を調整する必要がある。調整の結果について満足できるかどうかを判断するには、調整後の尺度性能が元の尺度に近いことを原則としなければならない。
- 5、中国語版初版テスト。ターゲット集団で中国語版を使用して患者インタビューを行い、尺度項目の理解可能性及び患者の認知度などを評価し、尺度性能に関する定量テストを行った結果、尺度の性能が原文の尺度と近い場合、中国語版を確定することができる。違いが大きい場合、尺度の性能が要求水準に達するまで、中国語版を修正し、中国語訳最終版を仕上げる。
- 6、中国語版研究開発報告書。中国語訳最終版が完成された後、研究開発報告書を作成し、研究開発の全過程を記録し、尺度性能を報告し、測定用尺度の説明書を作成し、必要に応じて中国語版ソフトウェアの著作権を申請する。

（三）患者報告アウトカム測定用尺度の改善

既存尺度が研究プロジェクトに完全に適合していない場合、改善した上で使用しなければならない。例えば、初期臨床試験（第Ⅱ相など）のデータ分析によって、使用された尺度は、研究に必要な信頼性及び（または）妥当性を持たず、尺度を改良し、もしくは新しい尺度を開発する必要があるとわかった場合、第Ⅲ相試験で使用される尺度が十分な信頼性と妥当性を持つよう確保するために、第Ⅲ相試験実施前に、尺度を再度テストしなければならない。

四、患者報告アウトカム測定用尺度の選択と評価

測定用尺度はPRO測定器具として良好な性能を持ち、信頼性も有効性持つべきである。実施予定研究プロジェクトに適用するPRO測定用尺度を正しく選択することは非常に重要であり、科学性と操作可能性を同時に考慮し、以下のポイントに注意するよう推奨する。

1、尺度の適用性：尺度の構造を考察し、その全体的概念が尺度開発の目的を満たすかどうかと適用集団に相応しいかどうか注目し、研究対象となる患者集団は原尺度の適用集団と一致しなければならない。

2、書類またはシステムの標準化：標準化した尺度関連書類とシステムがあるかどうか注目すべきである。書類とシステムとは、説明書類（特に尺度におけるスコア計算の説明）、ユーザー向けマニュアル、データ収集の標準フォーマット、重要な参照データ（設計時に使用するサンプル量の推定）などを含むが、それらに限るわけではない。

3、開発プロセス：尺度の使用目的が明確化されているかどうか、開発プロセスが厳格に規範化されているかどうか、尺度の構造（領域と項目及びその権限付与）が合理的であるかどうか、発表されたアウトカムが詳しいかどうか注目する。

4、権威性：研究開発の成果が同業界の定期刊行物で発表されているかどうか、広く引用され、応用されているかどうか、ガイドブックに推薦されているかどうか注目する。

5、言語と文化：尺度の有効性検証で、教育、文化、人種の差異を考慮しているかどうか、新しい外国語版が正確に翻訳され、検証されたかどうか注目する。翻訳及び（または）異文化間調整された尺度の性能は、原尺度に近いものでなければならない。

6、検証：十分なサンプル量で検証を行ったかどうか、項目の設計と計算方法が合理的であるかどうか、十分な信頼性と妥当性があるかどうか注目する。

7、実行可能性：尺度使用の実行可能性（実行プロセスの操作可能性、複数の尺度を使用する際の重複問題などが含まれるが、それらに限られるわけではない）。患者の回答時心理的負担が大きければ、回答漏れや回答拒否の現象が増加し、PROデータの質も低下する可能性がある。患者の負担を増やす要素は尺度の内容が多すぎる、尺度には重複する部分が多いこと、同時に複数の尺度を使用し、その中の1つまたはいくつかの尺度は意味が大きいこと、質問票がうまく設計されず、読みにくいこと、質問項目が回答しにくいプライバシーに関連すること、項目設計が合理的ではないことなどである。

五、臨床研究における患者報告アウトカム使用の考慮事項

（一）推定目標フレームワーク

ICH - E9 (R1) で提案された推定目標フレームワーク構築のガイドラインと方法は、PROをエンドポイントとする臨床研究にも適用される。推定目標フレームワークは、全体プランと統計分析計画で明確に定義する必要がある。

（二）患者報告アウトカムを臨床研究のエンドポイントとする場合

臨床研究で患者報告アウトカムを主要または副次的エンドポイントとして選択する場合、選択の理由と根拠を説明し、研究目的、適応症の機序、薬物作用機序及び臨床研究プランなども総合的に考慮すべきである。PROを主要または副次的エンドポイントとすることに関しては、以下の問題に注意しなければならない。1) 十分な根拠が必要であり、かつ研究目的と一致しなければならない。2) 研究設計で患者への盲検が設けられない場合、主観的評価で比較的に大きな偏差が発生するリスクもあるため、極めて慎重に設計を行わなければならない。3) 十分に長い観察期間を設け、臨床的意義のあるPRO変化を具現化すべきである。4) 全体でI類の過誤率をコントロールしなければならない。5) サンプル量の確定は予想の差異を十分に考慮し、臨床的意義を持たせるべきである。

選定された患者報告アウトカムは、患者の薬物作用に関する体験を反映しなければならない。薬物作用は有効性だけでなく、安全性、耐性または生活の質への影響などにもかわる。そこで、合理的に患者報告アウトカムを選択することは研究で患者の体験をよりよく把握し、薬物の研究と開発が患者中心とする理念に基づくことに役立つ。

(三) 研究プランと研究報告書における関連尺度の説明

質問票を使用して測定した患者報告アウトカムを主要または副次的エンドポイントとする際に、研究プランでそれに関する説明を行わなければならない。説明は尺度選択と使用の合理性、必要に応じて尺度開発と使用状況を簡単に紹介すること（特に使用率が比較的に高くない尺度について説明する）、尺度性能の評価方法と指標、尺度関連データの収集と品質制御、尺度関連データの分析方法、尺度使用に関する詳細説明と研修計画などを含まなければならないが、それらに限るわけではない。

臨床研究報告書は尺度関連データの収集状況（有効回答、欠落データなど）、尺度性能（信頼性、妥当性など）、原尺度と比較して差異が大きい場合における具体的な原因分析と研究結論への潜在的な影響の評価、尺度関連データの詳細分析結果及びその合理的な解釈を含まなければならないが、それらに限るわけではない。

(四) 質問票における有効回答

患者は質問票を記入する時に、回答漏れ、消極的な回答（例えば、Likert項目で5種類の選択肢があっても1つの決まった選択肢にしかチェックしない）などの現象が現れる可能性があり、それによってデータで実際の状況を反映できないこともある。そこで、質問票を使用する場合、いずれも有効回答の基準を設定し、尺度説明書で説明しなければならない。例えば、ある尺度に関する規定で、有効回答率が15%以上とされる。

（尺度によって定義が違う項目で回答せず、またはすべての項目で1つのレベル（「非常に満足している」など）にだけチェックした場合、当該研究対象の回答が無効と見なされる。研究プラン及び（または）統計分析計画で、有効回答の判断基準及びその理由を詳細に説明しなければならない。最終的に無効回答と判断されれば、無効回答と同様に欠落データとして扱う。場合によっては、質問票全体が有効回答であるかどうかを考慮することに加えて、尺度におけるある領域が重要な変数と見なされる可能性があり、その領域に関する回答が有効であるかどうかも事前に規定するかもしれない。

(五) 欠落データ

PROデータ、特に質問票で測定されたデータは、よく欠落する。そのため、研究の実施過程で品質制御に注力し、なるべく欠損を減らす必要がある。多領域を使用する質問票における項目別データの欠損に関しては、通常は補填という方法をとる。具体的には、原尺度説明書における方法を優先的に使用し、次に文献や記事で紹介される一般的な方法を使用し、また次に現在の研究データに関する探索的分析によって方法を確定する（通常は探索的研究で確定する）。補填を行わない場合、欠落が多すぎて無効回答とみなされる。また、原尺度の規定に従って欠落データの処理を行い、もしくは事前に研究プランで定めた規則に従い、項目単位でデータが欠落した場合にいかに関問票全体と各領域のスコアを計算するかを決める。試験設計段階で、欠落データに関する合理的な統計分析戦略を決定しなければならない。

(六) 多重性問題

PROが主要エンドポイントの一つまたは副次的エンドポイントに選定される場合、多重性問題がしばしば発生する。その一般的な対処原則は「薬物臨床試験多重性問題ガイドライン（試行版）」を参照するように。治験依頼者は臨床研究プランと統計分析計画で多重性問題の解決方法と調整方法を事前に定めなければならない。PROで使用する尺度は通常、複数の領域を含む。そのうちの1つまたはいくつかの領域が重要な臨床的意義を持ち、研究プランで重要な副次的指標（治験依頼者は当該特定メリットを説明書で主張

する予定)として指定されている場合、多重性問題にもかかわり、研究設計時に全体でのI類過誤制御を考慮しなければならない。質問票には多領域と多項目が含まれる特性から、質問票全体のスコア計算と分析に重点を置くことに加えて、各領域と項目の分析も必要である。広い意味から言えば、多重性問題に関連するが、それらが主要または副次的エンドポイントに選定されず、または尺度説明書で特定メリットとして主張されていない限り、多重性関連調整を行う必要はない。

(七) アウトカムの解釈

尺度に基づくPRO解釈は、治療利益を評価するために使用される他のエンドポイントと同様であり、ポジティブアウトカムは臨床的意義と統計的意義の両方を持つ必要がある。

臨床的に有意な最小差 (minimum clinical important difference, MCID) は通常、臨床的意義の閾値を定義するために用いられる (例えば、10点上限痛み視覚シミュレーションという尺度を使用して痛みの度合いを測定する時、介入前後の平均点数がどれだけ下がれば、臨床的意義があるか、またはベースラインより点数がいずれも下がった2グループの差がどれだけあれば、臨床的意義があるか)。MCIDを確定する時に、関連ガイドライン、専門家のコンセンサスなどの公認された基準を第一に選択して使用しなければならない。公認された基準がない場合、監督管理機関と速やかに意思疎通を行い、基準を選定しなければならない。統計方法は基準の選定に一定の根拠を提供する可能性がある。

統計でMCIDを推定するには、分布に基づく方法とアンカリングリングに基づく方法がよく用いられる。そのうち、アンカリング法はより信頼性が高く、試験間比較も行いやすいため、患者の臨床的意義に対する体験によって外部の全体的な指標 (改善なし、やや改善、著しい改善など) を設定し、関連尺度のスコア変化を確認できる。通常、全体的な指標 (レベル変数) と関連尺度スコアの変化量との相関係数は最小限に0.3以上になれば、有意義とされる。ある研究の結果によると、相関係数が0.3の場合は低度相関で、相関係数が0.5の場合は高度相関である。MCIDの推定で、ほかの統計方法を使用することもできる。例えば、混合線形モデルに基づく方法などがある。規制当局と意思疎通を行った上で、主に使用する方法を決定するように。

(八) PROまたはePROの品質制御

研究実施過程で、研究センター、患者、観察者データ収集の一致性を保障し、臨床研究のクオリティを高めなければならない。研究プランでは、少なくとも以下のことを明確に定めなければならないが、それらに限るわけではない。

- 品質管理の標準的な操作規程の確立。
- PROまたはePROデータ収集の時点と実施順序。
- 関係者を対象にPRO/ePRO測定器具を使用するための研修と指導 (尺度の完全性を判断する方法と基準、データ記入、保存、伝送の時間と方式などを含む) を実施し、尺度使用の目的、尺度説明書の具体的な内容及び尺度データ収集過程における品質制御を十分に理解させること。
- PROまたはePROのデータ管理計画。

なお、PROまたはePROを使用する臨床研究では、PROまたはePROデータ収集の完全性と正確性を保障するために、より持続的に、積極的にオンサイト監査を行う必要がある。

(九) リアルワールド研究におけるPROまたはePRO使用

リアルワールド研究におけるPROまたはePRO使用は、展望性のある研究 (展望性のある観察と研究や実効的な臨床試験などを含む) で多く行われる。収集されたPROまたはePROデータの管理または処理の具体的な方法については、「リアルワールド・エビデンス生成のた

めのリアルワールド・データに関するガイドライン（試行版）」を参照するように。

六、電子化患者報告アウトカム

（一）ePRO測定

紙版PROと比較して、ePROはデータ収集の効率、リアルタイム性、柔軟性、コンプライアンス遵守、安全性、患者プライバシー保護などの面で明らかにアドバンテージを持っている。ePROの不足点は一部の患者が電子機器の操作に困る可能性があることである（特に高齢者、児童、及び病気で体が不自由になり、機器操作に困る患者など）。

現在、ePROデータの収集は電話回線を使用する対話型音声応答システムとスクリーンを使用する対面報告システムという2つのツールで行われる。電話回線を使用する対話型音声応答システムは、自動呼び出し機能が特徴で、事前に録音された質問と回答用選択肢を用意し、患者はキーボードを使用して回答することができ、データは中央データベースに直接保存される。スクリーンを使用する対面報告システムは、スマートフォン、タブレット、コンピュータ、さらにはウェアラブル医療機器など、患者自身保有設備にインストールでき、患者は関連ウェブサイトにアクセスし、関連ソフトウェアをダウンロードしてセットアップし、自分の状況によって答えを選択し、記録を保存することができる。

ePROシステムは電子カルテシステムまたは電子データ収集システムとドッキングすることができ、個人レベルの完全なデータフローを生成させる。その時間記録機能は効果的に重複回答または早すぎる回答など、データの信頼性に影響する行為を防止し、識別することができる。その遠隔モニタリング機能は研究者、データ管理者がリアルタイムでオンラインデータ管理と遠隔データモニタリングを行い、疑問のあるデータにマークをつけ、速やかに被験者を再度訪問するのに役立つ。

（二）ePRO使用に関する一般的な考慮事項

薬物登録を目的とする臨床研究で、ePRO測定器具及びデータ収集と管理などは、薬物臨床試験データ管理、電子データ収集、リアルワールド・データ管理に関するガイドラインの基本要求事項に従うべきである。

オンラインプラットフォームに基づくePRO測定は紙版PRO測定器具と異なり、そのデータは通常オンラインデータ収集センターにアップロードされ、ユーザーはまとめて管理し、データの蓄積、監査と導出を実施する。そのため、研究者が電子版原データのメンテナンスと保存の権限を持つよう確保するために、研究機関は申請者による検査と監督管理機関による監査に原始ファイルを提供し、サポートすべきである。ePRO測定器具の使用にあたって、以下の原則に従わなければならない。

1、研究者は、ePRO原データの正確性、真実性を保障し、確認する権限を持たなければならない。研究者は監査軌跡を通じてePROデータが測定設備でアップロードされた後のあらゆるデータの変化と修正を把握し、申請者や第三者機関が独断的にePRO原データの収集と管理システムを制御することを防止する。ePRO原データとは、ePROシステムで最初に記録し、データベースに保存されたデータである。ePROシステムが最初の記録でEDCシステムを直接使用し、eCRFにデータを保存した場合、最初のeCRFが原データになる。

2、データセキュリティ管理システムとアクセス制御メカニズム。暗号化技術を使用してデータの収集、抽出、伝送と保存過程における完全性、機密性、追跡可能性を保障し、あらゆる個人または機関による原データ修正を防止し、患者が報告した有害事象、ハイリスク警報などのデータを削除し、関連アクセス制御メカニズムを確立し、計画外の盲検解除によるリスクを回避する。

3、データのバックアップ。テスト中にデータが破損し、または失われ、原データを再構築または検証できなくなるリスクを回避する。

4、データ保存。研究機関と研究者は、規制当局の検査スタッフが臨床研究の現場検査

で原データを確認、複製できるように、電子版原データまたは電子ドキュメントを保存しなければならない。

研究データの分析により、ePRO尺度の測定性能が原尺度と大きな差があると発覚された場合、ePRO尺度使用における潜在的な問題を考慮し、尺度を修正しなければならない。また、項目反応理論に基づくePRO測定器具は、コンピュータ適応テスト技術で、前項項目の答えによって後項項目を選択することによって、項目数を減らして患者の回答時心理的負担を軽減できるが、尺度の妥当性を確保する前提で項目数を削減するよう注意しなければならない。治験依頼者はこのようなePRO測定器具を使用し、概念的な枠組みの構築、項目プールの設計プロセス、プログラム設定規則及びアウトカム分析などに関する資料を提出しなければならない。

七、監督管理機関との意思疎通

治験依頼者がPROまたはePROを実証研究の主要または副次的エンドポイントとし、計画を立てる場合、監督管理機関と速やかに意思疎通を行う必要がある。意思疎通で議論する事項は、適応症とされる疾患の背景、PROを研究の主要または副次的エンドポイントに選定する理由と根拠、研究設計のタイプ、新規開発尺度（あれば）の概念的な枠組み及び説明書などの資料、PROまたはePRO改善及び（または）異文化間調整（あれば）とその根拠、信頼性と妥当性の検証、臨床的に有意な最小差及びその根拠、実施面における品質制御などであるが、それらの事項に限るわけではない。意思疎通を行う前に、治験依頼者は監督管理機関にPROまたはePRO統計分析を含む試験プランとPROまたはePROに関する資料を提供しなければならない。試験実施過程で、PROまたはePROの変更によって臨床試験プランが大幅に調整された場合、監督管理機関と速やかに意思疎通を行わなければならない。

参考文献

- [1] Acquadro C, Berzon R, Dubois D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient- Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health*. 2003; 6(5): 522-531.
- [2] Bukhari M. “PROMs vs. PREMs (Patient-Reported Experience Measures).”; *Patient Reported Outcome Measures in Rheumatic Diseases*. Ed. Miedany YE. London: Springer, 2016; 405-417
- [3] Byrom B, Watson C, Doll H, et al. Selection of and Evidentiary Considerations for Wearable Devices and Their Measurements for Use in Regulatory Decision Making: Recommendations from the ePRO Consortium. *Value Health*. 2018; 21(6): 631-639.
- [4] Calvert M, Blazeby J, Altman DG, et al, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013; 27; 309(8): 814-822.
- [5] Cohen J. A power primer. *Psychological Bulletin* 1992;112(1): 155–159
- [6] Coons SJ. ePRO systems validation: clearly defining the roles of clinical trial teams and ePRO system providers. *Value Health*. 2013; 16(4): 457-458.
- [7] Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health*. 2009; 12(4): 419-29.
- [8] Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J Off J North Am Spine Soc*. 2007; 7: 541–546.
- [9] Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the label claim. *Health Qual Life Outcomes*. 2010; 8: 89.

- [10] EMA. Reflection paper on the regulatory guidance for the use of health relate quality of life (HRQL) measures in the evaluation of medicinal products. 2005.
- [11] EMA. Reflection paper on expectations for electronic source data. 2010.
- [12] EMA. Reflection paper on the use of patient reported outcome (PRO) measures in oncology studies. 2014.
- [13] Fayers PM, Machin D. Quality of Life: The assessment, analysis and reporting of patient-reported outcomes (3rd Edit). John Wiley & Sons, Ltd. 2016.
- [14] Ferreira ML, Herbert RD, Ferreira PH, et al. A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. *J Clin Epidemiol*. 2012; 65: 253–261.
- [15] FDA. Clinical outcome assessment (COA) compendium. 2021.
- [16] FDA. Clinical outcome assessment (COA) qualification program.
- [17] <https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-qualification-program>.
- [18] FDA. Guidance for industry: Assessing COVID-19-Related Symptoms in Outpatient Adult and Adolescent Subjects in Clinical Trials of Drugs and Biological Products for COVID- 19 Prevention or Treatment. 2020
- [19] FDA. Guidance for industry: Electronic source data in clinical investigations. 2013.
- [20] FDA. Guidance for industry: Patient-Reported Outcome Measures: use in medical product development to support labeling claims. 2009.
- [21] FDA. Patient-Focused Drug Development: collecting comprehensive and representative input. 2020.
- [22] FDA. Roadmap to patient-focused outcome measurement in clinical trials. 2015. <https://www.fda.gov/media/87004>
- [23] /download
- [24] FDA. Plan for issuance of patient-focused drug development guidance. 2017.
- [25] FDA. Upper facial lines: developing botulinum toxin drug products. 2014.
- [26] Fiero MH, Pe M, Weinstock C, et al. Demystifying the estimand framework: a case study using patient-reported outcomes in oncology. *Lancet Oncol* 2020; 21: e488–94
- [27] Fleiss JL. Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*, 1975; 31: 651-659.
- [28] Fox MW, Onofrio BM, Onofrio BM, et al. Clinical outcomes
- [29] and radiological instability following decompressive lumbar laminectomy for degenerative spinal stenosis: a comparison of patients undergoing concomitant arthrodesis versus decompression alone. *J Neurosurg*. 1996; 85(5):793-802.
- [30] Hong K, Majercak KR, Villalonga-Olives E, et al. Patient- reported outcomes in breast cancer FDA drug labels and review documents. *J Patient Rep Outcomes*. 2021; 5(1):36.
- [31] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–415.
- [32] Lawrance R, Degtyarev E, Griffiths P, et al. What is an estimand& how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials? *J Patient Rep Outcomes*. 2020; 4(1):68.
- [33] Ly JJ, Crescioni M, Eremenco S, et al. Training on the use of technology to collect patient reported outcome data electronically in clinical trials: best practice recommendations from the ePRO Consortium. *Ther Innov Regul Sci*. 2019; 53(4): 431-440.
- [34] Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement Instruments. *BMC Med Res Methodol*. 2006; 6: 2.
- [35] Walters S. Quality of life outcomes in clinical trials and health-care evaluation: A practical

guide to analysis and interpretation. John Wiley & Sons, Ltd. 2009.

- [36] Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient- reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value Health*. 2005; 8(2): 94-104.
- [37] Guideline, ICH."Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials." E9 (R1). Step 4 (2019): 20.

別添1：用語集

基準妥当性 (Criterion Validity)：実証的妥当性とも呼ばれ、「ゴールド・スタンダード」と、現実を使う測定値との一致性 (相関) の度合いによって評価する妥当性である。ほとんどのPRO尺度は、「ゴールド・スタンダード」が存在しないため、その基準妥当性を測定できない。

測定器具 (Instrument)：データを取得し、その使用をサポートするすべての情報と文書を含むツール (質問票など)。通常は、実施に関する詳細なマニュアル、データ収集用標準的フォーマット、スコアリングと分析方法、及びターゲット患者群転帰説明用テンプレートなどが含まれる。

概念 (Concept)：関心のある概念 (concept of interest, COI) とも呼ばれる。監督管理では、概念はPRO尺度で把握または反映した個人の臨床、生物、生理、機能などの面での状態または体験である。PROでは、概念は患者の健康状況または治療に関連する機能または感覚を表す。

患者報告アウトカム (Patient-reported Outcome, PRO)：患者から直接報告され、他人に修正または解釈されていない自身疾患と関連治療体験に対する評価である。

患者中心型薬物開発 (Patient-focused Drug Development, PFDD)：一つの系統となる方法であり、当該方法は薬物の全ライフサイクルで、患者の体験、観点、ニーズと優先順位が把握され、薬物の開発と評価で活用されることを確保するのに役立つ。

想起期間 (Recall Period)：質問票に回答する際に、被験者が自身の体験などについて思い出す期間のこと。思い出は瞬間的 (リアルタイム) でも、少し前に遡ってもかまわない。想起期間は長すぎではないため、一般的には1週間以内とする。

変化発覚能力 (Ability to Detect Change)：測定器具でPRO測定を行う際に、条件 (介入前後の時間、患者グループ間差異など) の変化によって、測定結果も変化するため、それらの変化を発覚する能力が必要である。

構成概念妥当性 (Construct Validity)：構築妥当性とも呼ばれ、測定しようとする概念や特性をどれだけ適切に反映しているかを意味する。

Cronbach's α 係数：尺度内の一致性を評価するための信頼性指標。

尺度の概念的な枠組み (Conceptual Framework of a Scale)：今までの研究 (文献)、専門家の知識と経験、及び必要な調査に基づいて開発された尺度における領域と項目の枠組み。領域の数と命名は研究内容に対する理解によって設定され、各領域の下の項目数と項目内容はその所属する領域の内包と重要度を反映する (例えば、各項目が同様な重要性を持つ場合、項目数は領域の重要性を表す)。

内容妥当性 (Content Validity)：専門家の知識に基づく質的研究で、尺度が測定したい内容を測定できるかどうかを検証すること。

生活の質 (Quality of Life, QoL)：生命の質とも呼ばれ、生活各部分での体験を表し、全体的な健康状況を評価するために使用される。

項目 (Item)：患者が特定概念に対する評価を行う際に、使用される質問、陳述またはタスク (及び標準化された回答用選択肢)。

調整 (Adaptation)：人種間の言語と文化の違いを考慮して尺度を変更すること。調整はPRO尺度の構造を変更しない。一部の内容を調整することによって、尺度を別の国々の言語、患者、パターンに適用させるだけである。調整に関する研究は、PRO尺度が新しい環境または新しい言語で測定の機能を果たせるかどうかを検証するためのものである。

領域 (Domains/Dimensions/Factors)：尺度の第一層 (二重構造の場合) または第一と第二層 (三重構造の場合) を構成し、尺度のある内容 (概念) を表わす。1つの領域は1つまたは複数の項目で構成される。

妥当性 (Validity)：測定したいことがどの程度反映されているかを判断するための指標で、PRO尺度の有効性を評価するために用いられる。

信頼性 (Reliability)：類似の条件下で測定したアウトカムの一致性と再現性を指し、

PRO尺度の信頼性を評価するために使用される。

症状 (Symptom) : 患者のみが感知、体験できる疾患、健康の状況または治療効果で、主観的証拠である。

治療利益 (Treatment Benefit) : 治療が患者の生存、感覚または機能に与える影響。治療利益は有効性または安全性によって証明することができる。例えば、症状の改善または緩和、または治療による毒性の減少などによって治療利益を測定することができる。治療が患者の生存、感覚または機能に影響を与えて効果をあげたことを直接測定できないのは、治療利益の代替測定である。

臨床的に有意な最小差 (Minimum Clinical Important Difference, MCID) : 通常は臨床意義を定義するために用いられる閾値をさす。例えば、10点上限痛み視覚シミュレーション (VAS) を用いて痛みの度合いを測定する時、介入前後の平均点数がどれだけ下がれば、臨床的意義があるか、またはベースラインより点数がいずれも下がった2グループの差がどれだけあれば、臨床的意義があるか。

別添2：用語対照表（日本語と英語）

日本語	英語
基準妥当性	Criterion Validity
測定器具	Instrument
被験者内部信頼性	Intra-rater Reliability
再測定信頼性	Test-retest Reliability
電子化患者報告アウトカム	Electronic Patient-reported Outcome (ePRO)
電子データ収集	Electronic Data Capture (EDC)
尺度特性	Measurement Properties
概念的な枠組み	Conceptual Framework
患者報告アウトカム	Patient-reported Outcome (PRO)
患者中心型薬物開発	Patient-focused Drug Development (PFDD)
想起期間	Recall Period
尺度	Scale
簡素化した口腔乾燥度尺度	Summated Xerostomia Inventory, (SXI)
対話型音声応答システム	Interactive Voice Response Systems (IVRS)
スクリーンを使用する対面報告システム	Screen-based Reporting Devices
アンカリングリングに基づく方法	Anchor-based Method
変化発覚能力	Ability to Detect Change
構成概念妥当性	Construct Validity
内部一致性信頼性	Internal Consistency Reliability
内容妥当性	Content Validity
生活の質	Quality of Life
視覚シミュレーション尺度	Visual Analog Scale (VAS)

日本語	英語
項目	Items
項目プール	Item Pool
調整	Adaptation
領域	Domains/Dimensions/Factors
項目反応理論	Item Response Theory (IRT)
妥当性	Validity
信頼性	Reliability
回答時心理的負担	Respondent Burden
グループ内相関係数	Intraclass Correlation Coefficient (ICC)
症状	Symptom
重度うつ病症状尺度	Symptoms of Major Depressive Disorder Scale (SMDDS)
治療利益	Treatment Benefit
患者自身保有設備	Bring-Your-Own-Device (BYOD)
臨床的に有意な最小差	Minimum Clinical Important Difference (MCID)