政策研のページ

Patient Reported Outcomeの臨床適用と課題

医薬品の臨床評価において、従来の有効性（Efficacy）、安全性（Safety）に加えて、臨床成果（Outcome）を評価する指標Effectiveness（有用性）が注目されるようになっています。Effectivenessという概念は、日本で以前行われていた「有用度」と似ていますが、単なる印象に基づく「主観」ではなく、評価尺度の信頼性・妥当性を評価することで、試験成績に普遍性・一般性を与えています。中でも、医師を介さず患者自身が治療効果を評価するPatient Reported Outcome（PRO）が、近年、臨床試験で新薬評価基準として重要視されています。この背景には、患者参加型医療といった納得の医療の理解に加え、新薬の患者に立脚した治療上の意義を明確にすることで、そのポジショニングに寄与することが期待されています。

臨床試験登録データベースを用いたPRO利用状況調査

「患者による直接評価」に注目したプロトコールがどの程度あるかを調査するため、臨床試験の米国登録データベースであるClinicalTrials.govを用い、企業主体の第二相あるいは第三相試験に絞り込んで、PRO関連用語を抽出して検索[1]を行いました。絞り込まれた4,351試験のうち、各試験の臨床評価方法の記載内容を確認した結果、2,033試験を何らかの形でPRO評価を行っている臨床試験と判断しました。そのうち、PRO評価を主要評価項目に採用していた試験は582試験、副次評価項目のみで採用していた試験が1,447試験、4試験はその他のエンドポイントとして利用されていました。該当試験を疾患別に集計した結果を**図1**に示します。

「痛み」に関する評価は、以前から患者の主観に頼らざるを得ない分野であり、患者による直接評価が頻繁に用いられていることが今回の結果からも確認できました。また、「悪性腫瘍」に関しては、生活の

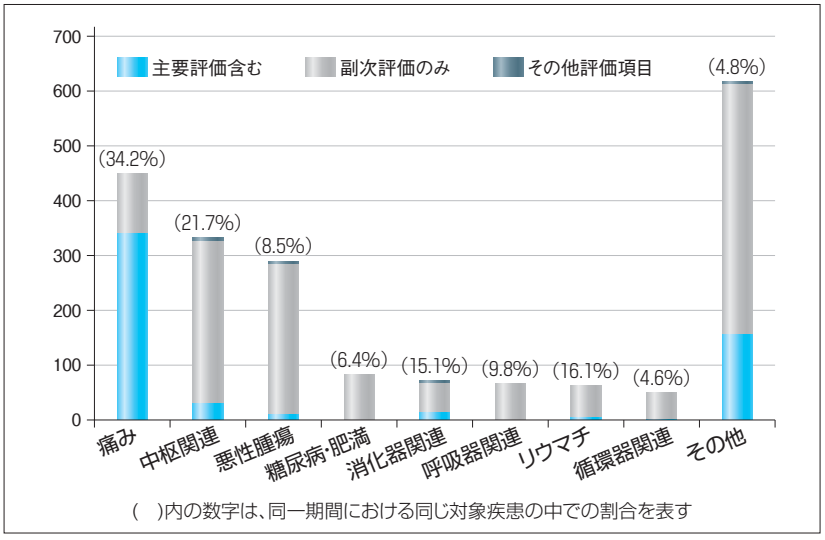


図1　PRO関連評価項目を有する試験数の内訳

1）検索条件：PRO関連用語として、“Patient Reported”“Patient self-reported”“ Patient Handling Questionnaire”“Patient satisfaction”“Patient's global evaluation”を含む試験を検索、さらにPatientの部分をSubjectおよびParticipantに変更したものも同様に検索、これに“Satisfaction Questionnaire”といった特定用語を追加検索し、重複した試験を整理してリストとした。対象期間は2005年1月～2012年8月。

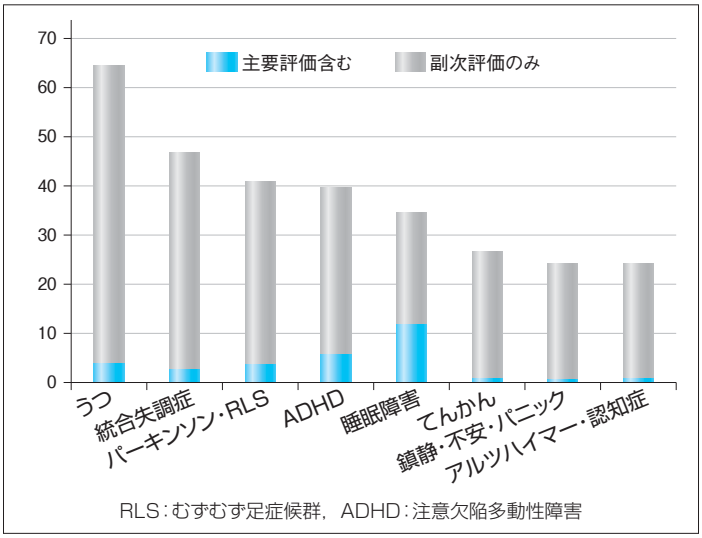


図2　中枢分野内の内訳

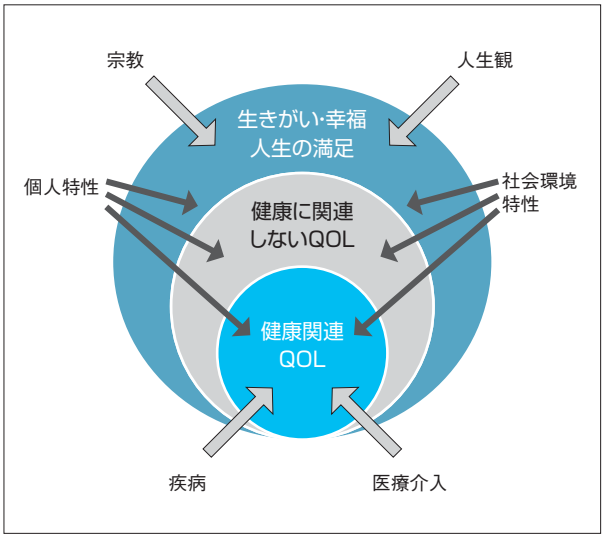


図3　QOL（広義）と健康関連QOLの関係概念図

質（Quality of Life）が重要であることが近年話題となっており、多くの試験でPRO尺度が利用されていましたが、そのほとんどは、副次評価でした。

「中枢関連」としてヒットした333件の疾患別内訳を**図2**に示します。中枢には、「睡眠障害」のように従来から患者の訴えが直接評価されている疾患もありますが、「うつ」「統合失調症」などのように、これまでは医師による評価が主流を占めてきましたが、この領域においても、近年、PRO評価を積極的に取り入れる傾向がみられます。

健康関連QOL（HRQOL）評価のためのPRO[2)]

QOLを広義に捉えると、患者本人の健康状態を表す健康関連QOLのための評価のほかに、生きがいなどを含めたさまざまな価値の評価を対象とすることができます（**図3**）。この中で、健康関連QOLは医薬品の臨床評価における重要なアウトカムであり、QOLの指標としてのPROは、アウトカムを患者自身が医師などほかの人の意見に左右されず算定するもので、患者中心医療の原動力となっています。

健康関連QOL評価は、一般的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ための「疾患・症状特異的尺度」に大きく分類することができ、さらに包括的尺度は、「プロファイル型尺度」と「選好による尺度（Preference-based measure）」に分類されます（**図4**）。

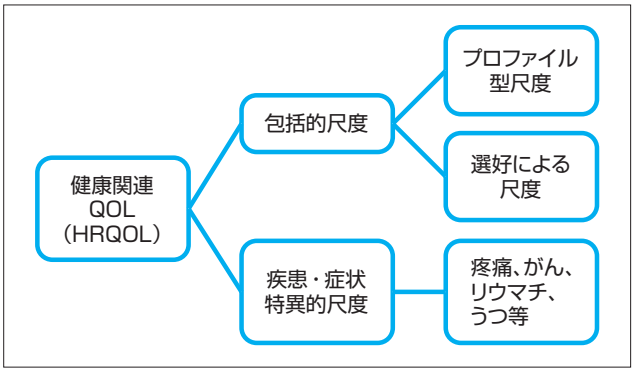


図4　健康関連QOLの分類

「包括的尺度」は、文字通り健康関連QOLを包括的に測定するためのツールであり、被験者の一般健康状態を知ることができます。この尺度を使えば患者から健康な人まで連続的に測定でき、疾患が異なっていても比較が可能になるという特長があります。包括的尺度の中の「プロファイル型尺度」は、QOLの構成要素を、身体機能，メンタルヘルスというように多次元に分けて評価する尺度であり、SF-36、Sickness Impact Profile、WHO-QOL などの尺度があります。「選好による尺度」は、健康状態全般について価値付けを行い、一つの数字（効用値）で表す評価尺度であり、EQ-5D, Health Utility Index, Quality of Well-Being Scaleなどが代表例です。最近、医療経済評価としてしばしば議論に上るQALY（Quality Adjusted Life Year）は、生存年（LY）を生活の質（Quality）で補正した数値ですが、QALY算定の際に用いられる包括的な生活の質は一つの数値により表す必要があるため、「効用値」を求めることが

2）下妻晃二郎他　「臨床のためのQOLハンドブック」　医学書院

できる「選好による尺度」が活用されます。

プロファイル型尺度と選好による尺度の代表例であるSF-36とEQ-5Dについて、それぞれの構成、特長などをまとめ、**表1**[3~7]に示しました。

一方、「疾患・症状特異的尺度」は、個別の疾患での患者の疾患特異的な訴えや随伴する症状に焦点をあてたPRO尺度のことをいいます。一般に「疾患特異的尺度」のほうが「包括的尺度」に比して情報量が多く、感度も高いため、特定の疾患に対する臨床評価には、「疾患特異的尺度」が多く利用されます。また、希少疾患を始めとする一部の疾患には標準的な評価基準がなく、新たに評価基準を開発する必要があることや、社会的な流れとして医師の主観的評価を患者主体の観点から見直す動きがみられることから、「疾患特異的PRO尺度」の開発・標準化が臨床評価の場で重要視されるようになっています。

表1　SF-36（プロファイル型尺度）とEQ-5D（選好による尺度）

SF-36[3, 4]	**開発の経緯**	SF-36は、1986年より実施されたMedical Outcome Studyに伴って作成された米国発祥の尺度である。1990年から「国際QOL研究プロジェクト」の対象として検討され、日本はその7カ国目として参加した。このプロジェクトにより、国際的な標準化が進み、多国籍臨床試験での評価項目として一段と汎用されるようになった。
	構　成	SF-36に基づく結果は、以下の8つの尺度で構成され、これらの尺度と2つの因子（身体的側面、精神的側面）の相関から、身体的評価スコアと精神的評価スコアで表される。 1. 身体機能：歩行、着替え、入浴など 2. 日常役割機能(身体)：仕事や日常業務の身体的問題 3. 体の痛み：痛みの程度やその障害 4. 全体的健康感：健康状態の評価 5. 活力：活力や疲れなど 6. 社会生活機能：家族や第三者とのつきあい 7. 日常役割機能(精神)：仕事や日常業務の心理的問題 8. 心の健康：神経質やゆううつ感など
	特長等	近年、質問項目を改訂したVersion 2.0が広く用いられており、更に短縮版であるSF-12, SF-8も開発されるなど尺度として発展してきた。SF-36及びSF-8については、日本も含め国際的に「国民標準値」が求められており、これらの数値と比較し健康状態を評価できる。更に、後述するEQ-5Dに見られるような選好による尺度、すなわち「効用値」に変換する方法も検討されており、単なるプロファイル型を超えた利用も期待できる。
EQ-5D[5~7]	**開発の経緯**	EQ-5Dは、欧州の研究機関グループ(EuroQOL Group)により5カ国語同時に開発され、1990年に発表された。その後、各国版が作られ、日本語版は日本語版EuroQoL開発委員会により1997年に認定された。
	構　成	調査票は 以下の5項目について3つのレベル（あてはまる、いくらかあてはまる、あてはまらない、など）で評価する領域と、健康状態全般をVAS（Visual Analogue Scale)により自己評価する2つの領域に分かれる。 1. 移動の程度　2. 身の回りの管理　3. ふだんの生活 4. 痛み・不快感　5. 不安・ふさぎ込み 5項目の質問に対する回答の組み合わせにより一つのスコア(効用値)が算出される。スコアは、1が健康、0が死を表す。スコア算出の際には、各国で作成・検討された「換算表」があり、日本語版換算表もEuroQOL本部にて妥当性が認められている。
	特長等	EQ-5Dは、医療経済効果を算出するため欧州で汎用されるQALY（質調整生存年）を算出するため、最も利用されている評価尺度であり、実際、2005年から2008年の間、英国NICEの医療経済評価で使用されたQOL評価尺度の49%がEQ-5Dであった。簡易な方法であること、国際的協力が得られていることから汎用されているが、243(35)通りで健康状態を完全に数値化することには無理があり（最近では5つのレベルの変更が検討されている)、臨床評価指標としては、EQ-5Dのみで絶対的な評価を下すよりも、他の健康関連QOL評価と組み合わせて総合的に評価することが重要である。

3）Ware JE, et al., Med Care：30：473-83, 1992
4）福原俊一 他、SF-36v2日本語版マニュアル：健康医療評価研究機構, 2004（iHopeホームページ参照）
5）EuroQol Group, Health Policy, 16, 199-208, 1990
6）日本語版EuroQol開発委員会, 医療と社会, 8, 109-123, 1998
7）Tosh, et al., Value in Health, 14(1) 102-109, 2011

課題：医師評価と患者評価の乖離

医薬品開発において、従来、多くの疾患で医師による評価が主流となっていました。しかし、近年徐々に、医師評価に代わり患者による直接評価が積極的に行われるようになっている一因には、症状に対する医師評価と患者評価の乖離があると考えられます[8]。リウマチ分野を例にとると、医師と患者による評価の乖離が報告され、その原因として、患者は疼痛レベルを重視するのに対し、医師は腫脹関節数をより重視する傾向がみられ、医師が患者の痛みを過小評価していると考察されています[9]。また、抗リウマチ薬増量と医師、患者それぞれの評価との相関を調査したところ、医師評価よりも患者評価とより相関が高く、増量は患者の訴えに依存していると報告されています[10]。

このような医師評価と患者評価との乖離は、疾患領域により異なりますが、今後のPRO評価重視にますます拍車をかけるものと予想されます。

課題：PROスコアと臨床的に意義のある差

臨床試験において、算定されたスコアに対し、どの程度の差を「臨床的に意義のある差」とするかがしばしば話題となりますが、このことは、PRO評価においても例外ではありません。数値の差が単なる統計処理上のものではなく、臨床的意義と連動することが重要となります。こういった点に対し、米国FDAは、スコアが臨床的意義を示す閾値を重視しており、最終的にプロトコールを合意する際、特にResponderの定義、およびClinical Benefitを見極めるための基礎的情報の収集と検討に大きな注意を向けています。たとえば、ぜんそくやCOPD（慢性閉塞性肺疾患）の呼吸器症状や症状に伴う活動や生活状態を評価するために国際的に汎用されるSGRQという尺度がありますが、この尺度は開発当初より、臨床的意義のあるスコアの最小変化が議論され、現在、0～100で示されるトータルスコアの中で、臨床的に意義のある最小変化は「4」が国際的なコンセンサスとなっています[11]。このような考え方はガイダンス作成時も同様で、2010年10月に出されたFDAガイダンス（案）：Qualification Process for Drug Development Tools（医薬品開発における新規バイオマーカーと新規PRO開発の際の留意事項を示したFDA Draft Guidance）、2012年5月の過敏性腸症候群(IBS)の臨床評価ガイダンス、2007年11月のCOPD臨床評価ガイダンス(案) などで、評価スコアと臨床的意義との関連性について言及されています。なお、FDAはこれらガイダンスの中で、言語や文化の違いによるスコアへの影響についても留意を促している点は注目に値します。

まとめと考察

今回、臨床試験におけるPROの使用実態を示すとともに、健康関連QOLを評価するため実際に使われているPRO尺度とその課題について解説しました。PRO尺度の中でも、治療の現場で汎用される評価尺度と臨床試験に利用される評価尺度は必ずしも同じではなく、特に、新薬のグローバル開発に際して、共通のPRO尺度を用いるためには、臨床的意義や翻訳版を含めた標準化が大きなカギとなります。新たなPRO尺度が開発され、信頼性・妥当性を確立し、言語や文化を超えて国際基準として臨床試験で汎用されるためには、長い年月を要します。現在、国際的に汎用されているPRO尺度はほとんど海外で開発されており、それらの中には、韓国語では翻訳版が存在するにもかかわらず、日本語版がないといった尺度も複数存在しています。このような状況から、国際臨床試験への日本の参画が不利になっていくことが危惧されます。

近年、Patient-focused Drug Developmentが注目され、既存PRO尺度の国際標準化や新規PRO尺度の開発に欧米では官民ともに力を注いでいます。この流れにより、新しいPRO評価の適用や普及が加速度的に進むものと予想されます。このような中で、汎用されるPRO尺度で日本語版が存在しないものについては、国際的に受け入れられる日本語版作成を急ぐ必要があることはいうまでもありませんが、今後、臨床開発の分野で日本がイニシアティブを取るには、さらに一歩進んで、日本発のPRO尺度を開発し、それを国際標準化していくという状況を作り出すことが、今後の日本にとって重要であると考えています。

（医薬産業政策研究所　首席研究員　小林 和道）

8）Litwin et al, J Urol. 59(6), 1988-92. 1998
9）Studenic P. et al, Arthritis Rheum. 64(9), 2814-23, 2012
10）Dougados et al, Rheumatology 52(2), 391-399, 2013
11）西村浩一、アレルギー科, 20(1), 47-53, 2005