



治験における **Patient Reported Outcomes**

～臨床開発担当者のための **PRO** 利用の手引き～

日本製薬工業協会

データサイエンス部会

タスクフォース 7

Ver 1.0

2016年6月（2025年4月再掲）

目次

はじめに.....	1
1 PRO とは.....	2
1.1 PRO の定義.....	2
1.2 臨床試験における PRO 評価の意義.....	3
1.3 各規制当局の動き.....	5
2 FDA の PRO ガイダンスの概要.....	7
2.1 PRO 尺度の概念的枠組み (CONCEPTUAL FRAMEWORK).....	7
2.2 内容妥当性 (CONTENT VALIDITY).....	8
2.3 新規 PRO 尺度の開発と妥当性評価.....	9
3 臨床試験計画のための PRO 尺度準備プロセス.....	11
3.1 社内体制.....	11
3.2 臨床パッケージとして PRO の必要性の検討.....	12
3.3 既存の PRO 尺度を選択する.....	12
3.4 新規 PRO 尺度の作成.....	13
3.5 言語バリデーション.....	20
4 PRO 尺度を用いたデータ収集の留意点.....	23
4.1 施設担当者のトレーニング.....	23
4.2 被験者のトレーニング.....	24
4.3 治験依頼者側でのデータレビュー.....	25
4.4 統計解析留意点.....	26
おわりに.....	28
付録.....	29
1) PRO 使用経験.....	29
2) 社内体制.....	32
3) 国際共同試験における PRO.....	34
4) 既存 PRO の選定.....	36
5) 新規 PRO の開発.....	36
6) 既存 PRO 尺度の言語バリデーション.....	38
7) トレーニング.....	38
8) 実施上の工夫.....	40
引用文献.....	43
資料作成者.....	46

はじめに

従来、治療効果の評価方法として、生化学的検査、生理学的検査、画像検査及び生存率などの医学的指標が重視されてきた。これらの指標は、被験者自身の主観的な要素が入りにくく科学性を担保することが容易であることから、世界共通で使用されてきた。

一方、適切な医学的指標がなく、被験者の主観的指標のみが評価の対象となる場合もある。すなわち、従来の医学的評価を補完する指標として、場合によっては主要評価項目として、治療に対する被験者の主観的評価が測定される場合があり、それは科学的に測定することが求められる。この被験者の主観的な指標は **Patient Reported Outcome (PRO)**：患者報告アウトカム)と総称されている。具体的には、治療満足度、症状や機能の程度、**HR-QOL (Health related quality of life)**：健康関連の生活の質)、治療の遵守度などがあり、主観的指標という特性を考慮した、信頼性と妥当性を担保できる適切な方法によって測定することが重要である。そのためには科学的に妥当な指標の開発に加え、施設担当者や被験者への実施方法のトレーニング・実施上の留意点の理解も不可欠である。

2015年度データサイエンス部会タスクフォース7で2015年8月にアンケートを実施したところ、7割を超える企業でPROの使用経験があることが確認され、そのアンケートにおいて、データの正確性・信頼性に対する漠然とした不安、国際共同治験での言語や文化の違いに起因するデータの信頼性に対する心配、施設担当者や被験者へのトレーニング不足、PRO尺度(PRO measure)の妥当性評価に関する社内での経験・知識不足、PROの結果の取り扱いに関する誤解、ePRO (electronic PRO) システム関連のコストおよび技術的な問題など、様々な課題が挙げられた。一方で、施設担当者や被験者へのトレーニングに関する様々な工夫を凝らしている会社や、新たにPRO尺度を作成しその妥当性評価の結果を公表している会社もあることもわかった。

本報告書は、治験でPROを用いたい、特に承認申請目的(効能効果など)で使用したいと考える臨床開発担当者を対象に、PROおよびPRO尺度使用の実際についての理解を目的としてまとめたものである。1章では、PROの定義やその評価の意義、各規制当局の動きについてまとめ、2章ではFDAのPROガイダンスの概略を紹介した。3章ではPRO尺度を治験で利用するまでに必要な準備、特に既存PRO尺度の選択や言語バリデーションの手順、既存PRO尺度がない場合の新規PRO尺度作成の手順についてまとめた。4章では実際にPRO尺度を使用する際に必要な施設担当者や被験者へのトレーニングやその資料、実施環境の留意点、統計解析の留意点を記載した。なお、医療経済評価など、承認申請以外の目的で用いるPROに特化した話題は他の成書を参考にされたい[1, 2]。また、付録として上記アンケート結果も収録した。本報告書が、PROの適正な利用の一助となれば幸いである。

1 PROとは

医薬品の開発においては、臨床試験によってその有効性、安全性を検証することが必要である。近年では被験者自身により直接新薬の有効性・安全性さらには有用性を評価するという考え方で、PROが重要視されている。

PROはQOL（Quality of Life：生活の質）や自覚症状（痛み、口渇、倦怠感、頭痛、うつ状態、不安感、睡眠障害等）、機能状態（身体機能（歩行等）、社会機能、認知機能、生活の制限等）、全般的な健康状態や幸福度、治療に対する満足度、治療遵守などを測定することができる。

QOLとは？

QOLとは、日常生活、身体的、精神的、社会的な機能と、病気とその治療の影響についての、被験者自身による幸福度の報告である[3]。QOLのうち、通常医薬品開発を目的とした臨床試験では、健康関連QOL（Health Related QOL, HR-QOL）を考える。健康関連QOLの定義は、1948年に世界保健機構（WHO）によってなされた健康についての定義「単に疾病がないということではなく、完全に身体的・精神的及び社会的に満足のいく状態にあること」に基づいている[4]。

1.1 PROの定義

PROは、「被験者の症状やQOLに関して、自分自身で判定し、その結果に医者をはじめ他のものが一切介在しないという評価方法」（後述のFDAガイダンス[5]の定義）である。PRO尺度は質問票や被験者日誌のような症状や機能を測定するツールを指す。PROという概念が最初に提案されたのは、2001年2月のPRO調和グループの会議[6]である。ここでの定義は、「臨床医などによる被験者の回答の修正、または解釈を介さない被験者の健康状態に関する被験者から直接得られた報告に基づく測定」であった。この頃、疾患及び治療に対する被験者の認識を理解することへの関心が非常に高まっていた。

また、被験者の印象を単に尋ねるだけでPROと称して臨床評価に用いられるケースも見られるが、信頼できる結果を得るためには、評価項目、質問方法、結果の重みづけなどが検証される必要がある。このプロセスはPROのバリデーションと呼ばれる。詳細は3章に示す

PRO尺度を大きく分類すると、一般的な健康状態を包括的に評価する包括的尺度と、特定の疾患やそれに伴う特定の症状の程度を評価する疾患・症状特異的尺度に分類することができる。表1-1に包括的尺度と疾患・症状特異的尺度の例を示す。

表 1-1 包括的尺度と疾患・症状特異的尺度の例

包括的尺度	SF-36, EQ-5D	
疾患・症状特異的尺度	疼痛	VAS, MPQ, PainDETECT
	関節炎, 腰痛	WOMAC, ODI, JOABPEQ
	がん医療・緩和ケア	BPI, EORTC-QLQ, FACT-G
	睡眠	PSQI, ESS
	関節リウマチ	BASDAI, ACR core set の一部, DAS28 の一部
	COPD, 喘息	SGRQ, AQLQ
	うつ症状	HADS, BDI, SDISS
<p>SF-36 : MOS(Medical Outcome Study) 36-item Short-Form health survey, EQ-5D : EuroQol 5 Dimension, VAS : Visual Analog Scale, MPQ : McGill Pain Questionnaire, WOMAC : Western Ontario & McMaster Universities Osteoarthritis Index, ODI : Oswestry Disability Index, JOABPEQ : the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire, BPI : Brief Pain Inventory, EORTC-QLQ : the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire, FACT-G : Functional Assessment of Cancer-general, PSQI : Pittsburgh Sleep Quality Index, ESS : Epworth Sleepiness Scale, BASDAI : Bath Ankylosing Spondylitis Disease Activity Index, ACR core set : American College of Rheumatology core set, DAS28 : Disease Activity Score 28, SGRQ : St George's Respiratory Questionnaire, AQLQ : Asthma Quality of Life Questionnaire, HADS : Hospital Anxiety and Depression scale, BDI : Beck Depression Inventory, SDISS : Sheehan Disability Scale</p>		

また、データ収集の頻度・実施場所によって、質問票（例えばベースライン評価時、来院ごとなど）と被験者日誌（毎日、朝と夜など）に大別され、収集法によって紙かデバイスを用いた電子的な収集（Electronic PRO [ePRO]システム）かに分けられる。ePRO システムには、質問票タイプの PRO について、施設からタブレットなどのデバイス通じてデータを収集する site PRO と被験者にデバイスを渡して毎日入力してもらうもの（いわゆる eDiary）とがある。ePRO システムは、被験者の入力やデータ管理が容易であるが、コストの増大も懸念されている。詳細は電子化情報部会からの報告書を参照されたい[7]。

1.2 臨床試験における PRO 評価の意義

臨床試験で収集するデータは、データソースにより、以下の 4 つの医学的指標に分類される[6]。

1. 患者報告アウトカム (PRO)
2. 保護者や介護者 (caregiver) が評価したアウトカム
3. 医療者が評価したアウトカム
4. 生理学的アウトカム (FEV1, HbA1c 等)

主要評価項目は、一般に対象とする疾患の性質や開発品の使用目的等に応じて選択される。その中でもエンドポイントの客観性は重要なファクターであり、これまで 3,4 のアウトカムが用いられていた。特に 4 の生理学的アウトカムは、客観的指標として多くの評価で用いられている。このように、より客観的な指標、たとえば、死亡率、心筋梗塞、脳卒中

の発症率等の、いわゆるハードエンドポイントを主要評価項目にする場合であっても、同時に1のPROも評価することで、医療に提供できる情報の幅が広がる。有効性のICH-E9ガイドラインでも、主要評価項目は「試験の主要な目的に直結した臨床的に最も適切で説得力のある証拠を与える」ものであるべきとしているが、QOLについても主要評価項目となる可能性があるとしている[8]。

実際に今日の日本の承認申請において、PROが主要評価項目として採用されているかを確認する目的で、平成22年1月～平成26年12月の5年間に日本で承認された医薬品208品目の申請資料概要を調査した。その結果、24品目(11.5%)で検証を目的とした試験の主要評価項目にPROが用いられていた。その内訳を表1-2に示す。24品目のうち、薬効分類別には中枢神経系用薬が7品目、代謝性医薬品が6品目であった。また、8品目で疼痛に関連したPRO尺度が用いられていた。そのほかにも表1-2で示した疾患・症状特異的なPRO尺度が多く用いられていることが確認できた。

表 1-2 主要評価項目にPROが用いられていた品目の薬効分類別一覧

薬効分類	薬効分類 番号*	品目数 (計**)	主な疾患名 (主なPRO尺度)
中枢神経系用薬	112,116, 119	7 (24)	パーキンソン病, 不眠, そう痒, 神経痛 (そう痒 VAS, 疼痛スコア, 睡眠調査票)
合成麻薬	821	2 (2)	疼痛を伴うがん (疼痛スコア)
消化器官用薬	239	2 (8)	機能性ディスぺプシア, 悪心・嘔吐 (SF-NDI : Short Form of the Nepean Dyspepsia Index, 被験者日誌)
泌尿生殖器官及び 肛門用薬	259	1 (2)	過活動膀胱 (排尿回数)
血液製剤類	634	1 (26)	血友病 A (出血エピソード)
代謝性医薬品	399	6 (46)	関節リウマチ, 変形性膝関節症 (HAQ : Health Assessment Questionnaire, 疼痛 VAS)
局所麻酔剤	121	1 (3)	局所レーザー照射療法 (疼痛 VRS : Verbal Rating Scale)
解熱鎮痛消炎剤	114	1 (1)	変形性関節症, 腰痛 (疼痛強度)
抗ウイルス剤	625	1 (14)	インフルエンザ (被験者日誌)
アレルギー用薬	449	1 (2)	そう痒 (Total four-symptom score)
眼科用剤	131	1 (5)	術後炎症 (眼痛評価スコア)
*厚生労働省薬効分類「87」に続く3桁			
**調査した医薬品208品目中, 各薬効分類に該当する品目数			

なお、2007年から2013年にかけて ClinicalTrials.gov に登録された試験について調査した結果では[9]、製薬企業が実施した臨床試験 30,012 試験（検証試験以外も含む）のうち、1887 試験（6.3%）で主要評価項目に PRO が設定されていた。

1.3 各規制当局の動き

米国 FDA は 2009 年に PRO の測定についてのガイダンス *Guidance for Industry Patient-Reported Outcome Measures*[5]を配布している。現在のところ日米欧の規制当局から出されている唯一の PRO に関する直接的なガイダンスであり、PRO をめぐる環境を大きく変えたものであるため、その詳細は 2 章で紹介する。本ガイダンスは PRO の結果を効能・効果として表示するための開発企業のガイダンスとして交付されたものであるが、それに加え、新たに PRO 尺度を開発する際の内容の妥当性、信頼性、臨床における取り扱いや、FDA の審査に際しての提出内容並びに留意事項なども記載されている。

近年米国においても、新薬の価値として、プラセボに勝るという絶対的価値に加えて、治療選択のためのエビデンスとなる相対的価値（Comparative Effectiveness）が重視され始めている。利用上意味のある価値基準を追及することが要求され、過去の有効性重視の反省から安全性とのバランスを考える上において、被験者が総合的にどう感じるかを定量化する必要性が生じてきた。これらのことが PRO という概念に結び付き、臨床試験での活用に注目が集まっていると考えられる。先日開催された FDA-DIA stat workshop [10]でも PRO が話題に上るなど、上記ガイダンスの配布後も精力的に議論が継続されている。また、Critical Path institute（C-PATH）の PRO コンソーシアムの活動も活発に行われている[11]。

一方、EMA には PRO に関する直接のガイダンスは存在しないが、2005年に制定された *Reflection Paper on the Regulatory Guidance for the use of health related Quality of Life* [3]や、2009年に制定されたバイオマーカーなどを含めた新たな手法を開発する際のガイダンス *Qualification of novel methodologies for drug development* [12]がある。抗がん剤に対しては、*Guideline on the evaluation of anticancer medicinal products in man* [13]が 2012年に公布され、その中で PRO の適用に関する概略が述べられている。その詳細については Appendix として添付されることとなっており、2016年11月に *The use of patient-reported outcome (PRO) measures in oncology studies* [14]が発効されることが決定している。

日本においては、現在のところ、PRO 関連のガイドラインを作成する動きは見られていない。当局から配布された関連したガイドラインとして、平成 3年に策定（平成 17年改訂）された抗悪性腫瘍薬の臨床評価方法に関するガイドライン[15]がある。ただし本ガイドラインでは、延命効果に加えて QOL の評価も併せて行うことが望ましいことが規定されているに留まっている。

今後、グローバル臨床開発が進むにつれて、PRO 尺度の日本語版作成の機会が増えていくことが予想される。さらには、欧米で被験者に焦点を当てた医薬品開発への取り組み

(Patient-focused Drug Development) が近年注目されていることから[16], 日本から新たな PRO 尺度が開発され, それがグローバルで活用されるようなケースも発生するであろう.

2 FDAのPROガイダンスの概要

本章では、2009年12月にFDAが交付した「FDA Guidance for Industry : PRO Measures : Use in Medical Product Development to Support Labeling Claims」[5]の概要を紹介する。まずFDAが重視しているPROを理解する上で重要な概念について説明し、次にPRO尺度開発の一般的なプロセスを紹介する。臨床試験デザインやデータ解析についても触れられているが、PRO尺度を利用しない臨床試験での留意点と大きな違いがないことからここでは割愛した。データ解析に関しては、4章のPRO利用時の留意点を参照されたい。なお、ISPOR(International Society for Pharmacoeconomics and Outcomes Research)日本部会がこの日本語訳である「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」を公開している[17]。

2.1 PRO尺度の概念的枠組み (conceptual framework)

本ガイダンスは、PRO評価を効能効果の根拠として用いることを念頭に、医薬品及び医療機器の開発企業に対して交付されたものであり、承認申請に用いられるPRO尺度および得られる結果をFDAがどのように審査・評価するかを記載している。

本ガイダンスでは、症状や機能、被験者が感じる効果などPRO尺度で測定しようとするものを「概念 (コンセプト)」と称している。例えば鎮痛剤の臨床試験であれば、評価しようとする概念は「痛み」である。この概念が、1つ以上の構成要素 (ドメイン) からなることもある。例えばリウマチの臨床試験では、腫れや痛みなどの構成要素に加えて、身体的機能の制限などの構成要素もリウマチの疾患の負荷を表す概念に含まれるだろう。ドメインはさらに複数の質問項目から成り、これらの質問の回答に重み付けやスコア化して、PRO尺度となる。

このようにまずは臨床パッケージ全体として、評価すべき概念を整理し、ドメイン構造を明記した全体的な枠組み (framework) を示すことが求められおり、これを概念的枠組み (conceptual framework) と呼ぶ (図 2-1)。

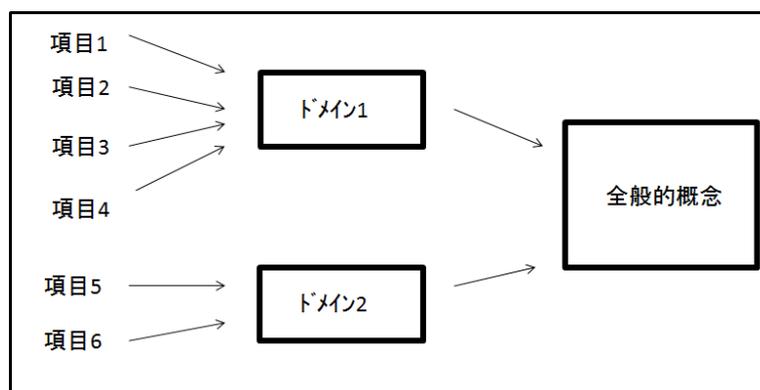


図 2-1 PRO 尺度の概念的枠組みを示すダイアグラム

ISPOR 日本部会「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」より引用[17]

ここでは全体像だけでなく個々の質問項目間の関係や、質問項目とドメインとの関係、さらにドメイン間や、ドメインと概念的枠組み全体との関係を明確に定義することが必要である。バリデーションの過程で、スコアの合計の変化がある特定の質問項目のみに起因するような場合は、この概念的枠組みを見直す必要性を考慮する。さらに、PRO が利用される被験者集団を明確に定義する必要がある。

この概念的枠組みが臨床試験の各エンドポイントでいかに充足されるかを示すことが重要であり、臨床試験で用いる PRO と他のエンドポイント間の関係(エンドポイントモデル)を整理する必要がある。ガイダンスで紹介されている例を図 2-2 に示した。PRO を主要評価項目として用いるのか、重要な副次評価項目、あるいは探索的な評価項目として補助的なデータとして用いるかなどを検討し、PRO が果たす役割を定義する。

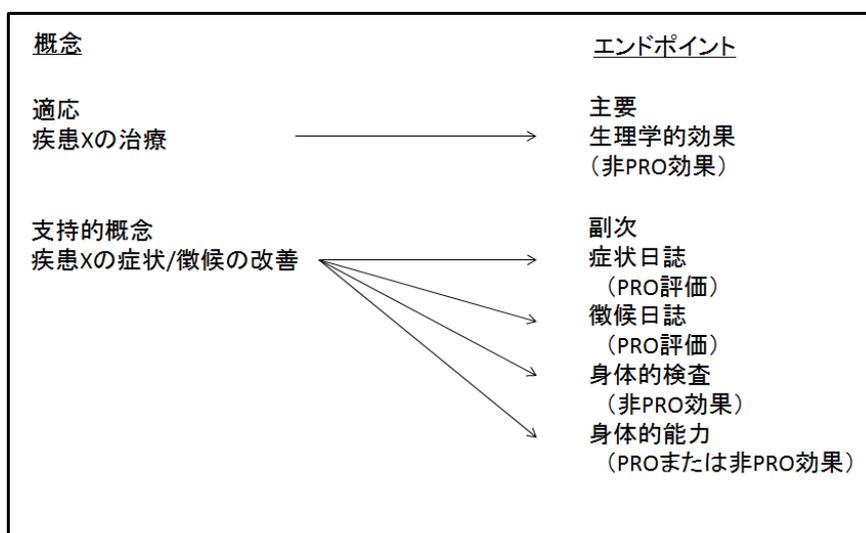


図 2-2 エンドポイントモデル：疾患 X の治療

ISPOR 日本部会「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」より引用[17]

2.2 内容妥当性 (content validity)

内容妥当性 (content validity) とは、問題や質問の内容が測定したい領域を反映しているか (領域の範囲内から選ばれているか、領域から偏りなく選ばれているか) である。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することやそれで調整することはできないため、PRO 尺度の妥当性評価 (validation) において、最も重要なポイントである。

PRO 尺度は、被験者にとって重要な概念がもれなくカバーされ、それが質問項目に適切

かつ十分に反映されていなければならない。そのため、PRO 尺度の作成プロセスにおいては、被験者からの情報が必須であり、以下のような情報及び検討結果が FDA の審査過程で要求されている：

- 項目の内容：臨床的に最も重要な概念や項目が反映されているか、また被験者が質問を理解できるかなど
- データ収集方法と尺度記入法：尺度の記入に関する手順は適切か
- 想起期間：被験者が状態を思い出せる最適な期間を設定しているか
- 回答方法：質問の表現が被験者に対して明瞭であるか、質問が適切な順序で設定されているか、スコアの分布が極端に歪んでいないか（天井効果や床効果）など
- スコアリング：スコアに適切な尺度（順序尺度、間隔尺度など）が設定されているか、スコアの重み付けは適切か、合計スコアの適切性など

2.3 新規 PRO 尺度の開発と妥当性評価

FDA ガイダンスでは、新薬および新医療機器の開発に当たり、対象とする概念を評価するための適切な PRO 尺度の有無を調査し、ない場合に限り、新規に PRO 尺度の開発が可能であると記載されている。この開発には既存の PRO 尺度の修正も含まれる。PRO 尺度の修正の場合、PRO 尺度は対象集団や疾患に特有なものであることから、異なる集団に対する利用の適切性の確認が必要になる。例えば、既存の PRO 尺度を小児などの新たな被験者集団で利用する場合は、内容妥当性などの点から慎重に検討することが必要となる。用途またはフォーマットなどの軽微な変更の場合には、必ずしも詳細な再検討が必要ではない場合もあり、個々のケースでの対応が必要である。

FDA ガイダンスでは、以上のそれぞれのプロセスにおいて、実施した内容・エビデンスなどを記録することが求められている。また、ガイダンスには、FDA が審査する際に企業に提出を求める情報、留意事項などが付録として添付されている。

PRO 尺度のバリデーションにあたっては、内容妥当性の評価に加え、実際に対象集団から収集したデータを用いた、尺度の信頼性や妥当性などの確認が必要である。PRO 尺度の開発の際に確認が必要とされる内容を表 2-1 に示した。

表 2-1 PRO 尺度の開発の際に確認が必要とされる内容

<p>信頼性 (reliability)</p>	<ul style="list-style-type: none"> ・ 一貫性のある結果を導けるか，治療の真の効果を再現性よく推定できるか ・ 再現性 (test-retest reliability): 同じ被験者で同じ PRO 尺度を繰り返し測ったとき，同じ結果が得られるかという再現性テスト．ただし，何度も繰り返し行くと前回の回答を記憶している可能性があるため，適度な間隔をあけて実施する[6]． ・ 実施者間の再現性 (Inter-interviewer reproducibility) : PRO 尺度が面接調査担当者によって使用される場合，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[6]．
<p>構成概念妥当性 (construct validity)</p>	<ul style="list-style-type: none"> ・ 項目・ドメイン，およびコンセプト間の関連性：事前に設定された仮説に関して，他の評価項目との間に論理的に妥当な関連性がある（ある項目に関して，一般的な1つの質問と具体的な複数の質問との間等）
<p>基準関連妥当性 (criterion-related validity)</p>	<ul style="list-style-type: none"> ・ 同じコンセプト内で PRO 尺度と他の診断や評価の精度が高いものとして広く容認された手法との間の相関の程度 (PRO vs 終夜睡眠ポリグラフィー検査等)． PRO 尺度と他の疾患や病態等の評価尺度との相関性の基準そのものは一般的には存在しない
<p>変化を検出できるか (change detection)</p>	<ul style="list-style-type: none"> ・ 経時的变化 ・ 他の同様な基準との比較

3 臨床試験計画のための PRO 尺度準備プロセス

臨床試験を計画するにあたり、痛みや眠気、QOL のような被験者からの報告以外に適切な評価方法がない場合、PRO を収集することになる。この決定から実際に治験で利用する PRO 尺度（患者日誌など）の選定（あるいは作成）までのプロセスを図 3-1 に示した。まずは既存の PRO 尺度の使用を検討し、治験を実施する国の言語版（例えばアメリカ用の英語版とスペイン語版、カナダ用の英語版とフランス語版など）が存在するかを確認する。必要な各国語版がない場合には、言語バリデーションを行う。

臨床パッケージとしての概念的枠組みに沿った「その臨床試験で証明したい概念」を示すのに最適な PRO 尺度が存在しない場合には、新たな PRO 尺度を開発し、バリデーション（妥当性評価）を行うことになる。これは通常何年もかかるプロセスであり、第一相試験前後から準備が必要である。さらに新規 PRO 尺度のバリデーションにはある程度の被験者数が必要となることから、通常は後期第二相試験で使用し、その結果を利用してバリデーションとし、第三相試験で利用できるよう準備することが多い。

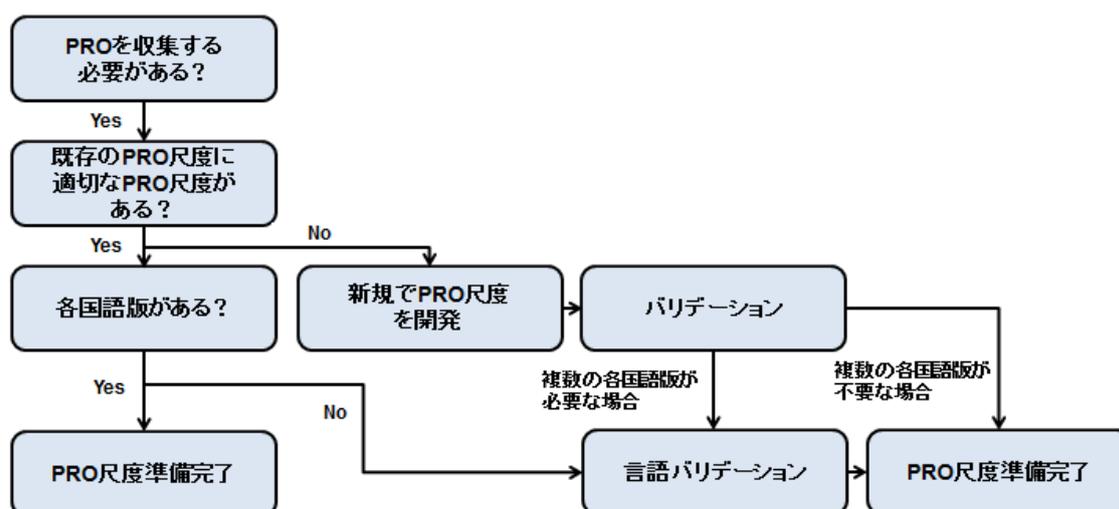


図 3-1 臨床試験における PRO 尺度の準備プロセス

本章では、このプロセスについて順を追って紹介する。

3.1 社内体制

本タスクフォースが 2015 年夏に実施したアンケート結果によれば、約 8 割の会社で PRO 専門部署がなかった（付録 Figure 4）。海外にも日本にも専門部署がなく、PRO 専門家もない場合には、外部専門家（業者）と契約する、または社内で専門知識を持つ担当者を育成する必要があるだろう。大手の CRO であればアウトカムリサーチやそれに準ずる部署

があり、PROの開発やバリデーションのノウハウもあるし、他にもPROに特化したような業者もある。また、より効率よく業務を推進するためには標準的に使用される手順書等（以下、Standard Operating Procedure : SOP）やそれに準ずる手順書の整備も進める必要があるだろう（付録 Table 3）。

3.2 臨床パッケージとしてPROの必要性の検討

前章で述べた通り、その疾患に対する対象医薬品の有効性・有用性や安全性を適切に評価するために、エンドポイントとしてのPROの必要性を吟味しておく必要がある。また、エンドポイントモデル、および概念的枠組み等について早期に検討することが重要である。この概念的枠組みを熟慮することで、例えば痛みを評価する場合でも、慢性的な痛みか、短時間の耐え難い痛みか、痛みの酷さか、痛みを感じている時間かなど、何が患者にとって意味があるのか（relevant）、何を指標として評価すべきなのかを明確化・具体化することができる。

PRO尺度の準備には、下記のようにある程度時間がかかることを念頭に置いて、臨床パッケージの企画・立案やエンドポイントモデルの策定の段階からPRO担当者との協議を開始し、検討を開始する。

3.3 既存のPRO尺度を選択する

臨床パッケージでPROが必要でも、むやみに新規のPRO尺度を作成しようとせず、まずは既存のPRO尺度の使用を検討する。PRO尺度の検索法としては、審査報告書や米国の臨床試験登録・公開サイト[clinicaltrials.gov]、論文などから過去の治験で使用されたPRO尺度を検索する、あるいはePROVIDE [<https://eprovide.mapi-trust.org/>]（旧PROQOLID）などのウェブサイトから検索するなどの方法がある。

一般的には後期第二相試験まではさまざまなPRO尺度を試し、または妥当性を評価し、その中からよりの確に医薬品の特性を示せるもの、被験者の反応を感度よく拾えるものを第三相試験に残せるようにする。そのため、後期第二相試験までは多くのPRO尺度の準備を同時に進めなければならないことが多い。

PRO尺度を選択するに当たっては、少なくとも下記のことを考慮する必要がある。

- ・ 臨床試験の対象集団（疾患、重症度、年齢など）
- ・ 臨床試験の目的とデザイン（エンドポイントモデル）
- ・ PRO尺度の概念的枠組み（何を示したいのか）
- ・ PRO尺度の測定特性

本タスクフォースのアンケートの結果では、PRO 尺度の選定の際に考慮する点として、「日本語版があるか/言語バリデーションされているか」、「過去の利用頻度」、「FDA で承認された試験で使用されているか」、「検出感度」などが挙げられた（付録 Figure 8）。その他としては、コスト・運用面での簡便性、学会や研究会が推奨しているか・臨床現場で広く認知されているか、という意見も挙げられた。

PRO 尺度を選択したら、原作者（著作権等を含む、当該尺度の権利を有す個人もしくは法人）と連絡を取り、使用の条件等を確認して、必要に応じて使用等に関する契約を締結する。医師や研究者が治療や臨床研究の目的で PRO 尺度を利用する場合と異なり、治験での使用は商業目的となることに留意する。契約等の手続きには時間がかかることもあるが、正式に使用許可を求め、治験の規模や使用期間など各種条件に応じて支払いを行う。著作権は研究者などの原作者が保持している場合もあるし、研究所や病院、企業が保持している場合もある。オリジナル言語版だけでなく、それをもとにした各国語版についても著作権は原則として原作者にある。例外的に翻訳版の著作権が各国の提携先に移管されていることもあるが、使用許可は原作者または原作者の委託先に求める。

3.4 新規 PRO 尺度の作成

既存の PRO 尺度が目的とする概念を十分に評価できないと判断した場合には、PRO 尺度を新たに作成することを考慮してもよい。2 章で紹介した FDA ガイダンスでは、以下のプロセスを例示している。PRO 尺度の開発は、案を作成し、試用して改訂し、改訂版をまたテストするというプロセスであり、以下を反復的に行うことにより達成される。

- i. 概念的枠組みの決定：特定の対象集団に対して、PRO 尺度の測定対象および添付文書上での適応を仮定し、スコアのタイプや回答方法など要素を決定する。文献調査や専門家のレビューを行い、概念的枠組みを開発する。
- ii. 概念的枠組みの調整および調査票（案）の作成：被験者からの情報を収集し、必要に応じて新規の質問項目を追加する。想起期間や回答フォーマットを選択し、実施方法、データ収集方法を決定する。質問を被験者が理解可能かなどの情報をインタビューから確認し、ドラフトを作成する。
- iii. 概念的枠組みの確認および他の測定特性の評価：スコア化と概念的枠組みを確認し、スコアの信頼性、構成概念妥当性、検出感度を評価する。PRO 尺度の内容、フォーマット、スコア法、実施手順、トレーニング資料を最終化する。
- iv. データの収集、解析およびデータの解釈：プロトコルおよび解析計画を作成し、データを収集・解析し、治療に対する反応を評価する。
- v. PRO 尺度の修正：得られたデータから、対象集団、質問項目の表現、回答の選択肢、

想起期間、実施・データ収集方法の変更の必要性を検討する。多言語で利用する場合は、翻訳し、文化的に適用できるように調整する。さらに、修正された内容が適切かどうかを評価する。

以下に上記の具体的な作業内容を参考までに紹介するが、これら一連のプロセスは専門業者に外部委託するのが一般的である。

i では、その疾患領域の医師やその他医療関係者にインタビューを行い、その疾患の重要な特徴を表す概念についての質問項目をできるだけたくさん挙げてもらう。さらに患者インタビューを行い、患者にとって重要な概念およびその概念を表す質問項目をできるだけたくさん挙げてもらう。インタビューの必要人数について特に規定はないが、できるだけ背景（性別、年齢、教育レベル、重症度など）が異なる患者から意見を聞くように工夫する。また、可能であれば国や地域にも多様性を持たせる。その後、質問項目を精査し、似たような質問項目をまとめ、多くの患者に共通して重要と考えられた概念に優先順位をつけるなどして、質問項目を絞っていく。このとき、概念的枠組みのどのドメインにどの質問項目が当てはまるのか、考えながら選択する。質問項目の数に決まりはないが、あまり多くの質問項目に回答するのは試験時に被験者の負担が大きくなることも考慮する。

ii では各質問の具体的な文言や、回答のオプションについて検討する（選択肢, VAS, Likert scale など[18]。質問は基本的に過去のことを思い出してもらうものなので、できるだけ生活に密着した質問とし、思い出してもらいやすいように工夫する。例えば「運動したら息が切れた」ではなく、「散歩するくらいの速さで歩いたら息が切れた」「雪かきをしたら息が切れた」のように、後からでもその場面をありありと思い出せるように、あるいはその期間にはその経験をしなかったことを思い出してもらえようにすることが肝要である。また、具体的にいつのことを思い出してもらうのか（この1週間、昨晚、今日1日など）も設定する。思い出す期間（recall period）は重要であり、したがってこの段階では紙で検討していても、最終的には ePRO にした方がよいかも合わせて検討する。案ができれば、臨床試験の対象集団に含まれるような患者、あるいはできるだけ近い背景を持つ患者に読んでもらい、読めるか、わかるか、意味があるかを確認するためのパイロット調査を行う。この段階では英語版しかないことがほとんどであり、患者パイロットも英語圏で行われることが多いが、もし日本語で作成を始めるなら、パイロットも日本語圏で行われることになる。患者数に決まりはないが、最初のインタビューと同様、さまざまな背景を持つ患者が含まれるよう工夫する。パイロット調査の結果、例えば理解が難しいと言われた単語を置き換える、自分の疾患にあまり関係がないと言われた質問を除く、もっと重要な概念がある場合はそれに関する質問を追加するなどして、質問票を改訂する。

iii, iv については、一般的には後期第二相試験でデータ収集を行うことが多い。これまでどのようにその質問票を作成し、テストし、改訂してきたかとその理由をまとめて当局と相談し、使用が妥当であるとの結論が得られれば実際に治験で使用し、ここで得られた

結果を用いて前章の表 2-1 のような内容について示すことになる。この時用いられる代表的な計量心理学的な検討方法には表 3-1 のようなものがある[19]。通常の相関係数の算出や分散分析等、通常の臨床試験で用いる統計解析手法も含まれるが、中にはあまり用いられない手法も含まれる。以下、代表的な手法について説明する。実際の使用例としては SF-36 の論文を参照されたい[20, 21]。

表 3-1 2004 年以降の PRO 関連論文 181 報に用いられた計量心理学的な検討方法[19]

計量心理学的な検討方法	使用件数 (%)
クロンバックの α 係数	157 (86.7)
相関係数 (構成概念妥当性)	136 (75.1)
分散分析	83 (45.9)
級内相関係数	81 (44.8)
因子分析, 主成分分析, 構造方程式モデリング	79 (43.7)
応答性 (検出感度)	50 (27.6)
効果サイズ	25 (13.8)
ラッシュモデル (1 パラメータロジスティックモデル)	19 (10.5)
多特性多方法行列	15 (8.3)
臨床における最小重要差, 反応者分析	12 (6.6)
回帰分析	9 (5.0)
項目応答理論	4 (2.2)

1. クロンバックの α 係数 (Cronbach's coefficient alpha) [22]

PRO 尺度の質問間の一貫性の程度、すなわち内的整合性を確認する際に用いる。信頼性そのものを算出しているわけではないことに注意が必要である。計算には以下の式を用いる。

$$\alpha = \frac{\text{質問数}}{\text{質問数} - 1} \left(1 - \frac{\text{各質問の分散の和}}{\text{合計点の分散}} \right)$$

伝統的に、0.7 以上の値が望ましいとされているが、1 に近い場合はすべての質問が全く同じものを測定していることになり、無駄に多くの項目を測定している可能性があると考ええる。また、各質問を除いた場合の α 係数を算出し、その値が大きい場合には構成要素として不適切である可能性がある。

2. 級内相関係数 (Intraclass correlation coefficient : ICC)

相関係数は 2 種類のデータの大小関係が一致しているかどうかを表す指標であり、データそのものが一致していなくても高い値を取りうる。そのため、データの値そのものが一

致しているかを調べる場合には級内相関係数 (ICC) を用いることができる。ICC は Case 1, Case 2, Case 3 の 3 種類が定義され[23], その下位モデルは, 以下に示すように ICC(k, 1), ICC(k, n) と表される (k = 1, 2, 3).

ICC(1,1) : 1 人の評価者が n 回評価した時の評価者内信頼性

ICC(1,n) : 1 人の評価者が n 回評価した時の評価平均の信頼性

ICC(2,1) : n 人の評価者が 1 回評価した時の評価者間信頼性 (変量効果モデル)

ICC(2,n) : n 人の評価者が 1 回評価した時の評価平均の信頼性 (変量効果モデル)

ICC(3,1) : n 人の評価者が 1 回評価した時の評価者間信頼性 (混合効果モデル)

ICC(3,n) : n 人の評価者が 1 回評価した時の評価平均の信頼性 (混合効果モデル)

ICC(1,1), ICC(2,1), ICC(3,1) は評価者内・評価者間信頼性を, ICC(1,n), ICC(2,n), ICC(3,n) は複数回測定した平均値を使用した際の評価者内・評価者間信頼性を求める場合に用いる。たとえば, 評価者 3 名が被験者 5 名の歩行速度を測定したとすると, 評価者 3 人の一貫性 (評価者間信頼性) は ICC(2,1) で求める。

変量効果モデルと混合効果モデルの違いは, 評価者を変量効果とするか固定効果とするかである。すなわち Case 2 では絶対的な一致を, Case 3 は相対的な一致を確認する。一般的に PRO 尺度の開発においては混合効果モデルを用いることが推奨されている[24]。ICC の判定基準は表 3-2 のようなものが存在するが, 理論的根拠はまったくない。また, ICC(3,n) はクロンバックの α 係数と一致する。

表 3-2 ICC の判定基準[25]

ICC の値	判定
0.0 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	almost perfect

3. 構造方程式モデリング (Structural Equation Modeling : SEM) [26]

構造方程式モデリング (SEM) とは, 各種変数間の理論的な関係性を仮定し, それをモデルとして描き, それが妥当であるかどうかを検証するための分析方法である。共分散構造分析とも呼ばれる。相関分析, 回帰分析, そして因子分析を統合した分析で, 実際には観測できない因子や, 複数の従属変数を分析内に含むことが出来ることが特徴である[27].

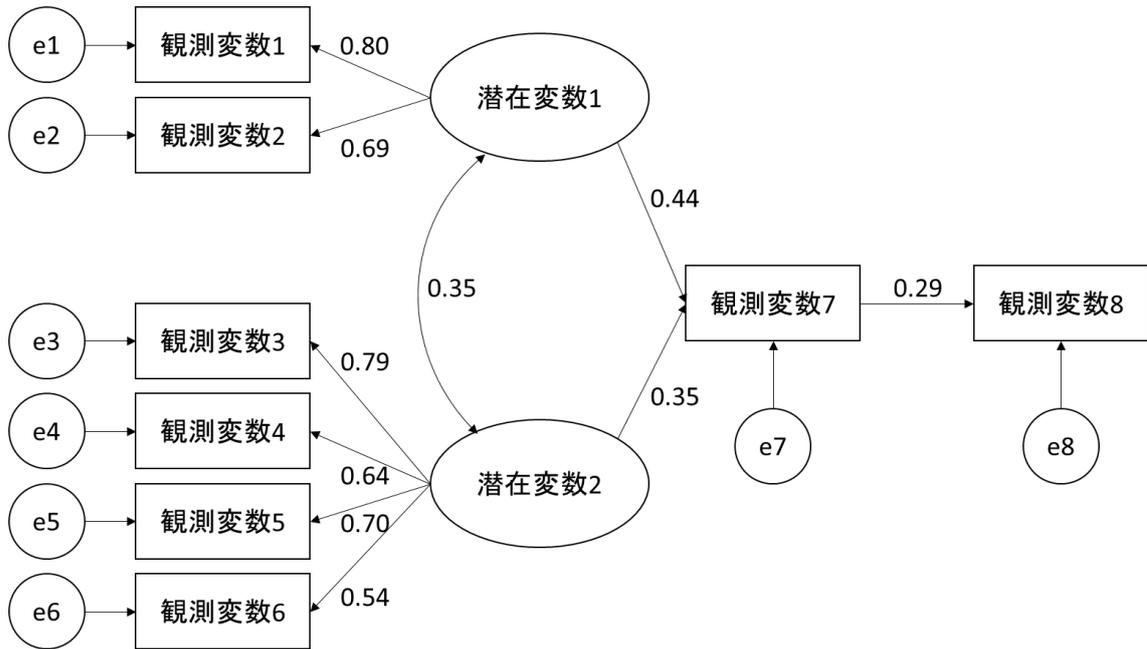


図 3-2 構造方程式モデリングの図例

図 3-2 内の正方形は観測変数，楕円は潜在変数を表す．単方向の矢印（パス）は変数間の予測の方向性を表すものであり，その上にある数字は重回帰分析の標準化偏回帰係数に相当する．一方，双方向の矢印上の数値は相関係数である．また単方向の矢印を受ける変数には必ず誤差変数が付属する（円 e1～e8）．

SEM は変数間の複雑な関係性をパス図で表すことが出来る分析手段である．あくまで理論的な仮説が正しいかどうかを検証する統計的手段であることから，探索的ではなく検証的な分析である[28]．先行研究等から変数間の仮説をモデル化し分析を行うことが SEM の理想的な使用方法である．しかしながら，「当初想定していた共分散構造モデルが，何の変更もしないでデータにぴったり適合する，ということは稀」[29]であるため，仮説モデルを修正しながら最適なモデルを探索的に検討することがしばしば行われる．ただし，修正されたモデルが科学的に意味があり，理論的に説明可能かどうかを検討し，それを適切に説明することが重要である．具体的な利用例は参考文献[30]を参照されたい．

4. 多特性多方法行列（Multitrait-Multimethod matrix : MTMM） [31]

多特性多方法行列（MTMM）は構成概念妥当性を評価する方法として提唱された．ある特性を測定したい際，用意した質問項目の信頼性が高くとも，その質問項目が必ずしも目的とする特性を測っているとは限らない．そこで MTMM ではある特性を複数の方法によって検証し，以下の項目の結果を相関行列として表す．

- i. 同一の特性を同一の測定法によって測定する (信頼性係数)
- ii. 同一の特性を異なる測定法によって測定する (収束的妥当性)
- iii. 異なる特性を同一の測定法によって測定する (弁別的妥当性: 異特性同一方法相関)
- iv. 異なる特性を異なる測定法によって測定する (弁別的妥当性: 異特性異方法相関)

		方法1			方法2			方法3		
方法1	A1	0.89								
	B1	0.51	0.89							
	C1	0.38	0.37	0.76						
方法2	A2	0.57	0.22	0.09	0.93					
	B2	0.22	0.57	0.10	0.68	0.94				
	C2	0.11	0.11	0.46	0.59	0.58	0.84			
方法3	A3	0.56	0.22	0.11	0.67	0.42	0.33	0.94		
	B3	0.23	0.58	0.12	0.43	0.66	0.34	0.67	0.92	
	C3	0.11	0.11	0.45	0.34	0.32	0.58	0.58	0.60	0.85
	特性	A1	B1	C1	A2	B2	C2	A3	B3	C3
		方法1			方法2			方法3		

信頼性係数
 収束的妥当性
 弁別的妥当性(異特性同一方法)
 弁別的妥当性(異特性異方法)

図 3-3 MTMM の図例

MTMM の基本原則として、値が信頼性 > 収束的妥当性 > 弁別的妥当性 (異特性同一方法相関) > 弁別的妥当性 (異特性異方法相関) となるのが理想的であり、これが証明された場合に、構成概念妥当性ありと判断される[32]. 具体的な方法は、使用例を参考にされたい[33, 34].

5. 項目応答理論(Item Response Theory: IRT) [35, 36]

項目応答理論 (IRT) は、被験者の特性や、評価項目の困難度、識別力を測定するための理論である。IRT では能力値や困難度のパラメータを推定し、データのモデル適合度を確認し評価項目の適切さを吟味することができる。一般的には、項目への離散的な応答の確率を、1 種類の被験者パラメータ、1 種類以上の項目パラメータで構成されるモデルを用いて表す。通常ロジスティック曲線を当てはめたモデルが用いられる。用いられる変数は以下のとおりである。

θ : 被験者パラメータ, 各被験者の特性の大きさを表す実数値

a_i : 識別パラメータ, 質問項目 i が被験者の能力を識別する力を表す実数値

b_i : 難易度パラメータ, 質問項目 i に適切に応答する難易度を表す実数値

c_i : 当て推量パラメータ, 多肢選択形式の質問票の場合に被験者が偶然適切に応答する確率を表す実数値

最も簡単な 1 パラメータロジスティックモデル（ラッシュモデルと呼ばれる）では、変数に θ と b_i のみを用いる。このモデルでは質問項目 i に適切に回答する確率は次の式で与えられる。

$$p_i(\theta) = \frac{1}{1 + \exp(-(\theta - b_i))}$$

2 パラメータロジスティックモデルではさらに a_i を用い、各項目が評価にとってどの程度適正な判断基準であるかを変数に組み込む。このモデルでは質問項目 i に適切に回答する確率は次の式で与えられる。

$$p_i(\theta) = \frac{1}{1 + \exp(-Da_i(\theta - b_i))}$$

ここで D に 1.701 を用いると、ロジスティック関数を累積正規分布関数に近似することができる。

3 パラメータロジスティックモデルでは、当て推量を考慮した c_i を組み込む。このモデルでは質問項目 i に適切に回答する確率は次の式で与えられる。

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp(-Da_i(\theta - b_i))}$$

一般的に、各パラメータの真値は未知であるため、最尤推定法やベイズ推定法を用い、それぞれのパラメータの値を推定する。具体的な使用例は参考文献を参照されたい[37]。

最後にこのような手法でバリデーションが取れた PRO 尺度を v で最終化する。国際共同治験で利用するためには、言語バリデーションを行う（詳細は次節で述べる）。また、その際にそれぞれの国や地域の被験者の生活にフィットするように **cultural adaptation**（フォークを箸に変えるなど）を行うこともある。

このように、新規 PRO 尺度を作成し、バリデーションするには、何年もの時間とコストがかかる。本タスクフォースのアンケート結果では、44 社中 16 社（36%）で新規 PRO 尺度の作成経験があるとの回答だったが、作成経験のある企業でもバリデーション試験は実施していないという回答が複数あった。日本の当局からは特にガイダンスは出されていないが、本タスクフォースのアンケートでも、「治験相談の中でバリデーションの実施を指示された」「バリデーションが取れていない PRO 尺度を用いてデータ収集した結果、審査報告書から PRO の結果のみ除外された」との回答もあり、審査の過程でバリデーション実施の有無が影響する可能性はある。

また、ePRO の場合には紙と比べて画面の文字が見えにくいこともあり得る。日本語版の

子供用では読み仮名を振ることもあるが、ePRO の小さな画面では字がつぶれてしまうこともあり、ひらがなにするなどの対応が必要な場合もあるだろう。アンケートの結果でも、特に高齢者に対して、文字を大きくする必要があったとの回答もあった。

3.5 言語バリデーション

言語バリデーションは単なる翻訳プロセスとは異なり、適切に作成され、妥当性を確認された PRO 尺度と「言語的に同等のもの」を作成し、保証するプロセスである[38]。したがって、単なる翻訳だけでは不十分であり、言語バリデーションの専門業者に外部委託する。オリジナルの PRO 尺度が適切に作成され、バリデーションされている場合で、治験に参加する各国語版が存在していない場合に行う。逆に言えばオリジナルの PRO 尺度が適切に作成されておらず、妥当性も確認されていないなら言語バリデーションには意味がない。PRO 尺度によっては言語バリデーションのプロセスを正式に踏んでいない、あるいはそのプロセスを踏んだかどうか不明確である言語版が存在することもあるが、国際共同治験において「そのエンドポイントで何を測っているか」が国によって異なるような事態は避けたいであろう。したがって、時間と手間がかかっても言語バリデーションのプロセスをおろそかにするべきではない。

本タスクフォースのアンケート調査では、PRO 尺度のオリジナル版が英語などの他言語版であった場合に、その日本語を作成した経験があるのは、44 社中 13 社 (30%) であった。このうち 10 社では「言語バリデーションを実施した」が、3 社では「翻訳したのみ」であった (付録 Figure 11)。また「当局から言語バリデーションについて確認された」とのコメントがあった。さらに「PRO の設問順が英語版と日本語版で異なっていたことがあった」との回答があった。これは単純なミスでないとすると、オリジナル版が何度か更新されており、いずれかの段階で作られた日本語版と更新された英語版のバージョンが異なってしまうためかもしれない。PRO 尺度は前節で述べたようなプロセスを経てバージョンアップされることがあるため、各国語版作成に当たってはオリジナル版のバージョンを注意深く確認することが必要である。

言語バリデーションの流れは以下の通りである。このプロセスは業者にもよるがおよそ 4 か月かかることに留意されたい。社内プロセスとして、プロトコル承認が下りないと、原作者や業者との契約や支払いができない場合もあるだろう。しかし、IRB や当局へ各国語版を提出する必要があることを考慮すると、言語バリデーションのプロセスはプロトコル承認より早く始めなければ間に合わないこともある。プロトコル承認前では、使用する PRO 尺度やそのバージョン、治験参加国が確定していないなど不確定な要素もあるが、チームでタイムラインをよく確認して進める必要がある。

1) 準備

まずは本当に必要とする言語版が存在しないかを確認する。国際共同治験時には、複数の言語版をまとめて準備するために見落としが発生することもあるので、日本語版の存在は念のために日本チームが確認するといいたい。言語バリデーションの結果が論文化されていなくてもその言語版がすでに作られていることもあるし、論文が日本語の場合に海外の担当者からは発見されない場合もある。必要とする言語版がない場合には、言語バリデーションの専門業者に外部委託するため、必要な国と言語の組み合わせ（1対1とは限らない）を準備し、Request for proposal を作成して業者に送付する。言語バリデーションの業者が原作者から管理を委託されている場合もあり、各国語版の情報を持っていることもある。原作者へのコンタクトや翻訳許可依頼なども代行してくれるので相談してみるとよい。原作者から作成途中の言語版の情報が得られることもある。別の企業が作成途中であることがわかって、治験の実施時期によっては完成まで待てないこともあるだろう。このような場合に複数の微妙に異なる言語バリデーション版が存在することになる。以下に外部委託内容となる言語バリデーションの具体的な作業を参考までに紹介するが、業者との契約に当たっては下記のステップのうち、患者パイロット以下も含めてもらうかを決めて、締結する。

2) Forward translation

オリジナルの言語版から各国語版への翻訳プロセスで、2名の翻訳者が独立に翻訳を行う。この際、もしも学会のガイドラインや、その領域で基本的に使われる用語集（被験者が使う用語でなければならないが）などがあれば、業者を通じて翻訳者に知らせておくことも有用である。

3) すり合わせ (Reconciliation)

独立に行われた2つの翻訳版を見比べて、違いを確認し、なぜその違いが生まれたのかを考察する。些細な文言の差であれば単純にすり合わせることができるが、時には元の言語版が異なる2つの意味に解釈できることが判明することや、そもそも違う文化圏・言語圏では意味不明なこともある。このような場合には原作者にもととの質問の意図を確認し、時には違う文化圏・言語圏に合うような変更について提案するなどして、1つの翻訳版を作成する。

4) Backward translation とすり合わせ

上記3)で作成した翻訳語版からオリジナル言語に逆翻訳する。独立2系統で行う場合もあるが、1人の場合もある。できた翻訳をオリジナル版と比較して、大きな齟齬がある場合にはその原因を検討し、反映する。誤解なく理解できるものになれば、翻訳語版を仮確定する。また、例えば日本の開発チームでも事前に翻訳を確認したい場合などは、このタイミングで可能であることがあるので、契約時に確認しておく。ただし、言語として同等な

ものを作成するプロセスを厳密に踏んでいるため、「日本語として不自然である」などの理由では変更は認められない。

5) Cognitive debriefing (患者パイロット)

各国・地域の対象疾患の患者にこの段階での言語版を試してもらう。このプロセスは必須ではなく、必要に応じて実施する。このプロセスの有無によって言語バリデーション全体の費用もタイムラインも異なるので契約時に確認する。依頼する患者数に決まりはないが、各言語に 5-6 名であることが多い。読めるか、わかるか、意味があるか、また意図せずその文化圏では失礼な質問や恥ずかしい質問になってしまっていて不快感を与えないかなどを専門のインタビュアーが確認する。日本での患者パイロットに対して日本語が話せるインタビュアーが派遣されるとは限らない。通訳者をはさんだ被験者パイロットは被験者もインタビュアーもかなり疲弊するので、契約時によく確認するとよい。

6) 最終化

被験者パイロットの結果を反映し、最終化された PRO 尺度の最終確認を行う。この段階では文法や改行のエラーなどごく些細な変更以外は認められない。上記すべてのプロセスに関する報告書とともに、業者から言語バリデーションの証明書が送られてくる。

4 PRO 尺度を用いたデータ収集の留意点

PRO 尺度を用いたデータ収集の妥当性を担保するためには、PRO 尺度そのもののデザインや妥当性と同等に、正しく使用してもらうための被験者および施設担当者へのトレーニングが必須である。また、PRO は基本的には過去を思い出してもらい、回答してもらうものなので、後から欠測や矛盾に気づいても被験者以外の誰かが勝手に回答を修正することはできない。したがってまずは欠測や矛盾が起こりにくいような試験デザインを工夫し、かつ治験中に頻発するようであればそのリスクを早めに検知し、是正できるようなプロセスを構築することが重要である。

本タスクフォースのアンケート結果でも、当局から被験者へのトレーニングの実施を指示されたとの意見があった。トレーニング資料は PRO 尺度とセットであり、PRO 尺度そのものと同様に重要である。中には PRO の意義や PRO 尺度の一般的な利用法など共通に利用できるトレーニング資料もあるので、事前に作成して、治験開始前にどの施設でもどの担当者でも同じように理解しておいてもらえることを目指す。基本的には国や地域、施設によらず共通のものを使用することで、異なる言語版の PRO 尺度であっても「同じトレーニングを施した」ことを担保するため、オリジナルのトレーニング資料が日本語版でない場合に英語のまま使用する場合もある（付録 Figure 12）。トレーニング資料については言語バリデーションを行う必要はないが、国や施設、担当者による差が出ないように注意する。

4.1 治験依頼者による施設担当者のトレーニング

PRO の基本的な概念や意義について、治験依頼者は被験者に直接説明する機会を得られないため、施設の担当者（CRC 等）にトレーニングをおこなう。正しく使用されてこそ PRO 尺度の価値を発揮できるのであり、その PRO 尺度で意図したデータを正しく収集し、また治験全体のデータの質を向上させるためにも、まずは施設担当者に PRO の意義や目的を明確に理解してもらい、それを被験者に伝えてもらうことが重要である。Investigators meeting で PRO 担当者が直接説明するのに加え、日本では施設セットアップの際にもトレーニングの機会があるので、治験依頼者側は誰がどの施設で説明しても同じトレーニングが施せるように、社内でトレーナーとなる人をまず初めにトレーニングする必要がある。例えば共通のスライドを用いるなど、investigators meeting や各施設セットアップで同じトレーニングが施されるように工夫する。本タスクフォースが 2015 年夏に実施したアンケート調査の結果でも、各社がトレーニング資料を工夫して準備している様子がうかがえるので参考にされたい（付録 Table 5）。

まず、施設担当者に理解してもらわなければならないことは、PRO が「被験者の回答について、臨床医や他の誰の解釈も介さず、被験者から直接得られる被験者の健康状態に関するすべての報告」という原則である。そのため、被験者から報告されたそのままを尊重

するよう徹底する。Visit 時に質問票に回答してもら場合など、質問票の文言について、被験者から理解を確認するような質問があるかもしれないが「思ったように書いてください」以外の回答をしてはならない。被験者がわからない、理解できないのであれば、それそのものがその被験者からの回答だからである。それぞれの施設担当者が自由に言い換えていては、苦勞して作成した元の質問の意図を保証できなくなってしまう。また持ち帰って回答してもら患者日誌などの場合には、なぜ事前に設定した思い出し期間（記載・入力期限）を被験者に守ってもら必要があるのか、その意義を十分に説明し、visit 時に無理に思い出させようとしたり、欠損値をなくすために後から回答させたり、矛盾する回答を後から無理に直させたりしてはならないことを十分に理解してもら。紙の患者日誌の日付はあらかじめ記載してから被験者に手渡してもら。

また、被験者が落ち着いて適切に評価できる院内の環境を整えることも重要である。被験者が質問票に回答する際に、評価の場所・状況と評価の順番が影響することも知られている[39]。例えば、有害事象の記録の後に QOL 調査を実施した場合は、それまでに経験した有害事象の記憶と印象から、PRO の結果が実際の状態と異なってしまいうことも起こり得るので[39]、被験者が回答する場所や環境は、治験期間を通じて可能な限り似た環境にするよう努める。さらに、PRO 評価の前に院内スタッフや家族が被験者の状態の印象をコメントすることや、回答時の他者の存在も、PRO 結果に影響を与えることが生じうる。質問票への回答は担当医との面談や検査結果の説明など、あらゆる臨床的なフィードバックの前に行うように徹底し、他者に影響されない静かな環境で実施できるようにする。被験者に対して強制にならないように注意し、被験者に不安感を与えないようにする。

4.2 施設担当者による被験者のトレーニング

まず被験者に対しては、正しい答えや間違った答えはないこと、回答が治療の意図や治療法に影響を与えることは決してないことを十分に理解してもらい、思ったままの回答を期待していることを施設担当者から伝える。トレーニング資料としては「入力・記載の手引き」が挙げられる。紙の被験者日誌や質問票の場合には裏表紙などに印刷して直接手渡し、さらに施設担当者からの説明や、施設で各被験者にその場で見せられるようなラミネート加工したごく簡単な手引きを示すこともある。ビデオなどの動画を準備することもある。思い出し期間の説明や VAS scale であればどのように印をつけてほしいかの例示などが含まれる。被験者が子供の場合や高齢者の場合には、被験者に応じたユーザーフレンドリーな資料の作成が望ましい。

ePRO デバイスを用いる場合には、実際にそのデバイスを用いたトレーニングを行う。この時、基本的な機器の取扱いを記したマニュアルや、ID やパスワードの管理についての留意点などを示す。電子化情報部会のアンケート調査でも、ePRO 導入経験のある 11 社中 10

社で ePRO デバイス自体の説明資料を作成していた[7]。また、デバイスの不具合や、アクセスコードを忘れるなどにより、データの欠測や、結果として紙媒体からの転記の発生によるデータの信頼性の低下を最小限にするため、ヘルプデスクを準備しておくことが望ましい[40]。アラート機能や入力画面構造についても、トレーニング時に伝えることが望ましい。トラブルシューティングマニュアルの作成、また事前のリスクの想定と対応を充実させておく必要がある。

被験者に PRO 尺度に慣れてもらうため、ベースライン来院以前にトレーニング期間を設定することもある。これは、PRO 尺度で自分の症状や行動を客観視することで自然に症状が安定・軽快する「トレーニング効果」を治験薬投与前に安定させ、日誌記載遵守状況がよくない被験者のスクリーニングにもなる。ePRO システムを使用する場合には、デバイスを簡便に利用できない被験者集団などに対して、このトレーニング期間を特に長く設けるのも良いだろう。患者日誌の場合には、被験者には実際に1週間か2週間分をあらかじめ記入させ、施設担当者には被験者と一緒に座ってその場で確認させることが重要である。

治験開始後の visit では、施設担当者と被験者とで一緒に PRO の回答を確認するが、欠損があっても決して無理に埋めさせようとはならない。確認している visit の時点では、もうその症状について思い出す期間としてあらかじめ設定されていた期間は過ぎているはずである。本タスクフォースのアンケート調査でも「CRC へのトレーニングが不十分で、欠測等を強制的に埋めさせている事例が発生している」との意見が寄せられた。

また、その日に rescue medication を使った、とあるのに、その日の症状は特にひどくなかったと報告しているような一見矛盾した回答に気づいた場合でも、無理に訂正を促すのではなく、その日のことを思い出せるか、思い出せるとしたらその回答で間違いなのか、あるいは単なるミスなのかなどを、あくまでも被験者と一緒に確認してもらう。もし本人から訂正の意思表示があれば二重線と本人のイニシャルなどで修正も可能だが、思い出せない場合には、次回から気を付けてもらえるように指導するに留め、決して修正を強要してはならない。矛盾した回答や回答欄のずれなど誤記載としか考えられないような回答もあるかもしれないが、施設担当者や医師が被験者のいないところで独自に回答を解釈して、修正するようなことがあってはならない。アンケートの回答でも「修正箇所があった場合、当局から修正時期、修正理由に関する問合せがあるため、あらかじめ確認しておく必要がある」「修正された場合、誰がいつ修正したかがわからないとの問合せがあり、修正のルールや手順を説明して納得してもらった」などのコメントがあった。

4.3 治験依頼者側でのデータレビュー

PRO データ収集時に、治験依頼者が確認すべきこと、対応すべきことは下記が挙げられる。治験開始時までに、PRO データ収集で起こりうるリスクを特定し、依頼者側であらか

じめ対応を決めておくことが望ましい。

- 妥当なデータ、信頼性におけるデータが得られていることの確認（事前に規定した手順が遵守されているかの確認・担保）
- 設定したデータ入力期間が実際の運用時に適切であるか（期間が短すぎて欠測が増加していないか）
- データの欠測の割合（未回答率）を確認し、回答者にかかる負担割合や、PRO 収集項目の妥当性を把握
- 不正確・欠測値の解析上の取り扱いについては、臨床担当者及び統計担当者で、治験の早い段階で取扱いを取り決める
- 送信エラー等の ePRO デバイスとサーバーの同期不良の解決、バッテリーの問題

具体的な対応としては、治験開始後直後からのデータレビューを実施し、変動がまったくない被験者、極端な変動がある被験者、誤記・欠測が多い被験者、誤記・欠測が多い設問項目などを特定し、治験期間中に施設に対して注意喚起する。最終評価時点になって PRO データをまとめて見直すことはできないため、施設担当者の役割を明確にし、被験者に適宜確認・注意喚起してもらえようとする。即時性が期待できる ePRO を利用する場合でも、デバイス入力時点とデータレビュー間にタイムラグが生じる可能性があるため、治験中データレビューが逐次可能となるプロセスを事前に構築し、データの質を担保することが望ましい。

ePROデータは、ePROシステムを利用する場合に遵守すべきGCPの条項とER/ESの要求事項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しておくことが必要になる[40]。

4.4 統計解析留意点

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。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おかなければならない。ただし、特定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。

臨床試験では、さまざまな原因によって欠測のあるデータが生じ、臨床試験にバイアス

を起こす。PRO はデータの欠損が生じやすいので、それを可能な限り減らすための試験計画や、欠測値を取り扱うための統計手法が重要になる。FDA のガイドラインも欠測の問題についても言及しており、感度分析 (sensitivity analysis) の重要性などについて述べている [5]。PRO の欠測は治療効果そのもの (例えば効果を実感できない気がするがそれを先生に言うのは申し訳ない気がする、生活上注意するよう言われていたことを守れなくてその結果悪化した、など) と相関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。解析計画書では、有効性の評価を行う際にどのような統計手法を用いて欠測のあるデータを扱うか事前に計画しておくべきである。

おわりに

以上、臨床開発担当者を対象とし、臨床試験における PRO 尺度の使用実態を示すとともに、各規制当局の動き、PRO 尺度の準備プロセス、PRO 尺度を用いたデータ収集の留意点、特に被験者や施設担当者へのトレーニングの重要性について述べた。また、付録に示したアンケート結果から、各社、PRO 尺度を用いる際に試行錯誤を重ねながら進めている現状が浮き彫りとなった。

近年、日本においても「患者視点」の重要性がうたわれ、今後「患者視点」に立った医療の提供について議論する機会が増えることが期待されている。また欧米を中心に、治験の中だけでなく、治験の計画段階あるいは承認申請や薬価収載の段階で、どのように被験者の意見を取り入れていくのかといった「患者エンゲージメント」という概念が注目され、その中で ePRO の様な電子ツールの普及がそれを後押ししている。また、糖尿病診療において patient-centered approach といった治療を個別化する流れの中で「患者中心医療」という言葉が日本でも浸透してきている。その際、PRO 尺度は被験者の意見を聞く貴重なツールであり、被験者自身の治療への参加が進んでいく中で、PRO 尺度の利用頻度は一層増えるであろう。そのためにも、PRO 尺度を正しく設計して信頼性・妥当性を担保し、かつ、医療現場で正しく使われることが非常に大切である。

付録

2015年7月に「治験における Patient Reported Outcomes (PRO) に関する諸問題」について製薬協データサイエンス部会参加企業 67 社のデータサイエンス部門又は臨床試験の企画担当者を対象にアンケートを行った。アンケートには ePRO を含む PRO に関する自社での経験、日本が参加した国際共同治験での経験を含めて回答することとし、回答は 1 社 1 回答とした。アンケートは Web アンケートシステムを利用して実施し、67 社中 60 社から回答を得た。また、別に電子化情報部会の 2015 年度タスクフォース 3 で実施した「電子日誌 (ePRO) の普及状況」に関するアンケートについて結果を入手してまとめ、一部引用した [7]。アンケートは 2015 年 9 月に電子化情報部会会員会社を対象に Web アンケートシステムを利用して実施され、回答は 1 社 1 回答とし、30 社から得たものである。Table 1 にこれら 2 つのアンケートの概要を示す。

Table 1 2 つのアンケートの概要

アンケート	治験における Patient Reported Outcomes に関するアンケート	「電子日誌 (ePRO) の普及状況」に関するアンケート
調査目的	治験における Patient Reported Outcomes (PRO) に関する諸問題のまとめ	国内の ePRO の普及状況の調査
実施タスク フォース	データサイエンス部会 2015 年度タスクフォース 7	電子化情報部会 2015 年度タスクフォース 3
対象企業	データサイエンス部会 所属会社 67 社	電子化情報部会 所属会社 51 社
回答対象者	データサイエンス部門、臨床試験 の企画担当者	電子化情報部会所属会社の担当者
実施期間	2015/7/27~2015/8/14	2015/9/7~ 2015/9/18
回答方法	Web アンケートシステムを利用	Web アンケートシステムを利用

1) PRO 使用経験

回答を得た 60 社の内訳は、内資系 46 社、外資系 14 社であった。このうち、治験で PRO を使用した経験があった会社は 60 社中 44 社 (73%) で、内資系が 33 社、外資系が 11 社であった。PRO を使用した経験のなかった会社は 60 社中 16 社 (27%) で、内資系が 13 社、外資系が 3 社であった (Figure 1)。また、電子化情報部会のアンケートでは、ePRO 導入経験のある会社は約 1/3 であった (10/30 社)。

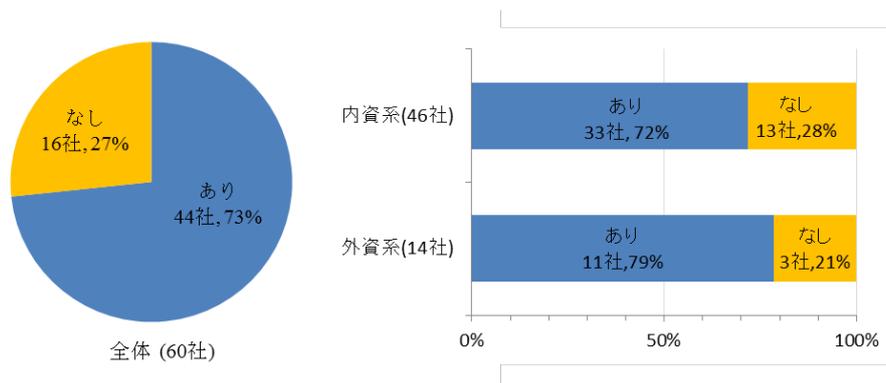


Figure 1 (質問2) 治験でPROを使用した経験がありますか

PROを使用した経験のあった44社に対しPROの使用目的を聞いたところ、ほぼすべての会社で「有効性評価項目」との回答であった (Figure 2)。「HTA 目的」は内資系で33社中3社 (9%) であったのに対し、外資系では11社中4社 (36%) と外資系のほうが割合が多かった。

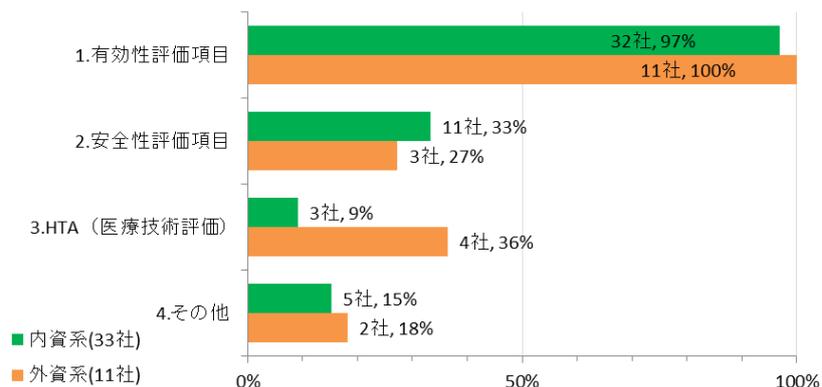


Figure 2 (質問4) PROの使用目的を選択してください (複数選択可)

また、PROをPrimary endpoint又はKey secondary endpointで使用した経験があると回答した会社は全体では43社中35社 (81%) であった (Figure 3)。内資系では27社 (84%)、外資系では8社 (73%) が使用経験があった。多くの会社でPROを臨床試験の主要なデータとして利用した経験があった。

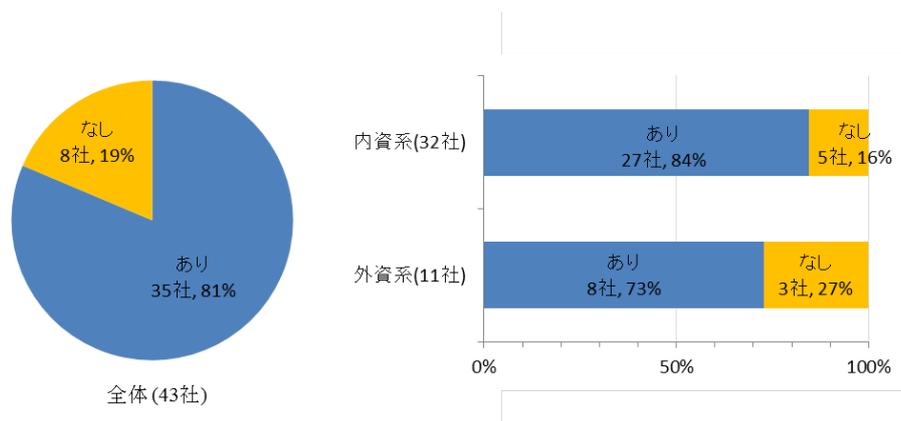


Figure 3 (質問 5) Primary endpoint または Key secondary endpoint で使用されたことがありますか？

一方、PRO の使用経験がないと回答した会社 16 社に、使用しなかった理由を自由記載で聞いたところ、PRO の活用が有用あるいは必要とする医薬品の開発を行っていないなど、必要なかったとの回答が多かった (Table 2)。PRO に対する懸念点として、バリデーションや費用、データへの信頼性などがあった。また、PRO を聞いたことがない、なじみがない、知識がないとの回答もあり、PRO に対する認識が会社によって異なっていることが伺えた。

Table 2 (質問 3) PRO を使用しなかった理由、及び懸念点を教えてください (抜粋)

- ✓ PRO を使用する医薬品の開発を行っていないため。
- ✓ PRO の活用が有用と思われる領域の試験を実施したことがないため。
- ✓ 各評価項目は医師による評価・判定を採用しているため。
- ✓ バリデーションや費用に対する懸念。
- ✓ データの信頼性への懸念。
- ✓ 費用対効果、実施手順の調査/知識不足のため。

(質問 2 で「なし」と回答した会社が対象、自由記載)

また、電子化情報部会のアンケート結果では、ePRO を導入しなかった理由として ePRO のベネフィットが見いだせない、コストが大きい、という理由が挙げられていた。実際、ePRO 利用経験のある会社は、Phase2 での構築費用 1000 万～3000 万円及び運用費用 1000 万～3000 万円に対してコスト感として高いと感じており、Phase3 でも構築、運用費用がそれぞれ 3000 万円以下なら概ねコスト感として「適正」だがその他は「高い」と感じていた。概して Phase2 のほうが Phase3 より入力期間は短く症例数が少なかったが、費用は Phase3 と同じぐらいであったところからコストが「高い」と感じていると思われる。

以降、「質問 2. 治験で PRO を使用した経験がありますか」で「あり」と回答のあった 44 社 (内資系 33 社、外資系 11 社) の回答について集計した結果を示す。

2) 社内体制

過去5年に遡って社内でPROに関する勉強会を開催したと回答のあった会社は44社中12社(27%)、なしと回答のあった会社は32社(73%)で、勉強会が開催されなかった会社が全体の3/4を占めた(Figure 3)。経験ありの会社は内資系では7社(21%)、外資系では5社(45%)で、外資系のほうが割合が高かった。

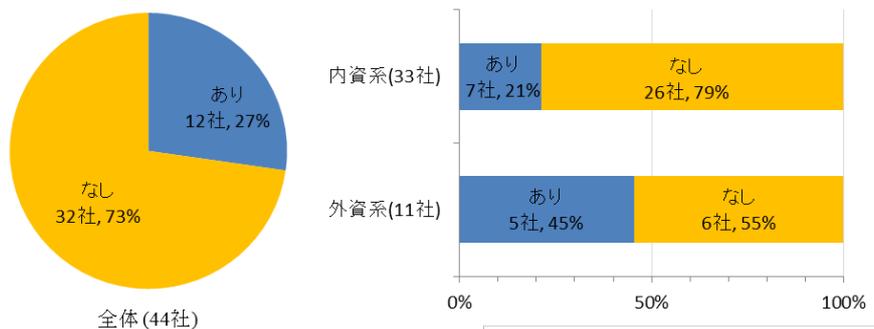


Figure 3 (質問6) 過去5年に遡って、社内でPROに関する勉強会が開催されましたか？

次に、社内にPRO専門の部署があるかを聞いたところ、「なし」と答えた会社が44社中36社(82%)であった(Figure 4)。「あり」と回答した8社(18%)で、海外に部署がある会社が8社、日本に部署がある会社が2社であった。内資系でPRO専門の部署があると回答した会社1社は海外に持っていた。外資系では11社中7社(64%)で海外に部署があり、そのうち2社は日本にも部署を持っていた。

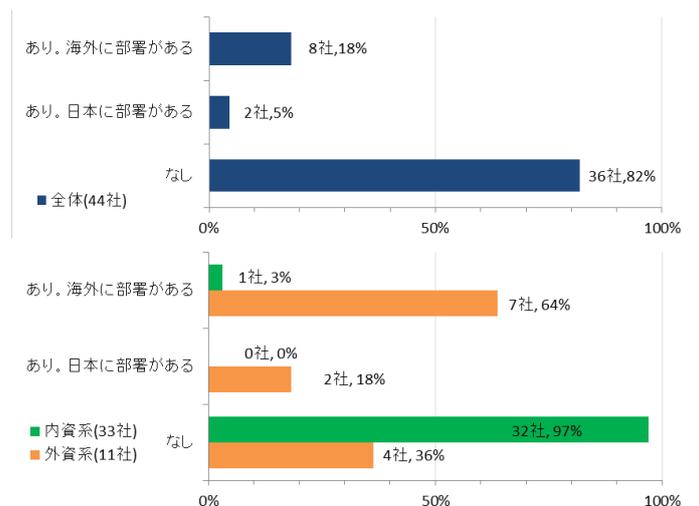


Figure 4 (質問7) PRO専門の部署がありますか？(複数選択可)

PRO を使用する際に DM・解析担当者がどのように関わるかについて、43 社中 31 社 (72%) が「開発担当者と協力して PRO 設定や資料レビュー時から」との回答であった (Figure 5)。「PRO 候補選定時から関わる」との回答が、内資系では 7 社 (21%) に対し外資系は 4 社 (40%)、「電子データ受領時の仕様書レビュー時から関わる」と回答が、内資系では 13 社 (39%) であるのに対し、外資系では 4 社 (40%) であった。外資系では DM・解析担当者が PRO 選定の初期からかかわりを持つ傾向があるようであった。

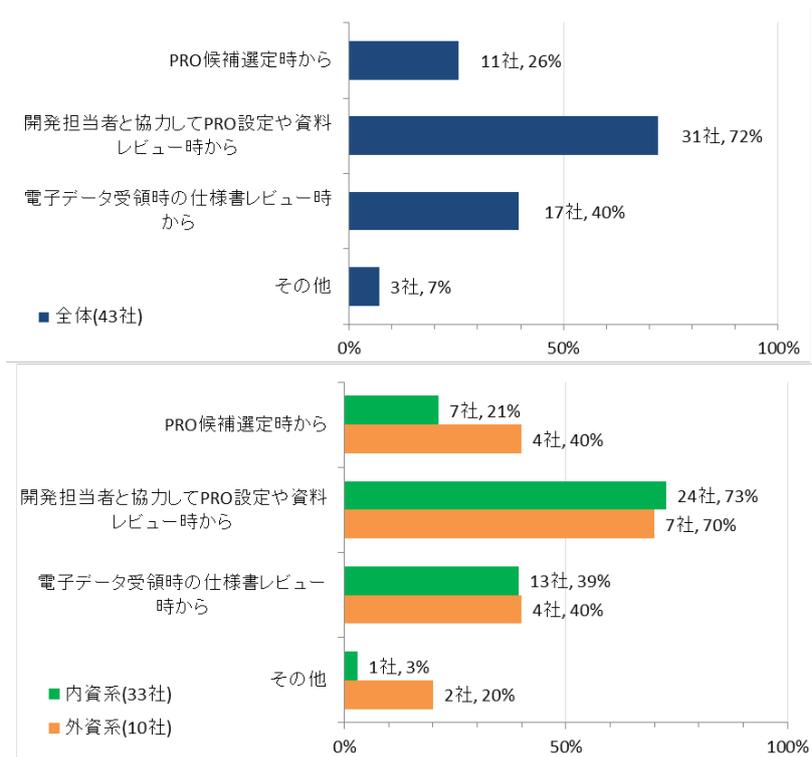


Figure 5 (質問 13) 治験で PRO を使用する際に、DM・解析担当者はどのように関わることになっていますか？ (複数選択可)

SOP やそれに準じた標準手順について確認した (Table 3, Figure 6)。アンケートで挙げた SOP またはそれに準じた標準手順をすべて揃えている会社はなく、専門部署を持っていることが多い外資系でも多くて 1 社当たり 3 種までであった。PRO の使用経験はあるが、多くの会社で SOP を所持していない現状が伺えた。一方でバリデーション手順やデータの取り扱いなどについて詳細に回答している会社も多く、必ずしも SOP 等で明文化していなくても、手順を踏んで PRO を使用しようとしている面も伺えた。

Table 3 アンケート内で聞いた SOP またはそれに準じた標準手順の種類

(質問 14)	PRO 作成に関する SOP や SOP に準じた標準手順
(質問 19)	言語バリデーションに関する SOP やそれに準じた標準手順

-
- (質問 23) PRO のトレーニングに関する SOP やそれに準じた標準手順
- (質問 26) PRO で特に取り扱いルールが定義されていない入力 (誤字脱字, 択一質問で複数回答された際の取り扱い等) に関する SOP やそれに準じた標準手順
- (質問 28) PRO で特に取り扱いルールが定義されていない欠測値, 不正值に関する SOP やそれに準じた標準手順
-

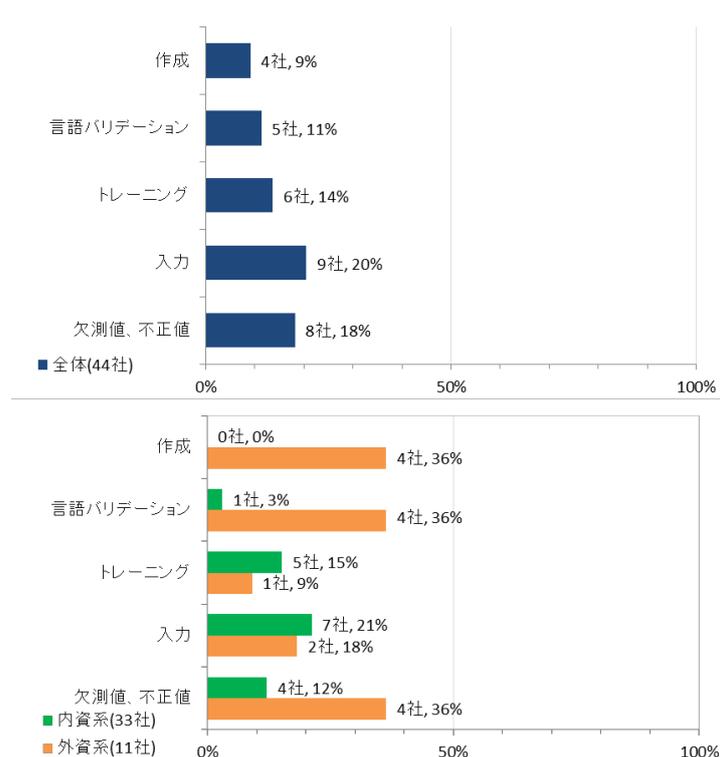


Figure 6 PRO に関する SOP またはそれに準じた標準手順の所持

3) 国際共同試験における PRO

国際共同試験で海外と同じ運用 (手順及びルール) で PRO を使用したかを聞いた (Figure 7). 国際共同試験での PRO の使用経験自体がない会社が 27 社 (61%) であり, 経験のあった 17 社のうち 16 社が海外と同じ運用で実施していた. 経験があるが海外と同じ運用で実施しなかった 1 社は, 「Primary endpoint ではなかったもので, 日本語でバリデーションされていない指標は日本人に限っては実施しなかった」とのことであった.

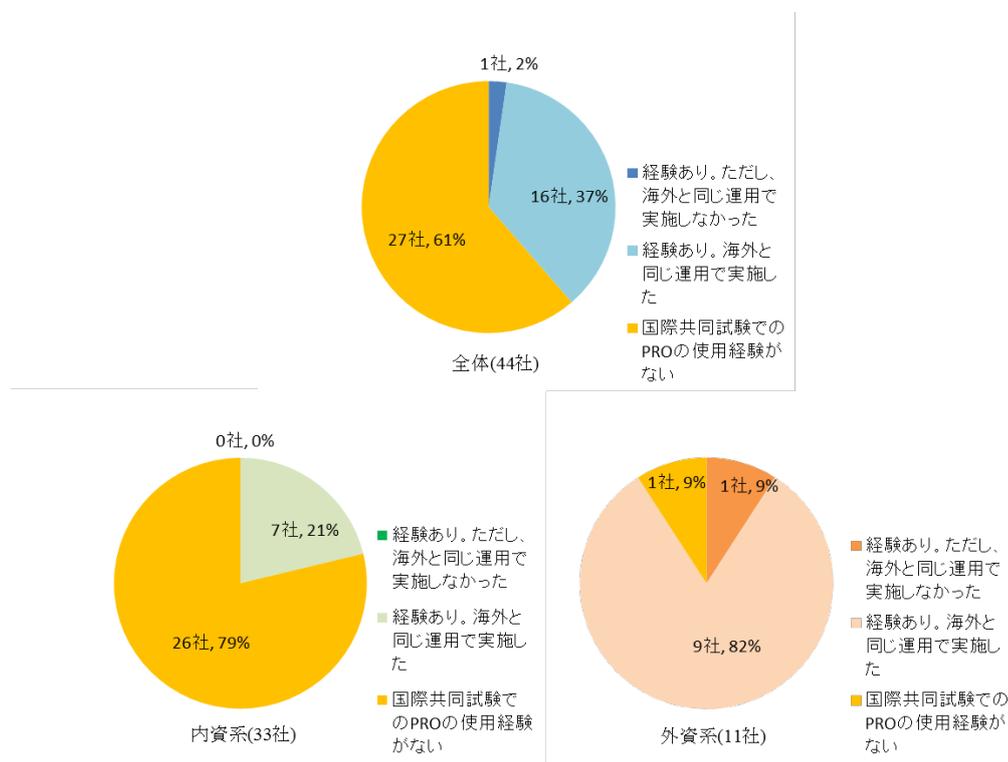


Figure 7 (質問 8) 国際共同試験で PRO を使用した経験がある場合、海外と同じ運用 (手順およびルール) で実施しましたか？

国際共同試験で日本の文化・環境にそぐわない質問項目を含むと感じられた PRO に対し「何か対応した経験がある」会社は 44 社中 5 社 (11%) であり、対応した経験のある会社は少なかった。対応した内容は、日本の文化・環境に合わせた翻訳になっているものを選択した、翻訳の際「フォーク」を「箸」と読み替えるなどであった (Table 4)。

Table 4 (質問 11) 日本の文化・環境にそぐわない質問項目を含むと感じられた PRO に対し、対応した経験

- ✓ 日本語の翻訳自体が日本の文化・環境に合わせた翻訳になっている PRO がある。
- ✓ 該当 PRO は特定できないが、言語バリデーション実施時に海外の専門業者と協議の上、文化に依存する英単語を適切な日本語に置換した。
- ✓ 通常利用されている薬剤名の変更等 (翻訳バリデーションを実施)
- ✓ 英語から日本語にする際に、下記のような読み替えを行った。
フォーク→お箸, yard→メートル

4) 既存 PRO の選定

PRO の選定の際に考慮される点として、多くの会社で「日本語版があるか」、「過去の利用頻度」、「言語バリデーションされているか」があがった (Figure 8)。外資系では「FDA で承認された試験で使用されているか」、「検出感度」を考慮している会社の割合が内資系に比べて多かった。その他の内容ではコスト・運用面での簡便性、学会や研究会が推奨しているか・業界で広く認知されているか、があがった。

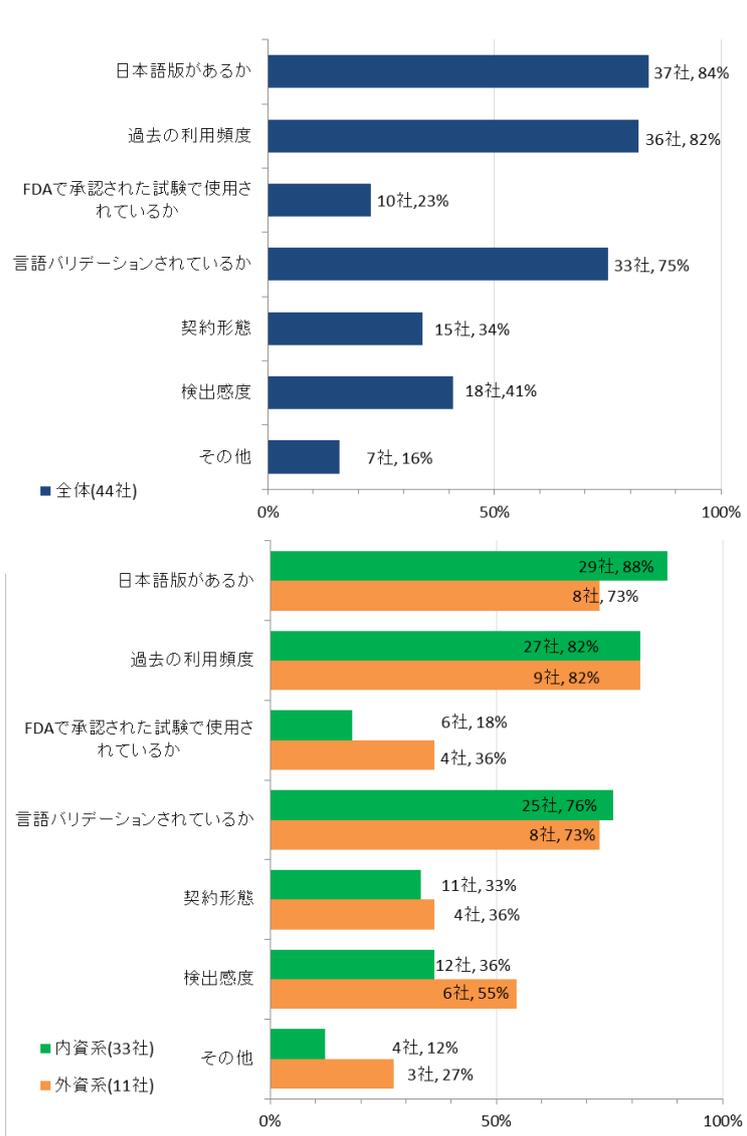


Figure 8 (質問 12) PRO の選定で考慮する点を教えてください (複数選択可)

5) 新規 PRO の開発

既存の PRO を使用せず新たに質問票・被験者日誌を作成した経験があると回答した会社

は44社中16社(36%)で、内資系が13社(39%)、外資系が3社(27%)で、内資系のほうが多かった(Figure 9)。PRO作成に関するSOPは持っていないが、作成経験のある会社が多かった。

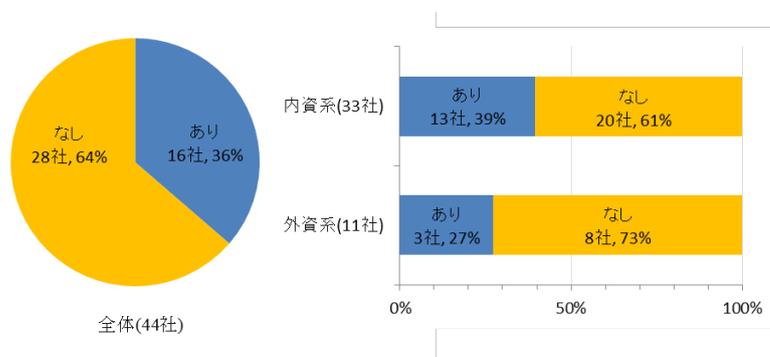


Figure 9 (質問 15) 既存の PRO を使用せず、新たに PRO (質問票・被験者日誌) を作成した経験がありますか?

また、PRO を作成した経験のある内資系 13 社、外資系 3 社に、PRO を作成した際は社内で作成したかを聞いたところ、9 社 (57%) が社内で作成していた (Figure 10)。

次いで、「医師と協力して作成した」が 5 社 (31%) で、PRO を作成した際は社内又は医師と協力して作成したが大部分を占めた。その他と回答した 1 社は全ての項目に当てはまるものであった。

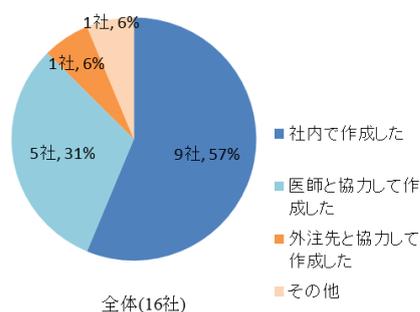


Figure 10 (質問 16) 新たに PRO 作成した際、社内で作成した経験がありますか?

新たに PRO を作成した際、バリデーションをどのように実施したかについて自由記載で聞いた。PRO 尺度に対するバリデーション後の言語バリデーションについてはいくつか回答があったが、専門医によるレビューやパイロット試験を行ったとの回答もあった一方、バリデーション試験は実施していないという回答も複数あった。

さらにこのバリデート結果について、公表 (パブリケーション) しなかった会社は回答のあった 15 社中 8 社で、うち内資系が 6 社、外資系が 2 社であった。内資系外資系ともに半数で公表しなかったとの回答であった。

6) 既存 PRO 尺度の言語バリデーション

他言語 PRO の日本語版を作成した経験が「あり」と回答したのは 44 社中 13 社 (30%) で、そのうち「言語バリデーションを実施した」が 10 社、「翻訳したのみ」が 3 社 (7%) であった (Figure 11)。内資系で「あり」と回答のあった会社は 33 社中 6 社 (18%) で、経験のない会社がほとんどであった。一方、外資系では「あり」が 11 社中 7 社 (64%) で、そのうち「言語バリデーションを実施した」が 5 社 (46%)、「翻訳したのみ」が 2 社 (18%) で、内資系と異なり経験のある会社が多かった。

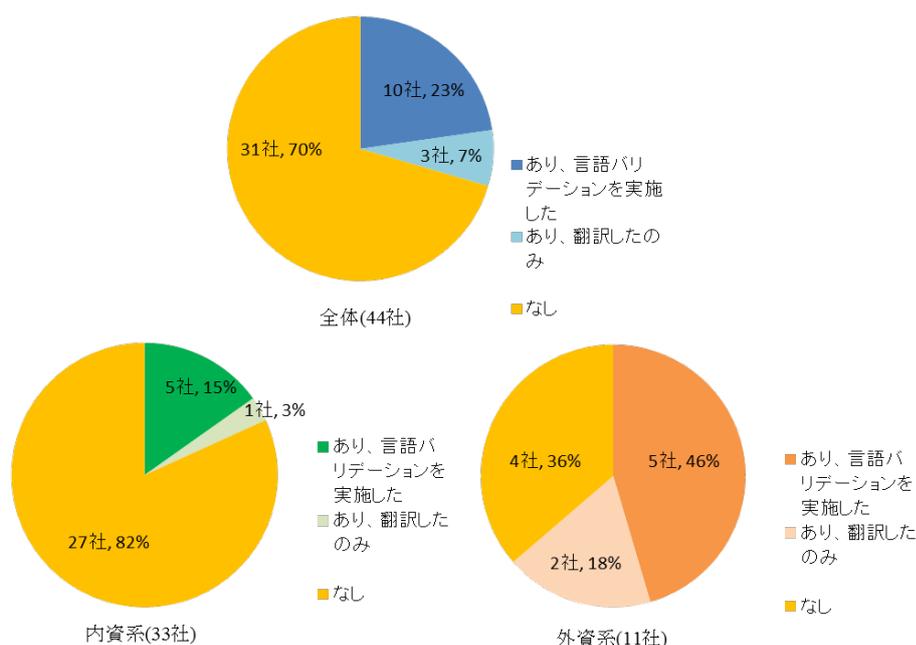


Figure 11 (質問 20) 他言語 PRO の日本語版を作成した経験がありますか？

さらに言語バリデーションを実施した経験があるかを聞いたところ、回答のあった 10 社中、「社内で実施した」が 1 社、「外注先に依頼した」が 8 社であった。内資系 5 社は全て外部委託先に依頼していた。外資系 5 社では、「社内で実施した」が 1 社、「外注先に依頼した」が 3 社であった。その他と回答した 1 社は外資系で、海外の専門業者に委託したとのことであった。また、成人用 PRO を小児へ利用した経験がある場合のバリデーションの方法を聞いた (質問 22)。43 中 37 社 (86%) で利用した経験がなく、経験のある 6 社 (内資系 3 社/外資系 3 社) はそのまま変更せず使用していた。

7) トレーニング

回答者及び施設担当者に実施したトレーニング方法の詳細を自由記載で聞いた (Table 5)。

CRF/調査票・入力の手引き・記入方法を施設用，被験者用にそれぞれ作成するとの回答が多かった。また，施設担当者に対してはマニュアルやトレーニング資料などを用いて説明する，被験者に対しては施設担当者を通じてトレーニングする，説明用のビデオを作成し配布する，などの回答があった。

Table 5 (質問 24) 回答者及び施設担当者へ実施したトレーニング方法を可能な限り詳細に教えてください

-
- ✓ **CRF/調査票 入力手引き，記入方法 の作成**
 - ・ PRO 自体に，記載の仕方のページを作成する。
 - ・ PRO 記載の手引きを施設用，被験者用それぞれ作成。・・・同様の回答が 17 件
 - ・ PRO の著作権元が作成した説明資料の配付。
 - ・ 施設用の PRO 手引きには被験者へのトレーニング方法，トレーニング達成の基準，未達の場合の手順等を記載。被験者用には PRO 記入方法，各症状の定義，FAQ などを記載。
 - ✓ **マニュアルやトレーニング資料などを用いた Investigator Meeting やトレーニングでの説明**
 - ・ 記載上の手引きの作成，Site training，Investigator Meeting 及び CRC training における説明（専門家・モニターからなど）。・・・同様の回答が 8 件
 - ✓ **施設担当者を通じて被験者へトレーニングを実施**
 - ・ 被験者に対しては，施設担当者がトレーニングを実施する。トレーニング後は署名をもらい，記録が残るようにしている。
 - ・ 被験者に対して，共通の資料を用いて施設担当者から説明。
 - ・ VAS の線の引き方について，記載要綱を日誌の見開きページに記載し，CRC を通じて指導するよう依頼した。
 - ✓ **その他**
 - ・ run-in 期間に回答者が練習する期間を設定。
 - ・ ePRO の際は電話にて業者がトレーニングを実施。
 - ・ ビデオ又はネットにおける映像によるトレーニング。
 - ・ Q&A 集，ビデオを日本語で作成。
 - ・ 被験者には ePRO training モジュールを提供。
 - ・ 個々の PRO の記載方法については，Protocol に詳細を記載しかつ ePRO 画面で説明しているので特に資料などは作成していない。
-

トレーニング資料を日本語訳することになっているかを聞いた (Figure 12)。外資系ではオリジナルが日本語でないため，社内又は外部委託で日本語訳の作業を実施しているよう

であるが、英語のまま使用することもあると回答した会社が4社（36%）あった。

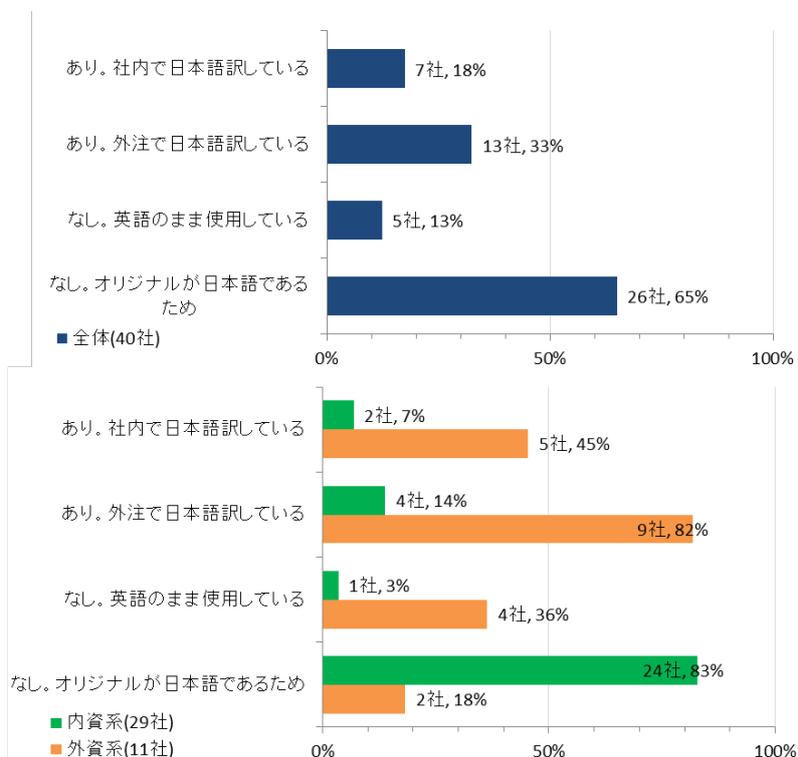


Figure 12 (質問 25) トレーニング資料は日本語訳することになっていますか？（複数選択可）

また、電子化情報部会のアンケート結果では、ePRO 自体の説明資料について、オリジナルが日本語であるか、またはオリジナルが英語の場合は日本語に訳しており、英語のまま使用しているという回答はなかった。

8) 実施上の工夫

被験者が PRO を施設で記入する際、医師や CRC が被験者の回答に対して介入しないために何か対応した経験があるかを聞いた (Figure 13)。「トレーニングで伝えている」が最も多かった一方、「特に対策していない」との回答も次いで多かった。特に内資系では半数以上 (17 社, 52%) が特に対策をしていないようだった。その他の詳細として、被験者が記入後、封筒に入れ封をして持参してもらい、被験者から直接データ入力業者に郵送してもらい、医師の評価前に記入してもらい、Visit 時ではなく、その日来院前に自宅で記入して持ってきてもらう、プロトコルで医師の評価前に実施するよう規定している、などのコメントがあった。

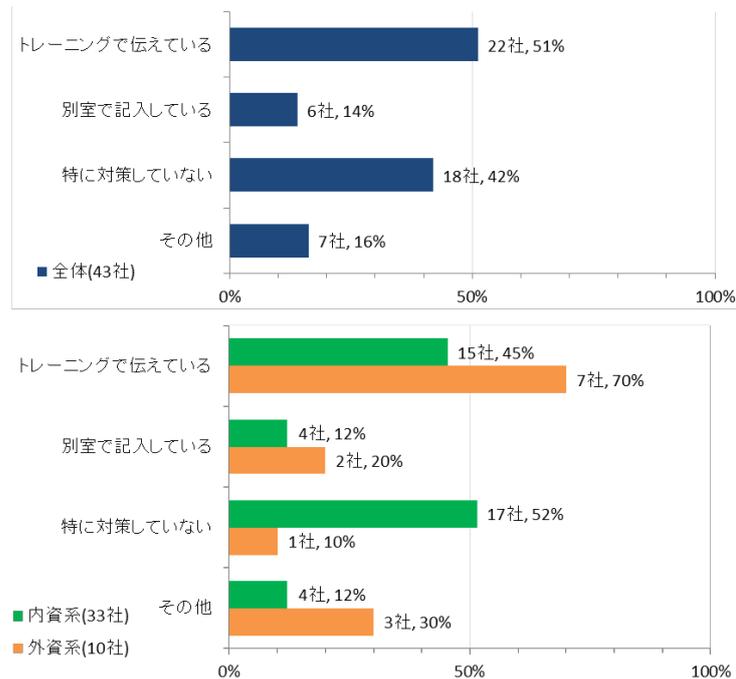


Figure 13 (質問 27) 被験者が PRO を施設で記入する際、医師や CRC が被験者の回答に対して介入しないために何か対応した経験がありますか？（複数選択可）

欠測値、不正値の取扱いルールについて自由記載で聞いた。欠測値や不正値があった時に何らかの数値を補う、欠測扱いとするなど、さまざまであった (Table 6)。

Table 6 (質問 29) 支障のない範囲で欠測値、不正値の取り扱いルールについて教えてください

- ✓ 欠測値は感度分析を実施する。部分的な欠測な場合は他の平均値を入れることが多い。ある時期すべてが欠測の場合は感度分析を実施する。不正値は基本的に利用しないが、探索的な試験の場合は不正の状況を見極めて、利用する場合もある。
- ✓ 解析計画書で記載した。・・・同様の回答が 2 件
- ✓ 不正値の記載方法により、個別検討。
- ✓ PRO によりさまざまなので一概には言えないが、プロジェクトや領域別に社内ですタンドを作成し、解析計画書で定義している。例えば明らかに入力ミスやエラーと思われるものは異常値として欠測扱いとする、など。
- ✓ 各質問に 2 つ以上のチェックがある場合、または空欄の場合は invalid な記載として扱う。1 日の日誌に 5 つ以上、invalid な記載がある場合は、その日

の日誌自体を **invalid** な日誌として扱う。

- ✓ 欠測については、欠測. 不正値については、それぞれ、欠測や中央値を入れる等のルールを作成。
 - ✓ 各試験の目的に応じたルールであり、一律のルールがあるわけではない。
 - ✓ **PRO** で特に取り扱いルールが定義されていない欠測値については、解析時に複数の取り扱いを行う場合がある (**Key secondary** など重要な **PRO** の場合)。被験者の入力したデータは原則修正しない。その他、システム上の不正値等は、別途手順書等で取り扱いを規定。 **PRO** を **DM** で入力した試験は、医療機関への問合せを行う (判読困難等)。また、別途手順 (評価として不利な方向 (重症な方) を採用等) に従い入力。
 - ✓ ある領域で出血に関する欠測値の取り扱いについて、ある一定の期間で欠測値が 2 つあった場合、補完できるというルールがある。一定の期間の定義は各 **project / study** で決定することになっている。
-

引用文献

1. Drummond, M.F., M.J. Sculpher, and K. Claxton, Methods for the Economic Evaluation of Health Care Programmes, Fourth Edition. Oxford Medical Publications, 2015.
2. 坂巻弘之, やさしく学ぶ薬剤経済学. じほう, 2003.
3. European Medicines Agency, Reflection Paper on the Regulatory Guidance for the Use of Health Related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products [http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf]. 2005.
4. World Health Organization, Constitution in basic documents. 1948.
5. Food and Drug Administration, Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009.
6. Food and Drug Administration, 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. 2001.
7. 日本製薬工業協会 医薬品評価委員会 電子化情報部会 タスクフォース 3, ePRO の普及に向けて [<http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/pdf/eipro.pdf>]. 2016 年 3 月.
8. 厚生省医薬品安全局審査管理課長通知 (医薬審第 1047 号), 臨床試験のための統計的原則 [<https://www.pmda.go.jp/int-activities/int-harmony/ich/0031.html>] 平成 10 年 (1998 年) 11 月 30 日. 1998.
9. Vodicka, E., et al., Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). Contemp Clin Trials, 2015. **43**: p. 1-9.
10. DIA/FDA Statistics 2016 Forum [<http://www.diaglobal.org/ja-jp/conference-listing/meetings/2016/04/dia-fda-statistics-2016-forum>]. 2016.
11. Critical Path institute, The Patient-Reported Outcome Consortium [<https://c-path.org/programs/pro/>].
12. European Medicines Agency, Qualification of novel methodologies for drug development: guidance to applicants. Revision 3. 2014.
13. European Medicines Agency, Guideline on the evaluation of anticancer medicinal products in man [http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137128.pdf]. 2012.
14. European Medicines Agency, The use of patient-reported outcome (PRO) measures in oncology studies

- [\[http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf\]](http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf)
]. 2016.
15. 各都道府県衛生主管部（局）長あて厚生労働省医薬食品局審査管理課長通知, 「抗癌性腫瘍薬の臨床評価方法に関するガイドライン」の改訂について 薬食審査発第1101001号. 平成17年11月1日.
 16. Food and Drug Administration, Webinar: Background on FDA and Patient-Focused Drug Development
[\[http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm349133.htm\]](http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm349133.htm). 2016.
 17. 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, 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 [\[http://ispor-jp.org/pdf/kankou/kho.pdf\]](http://ispor-jp.org/pdf/kankou/kho.pdf). 2011.
 18. Guyatt, G.H., et al., A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis*, 1987. **40**(12): p. 1129-33.
 19. Coles, T., et al., Current sample size practices in the psychometric evaluation of patient-reported outcome measures for use in clinical trials.
[\[https://www.rtihs.org/sites/default/files/T.Coles_ISPOR.17_Poster.pdf\]](https://www.rtihs.org/sites/default/files/T.Coles_ISPOR.17_Poster.pdf) Poster presented at the 17th Annual European Congress of the International Society for Pharmacoeconomics and Outcomes Research., 2014.
 20. Fukuhara, S., et al., Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. *J Clin Epidemiol*, 1998. **51**(11): p. 1037-44.
 21. Fukuhara, S., et al., Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. *J Clin Epidemiol*, 1998. **51**(11): p. 1045-53.
 22. Cronbach, L., Coefficient alpha and the internal structure of tests. . *Psychometrika*, 1951. **16**: p. 297-334.
 23. Shrout, P.E. and J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*, 1979. **86**(2): p. 420-8.
 24. Cappelleri, J., et al., Patient-Reported Outcomes: Measurement, Implementation and Interpretation. CRC press, 2013.
 25. Landis, J.R. and G.G. Koch, The measurement of observer agreement for categorical data. *Biometrics*, 1977. **33**(1): p. 159-74.
 26. Westland, J.C., Structural Equation Modeling: From Paths to Networks. New York: Springer. 2015.
 27. 平井明代, 教育・心理系研究のための データ分析入門 東京図書.
 28. Tabachnick, B.G. and L.S. Fidell, Using multivariate statistics. Boston: Pearson Education. 2007.

29. 狩野裕, 三浦麻子, AMOS, EQS, CALIS によるグラフィカル多変量解析 : 目で見ると分散構造分析. 現代数学社, 増補版 3 刷. 2007.
30. Carter, S.R., et al., The impact of patients' perceptions of the listening skills of the pharmacist on their willingness to re-use Home Medicines Reviews: a structural equation model. *Res Social Adm Pharm*, 2015. **11**(2): p. 163-75.
31. Campbell, D.T. and D.W. Fiske, Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull*, 1959. **56**(2): p. 81-105.
32. Trochim, W., The Multitrait-Multimethod Matrix [<http://www.socialresearchmethods.net/kb/mtmmmat.php>]. 2006.
33. Nagaraja, V., et al., Construct validity of the Patient-Reported Outcomes Measurement Information System gastrointestinal symptom scales in systemic sclerosis. *Arthritis Care Res (Hoboken)*, 2014. **66**(11): p. 1725-30.
34. Jacobs, K.E. and J. Roodenburg, The development and validation of the Self-Report Measure of Cognitive Abilities: A multitrait-multimethod study. *Intelligence*, 2014. **42**: p. 5-21.
35. Lord, F.M., *Application of Item Response Theory to Practical Testing Problems*, Lawrence Erlbaum Associates. 1980.
36. Nguyen, T.H., et al., An introduction to item response theory for patient-reported outcome measurement. *Patient*, 2014. **7**(1): p. 23-35.
37. Hsueh, I.P., et al., Development of a computerized adaptive test for assessing activities of daily living in outpatients with stroke. *Phys Ther*, 2013. **93**(5): p. 681-93.
38. Wild, D., 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): p. 94-104.
39. Fairclough, D., *Design and Analysis of Quality of Life Studies in Clinical Trials*, 2nd edition. CRS, 2010
40. 日本製薬工業協会 医薬品評価委員会, 臨床試験データの電子的取得に関するガイドライン追補. 2012 年 1 月 10 日.

資料作成者

日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2015年度タスクフォース7

阿達 則昭 MSD 株式会社 (2016年3月まで)
東 理恵子 あすか製薬株式会社
奥 玲子 帝人ファーマ株式会社
高田 順美 バイエル薬品株式会社

タスクフォースリーダー兼推進委員

河田 祐一 中外製薬株式会社
木村 友美 アステラス製薬株式会社

担当監事

三沢 秀敏 ファイザー株式会社