

ISPORタスクフォース報告書 臨床アウトカム評価: 概念的基盤 —ISPOR臨床アウトカム評価報告書— アウトカム研究のための新たな実施基準タスクフォース

ISPOR TASK FORCE REPORT Clinical Outcome Assessments: Conceptual Foundation — Report of the ISPOR Clinical Outcomes Assessment — Emerging Good Practices for Outcomes Research Task Force

Marc K. Walton, MD, PhD^{1,*}, John H. Powers III, MD, FACP, FIDSA², Jeremy Hobart, PhD, FRCP³,
Donald Patrick, PhD, MSPH⁴, Patrick Marquis, MD, MBA⁵, Spiros Vamvakas, MD⁶, Maria Isaac, MAsc, MD, PhD⁶,
Elizabeth Molsen, RN⁷, Stefan Cano, PhD, CPsychol, AFBPsS⁸, Laurie B. Burke, RPh, MPH^{9,10}

訳 山口拓洋 (東北大学大学院医学系研究科 医学統計学分野)
川口 崇 (東京薬科大学 医療実務薬学教室)
宮路天平 (東京大学大学院医学系研究科 臨床試験データ管理学講座 /
国立がん研究センター中央病院 支持療法開発部門)
全田貞幹 (国立がん研究センター東病院 臨床研究支援部門 支持・緩和研究開発支援室 /
国立がん研究センター中央病院 支持療法開発部門)

抄 録

エンドポイントに用いられる患者評価であるアウトカム評価は、評価またはスコア（カテゴリ変数または連続変数）を提供する測定ツールであり、患者の健康状態のある側面を表している。アウトカム評価は、ある疾患または状態に対する治療法を開発するときの有効性のエンドポイントを定義するために用いられる。大部分の有効性のエンドポイントは、患者の特定の臨床評価に基づいている。臨床評価が臨床試験のアウトカムとして用いられる場合は、臨床アウトカム評価

(Clinical Outcome Assessments: COA) と呼ばれる。COA には選択、判断または動機に影響を受ける可能性のある評価が含まれる。COA は明確に定義され、治療のベネフィットを（直接的または間接的に）立証する十分な測定特性を保持していなければならない。対照的に、バイオマーカー評価は、患者の動機や評価者の判断にはほとんど影響を受けない評価である。本報告書は「ISPOR (The Professional Society for Health Economics and Outcomes Research) 臨床アウトカム評価報告書—アウトカム研究のための新たな実施基準タスクフォース」と題する2つの報告書のうちの1

¹Janssen Research and Development, Titusville, NJ, USA; ²Leidos Biomedical Research in support of the Division of Clinical Research, National Institutes of Health, Bethesda, MD, USA; ³Plymouth University Peninsula Schools of Medicine and Dentistry, Devon, UK; ⁴Seattle Quality of Life Group, Department of Health Services, University of Washington, Seattle, WA, USA; ⁵Modus Outcomes, Newton, MA, USA; ⁶European Medicines Agency, London, UK; ⁷International Society for Pharmacoeconomics and Outcomes Research, Lawrenceville, NJ, USA; ⁸Modus Outcomes, Stotfold, UK; ⁹LORA Group, LLC, Royal Oak, MD, USA; ¹⁰Department of Health Services Research, University of Maryland School of Pharmacy, Baltimore, MD, USA
責任者著者連絡先: Marc K. Walton, 125 Crystal Spring Drive, Ashton, MD 20861.
E-mail: mwalton9@its.jnj.com.

つ目である。本報告書は COA 測定の原則を理解するために重要で基本的な定義を提供している。本報告書で提供される基盤として、ベネフィットのある効果を立証するとはどういうことか、患者の評価が治療のベネフィットを示すという目的にどのように関係しているか、そしてこれらの評価が臨床試験のエンドポイントでどのように使用されているかが含まれる。さらに、本報告書は測定特性に影響を与える患者評価および臨床試験の要因の本質的な属性について解説している。評価を開発するとき、または改良するときこれらの要因を考慮すべきである。これらの考慮事項は、研究者が試験をデザインする際に、既存の評価を用いるか、それとも新しいアウトカム評価を開発するかを選択するのに役立つだろう。

本報告書は主として臨床試験のエンドポイントを定義するための新たな COA の開発について記述しているが、これらの原則はもっと一般的に適用される場合がある。COA を吟味、開発するうえでクリティカルな要素は、患者がどのように感じているか、または機能しているかという明確に同定された側面に与える効果として、意図する治療のベネフィットを示すことである。この側面は患者にとって重要であり、また患者の日常生活の一部でなければならない。健康にとって意義のある側面は、直接測定することができ、また、直接評

価するのが実際的でない場合、もしくは測定することが難しい場合は間接的に測定することができる。間接的な測定では、関心のある概念 (Concept of Interest: COI) を同定することができる。COI は患者がどのように感じているか、または機能しているかに関連している必要がある。次に、COI を測定するための方法を開発する。COA の意図した場面および使用方法 (使用場面) において、これらの測定法と患者がどのように感じているか、または機能しているかとの関連性を定義することができる。COA は、エンドポイントに用いる場合に COA の測定特性に影響を与える同定可能な属性または特徴を有する。これらの特徴の 1 つは、判断が測定に影響を与えるかどうかであり、影響を与える場合には、誰の判断であるかである。この属性は COA の次の 4 つのカテゴリを定義する: 患者報告アウトカム、臨床家報告アウトカム、観察者報告アウトカムおよびパフォーマンスアウトカムである。本報告書には、その他の重要な COA の特徴の解説と、詳細な説明が含まれている。本報告書の情報は COA の開発、改良および標準化を支援し、最終的には測定特性を改善させるものである。

キーワード: 臨床アウトカム評価, 関心のある概念, 使用場面, 治療のベネフィット

タスクフォースの背景

ISPOR は 2009 年以來、患者報告アウトカム (Patient Reported Outcome: PRO) の開発および適用に関して取り組んできた内容に基づき、PRO に関するタスクフォース報告書を 8 件発表してきた。これらの報告書は、米国食品医薬品局 (Food and Drug Administration: FDA) が臨床試験における有効性のエンドポイントとして用いられる PRO 測定の妥当性と適切性をどのように評価するかが記載されている、FDA 業界向け指針である「患者報告アウトカムの測定法: 医薬品/医療機器における適応申請のための方法」と一致している。

治療のベネフィットを立証するための主要評価項目または副次評価項目を裏づけるために用いられてきた臨床アウトカム評価 (COA) に関する FDA のレビューおよび認定制度の進展とともに、元 PRO タスクフォースのメンバーが特定の種類の COA、すなわち臨床家

報告アウトカム (Clinician Reported Outcome: ClinRO) を中心に取り組むことを提案した。2013 年 1 月に ISPOR の健康科学政策評議会は、アウトカム研究・タスクフォースのための ClinRO 実施基準の制定を推奨した。続いて、理事会がタスクフォースを承認した。

政府 (米国 FDA)、学術機関、研究機関および製薬企業を含む、さまざまな視点を代表するメンバーおよび一次査読者が選ばれた。タスクフォースの運営グループは、PRO やその他の評価・開発、心理測定、臨床試験データ収集、規制関連の専門家で構成された。さらに、タスクフォースには欧州医薬品庁の職員や査読者といった国際的な人材が登用された。

タスクフォースは 5 週間ごとに電話会議を行い、概要を作成し、報告書に含める検討事項について議論した。さらに、タスクフォースのメンバーは ISPOR の国際会議で、対面で会議を行った。一次査読者とともに

にタスクフォースのメンバー全員が、報告書の多数の草案を査読し、口頭、文書両方のコメントの形式でフィードバックを頻繁に行った。

タスクフォースの審議の過程で、査読者からの特定のコメントや提案、報告書のページ数が多くなることに対する懸念を受けて、必要な内容をすべて網羅しながらも、読みやすく、わかりやすい内容で、論文を2つに分ける必要があることが明らかになった。Value in Health 誌の編集者の許可を得て、論文は2つに分けられた。本稿は COA を支える重要な概念および定義を説明する土台となる。2つ目の論文は ClinRO 評価の特定の要素を掘り下げており、優れた測定の実施基準に関し推奨を行っている。

2013～2015年に開催された ISPOR 欧州年会および国際会議での公開討論会およびワークショップにおいて、予備的な結果および推奨事項を4回発表した。これらの発表時に受け取ったコメントは、その後の報告書の草案作成時に反映された。さらに、タスクフォース報告書の草案は、ISPOR の PRO 査読グループの約 500 名に2回送付されている。

コメントはすべて検討し、大部分は実質的で建設的であった。このコメントは、一連の電話会議でタスクフォースによって議論され、報告書の改訂版の草案で適切に対処された。文書によるコメントはすべて ISPOR のウェブサイトにあるタスクフォースのウェブページ (<http://www.ispor.org/taskForces/Clinical-Outcomes-Assessment.asp>) で公表されている。

タスクフォースの報告書とウェブページは、ISPOR のホームページ (www.ispor.org) から、紫色の Research Tools メニュー、ISPOR Good Practices for Outcomes Research の見出しからアクセスすることも可能である。Patient Reported & Clinician Reported Outcomes Methods (患者報告および臨床家報告によるアウトカムの方法) とリンク http://www.ispor.org/workpaper/practices_index.asp からアクセスすることができる。運営グループの一覧もタスクフォースのウェブページからアクセス可能である。

タスクフォースのメンバー全員がコンセンサスに達した後の 2015 年 8 月に、最終報告書は Value in Health 誌に投稿された。

序 論

患者、医療提供者、規制当局、支払者（健康組合/保険会社）といった複数の関係者の視点から医学的介入の価値を確立することは、治療の利用可能性と適応に必要不可欠である。ある治療法の価値を確立する情報の重要な要素は、介入効果を評価している臨床試験から得られたエビデンスである。（本報告書では、「治療法」という用語は、すでに確立した疾患を有する患者に対する介入によって、改善する方法を意図している。またはさらなる副作用の発現を予防する方法を意図している。あるいはまだ疾病に罹患していない人に対して発症を予防することを意図しているか否かにかかわらず、ある個人に行われた介入を意味するために用いられている）。ベネフィットのある効果の測定は研究対象患者の特定の評価、すなわち患者がどのように感じているか、機能しているか、または生存しているかに依存する。この評価法は、試験のエンドポイントを定義するため、また研究の患者集団を比較するために用いられる。

多くの疾患では、関心のある特定の疾患の徴候を評価するための合意の得られた、十分に定義された信頼性の高い方法がなく、また、いくつかの疾患では重要な徴候が見られないこともある¹⁻¹¹⁾。他の疾患では、現在のアウトカム評価法は信頼性が低く、変化を検出する能力が限定的で、かつ変化の解釈が不十分であり¹²⁻¹⁵⁾、また妥当性が不確実である¹⁶⁻¹⁸⁾といった弱点が知られている。一部の臨床分野では、公表されている臨床試験において使用した方法の特性が十分に理解されていない、複数のアウトカム評価法が用いられている^{19,20)}。

アウトカム評価とその属性が測定にどのような影響を与えるかの理解不足に加えて、疾患の徴候のアウトカム測定のための優れたツールが不足していることによって、治療法の開発が阻害されている。患者を評価する新しく改良された、よりよく理解された評価方法は、治療法の開発を前進させることができる。したがって、多くの研究者はアウトカム測定のための新しいツールを開発、もしくは既存のツールの質を評価しようとしている。残念ながら、臨床

試験の患者評価のための新たなツールを開発、評価すること、あるいは既存の評価法を改良することは困難な場合がある。本報告書がCOAの理解を促す基礎として役立つこと、最終的には信頼性の高い測定能力を持った新たなツールの開発や改良につながることを期待される。

本稿は「ISPOR 臨床アウトカム評価報告書—アウトカム研究のための新たな実施基準タスクフォース」による2件の報告書のうちの1件目である。本報告書では、評価がエンドポイントに用いられた場合にその測定特性が影響を受ける属性または特性を掘り下げて解説する。ベネフィットのある効果とは何を意味するのか、患者評価が治療のベネフィットの立証にどのように関係するのか、そしてこれらの評価が研究のエンドポイントでどのように使用されるのかを明確に提供する。さらに本報告書では、患者評価ツールを開発または改良する際、考慮すべき測定特性に影響を与える評価の、本質的な属性と臨床試験の要因について説明する。

これらの概念は、臨床試験のどのような種類の患者評価ツールを開発するためであっても必要不可欠である。本報告書で定義され、議論されている用語、タイプ、および識別されている特性は、新しいアウトカム評価を開発し、十分に定義され、信頼性があることを示すための基礎を提供するものである。本報告書の解説および推奨は、研究者が臨床試験のデザインに新たなアウトカム評価を検討するうえでも有用である。

このタスクフォースの2件目の報告書である「治療のベネフィットに関する臨床家報告アウトカム (Clinician Reported Outcome: ClinRO) 評価を開発・評価する—新たな測定法の実施基準」(J.H. Powers, D.L. Patrick, M.K. Walton 他未発表の報告書)は、ClinRO 評価法の新たな開発および既存のClinROの評価の原則について記述している。これらの2件の報告書に記述されている概念の一部は、他の出版物²¹⁻²³⁾や米国食品医薬品局(FDA)のガイダンス²⁴⁾でも考察されている。本報告書で提示されている付加的な概念は、研究のエンドポイントに用いるための優れた臨床評価の開発をさらに促進することになるであろう。しかし、優れた測定特性を持つアウトカム評価を開発するプロセスを導くための推奨は、

本報告書のスコープ外としている。

治療のベネフィット

有効性を検討する臨床試験の主要な目標は、治療が有効であるというエビデンスを提供することである。ある治療法が有効であるという結論は、おそらく治療法の使用に起因する治療のベネフィットがあることを意味する。

治療のベネフィットとは、日常生活で患者がどのように感じているか、または機能しているかに関する意義のある側面、または生存に対するベネフィットのある効果である。はっきりと理解できるように、この定義における2つの表現について説明する。1つは「意義のある側面」、すなわち患者がどのように感じているか、または機能しているかに関する効果は、患者にとって意義があるべきである。その効果が患者にとって意義がない場合は、患者に対するベネフィットはない。より具体的に述べると、このことは、治療効果が患者の感覚や機能を変化させる疾患の影響を受ける健康上の側面に対して、正の影響を与えることを意味する。患者が気にかけている健康上の側面であり、患者はこの側面が1)悪化しない、2)改善する、または3)予防することを望んでいる。

定義の2つ目は患者の生活の中にある。これは治療のベネフィットが患者の普段どおりの(日常)生活でみられる側面に影響を与えることを意味する。治療効果は、医療機関でのみ発生する特定のタスクを行う際の変化であり、臨床試験の場面以外の生活の中で患者が行う(または行いたいと思う)普段どおりの動作と定義された関係がない場合には、治療のベネフィットとはならない。測定と患者の普段どおりの生活との間に十分に理解された関係を確立することが、観察された効果が実際に治療のベネフィットであるという結論の要となる。

しかし、単に医療機関で実施する方法によって、患者の生活機能と関連することがよくわかっている簡略化された患者の活動を測定できるのであれば、その測定法に対する治療効果は治療のベネフィットであると解釈できる。たとえば、喘息患者に実施する肺機能検査の明確な差が、喘息患者の普段どおり

の生活において類似した患者に対する意味のある効果と関連性を持つことが経験的に示されている。

同じ疾患に罹患している患者が、異なる症状や徴候を発現することがある。ある徴候を発現している患者にとってもっとも重要であるその徴候は、幅広い患者集団を対象にした比較試験では、その徴候に対する治療のベネフィットを立証するにはあまり適していない。さらに、疾患と関連のない患者の生活環境の差が、これらの症状のどれが一人一人の患者にとって、「もっとも重要であるか」を判断するのに影響を与える可能性がある。

そのうえ、通常は、異なる疾患の徴候に対して、治療法の効果の程度はさまざまである。このことは、有効性は一部の徴候でしか容易に示すことができないことを意味する。これらの症状は患者にとってもっとも重要であるかもしれないし、そうでないかもしれない。したがって、臨床試験の評価法のために選ばれた疾患の特定の側面は、研究において患者にとってもっとも重要なものではないかもしれない。多くの試験では、患者にとって重要度の異なるいくつかの疾患の側面が評価され、治療のベネフィットをより包括的に理解できるようになっている（たとえば、臨床試験の副次的エンドポイントなど）。それにもかかわらず、患者がどのように感じているか、または機能しているかに関して評価した側面は、その疾患に罹患している患者にとって意味を持たなければならない（すなわち、重要性の程度）。

比較臨床試験、たとえば第Ⅲ相試験は、比較治療（しばしばプラセボ）を受けた患者と試験治療を受けた患者の研究のエンドポイントの差を示すようにデザインされている。これらの厳格な研究は、治療のベネフィットの結論を裏づけるためのエビデンスの源である。したがって、治療群の患者と対照群の患者間の研究のエンドポイントの差によって、患者がどのように感じているか、機能しているか、または生存しているかに関する意義のある効果が示されているか、確実に解釈される必要がある。治療法によって提供される治療のベネフィットの重要性は、ベネフィットの種類（評価の特徴）と発生するベネフィットの量（連続尺度での差の大きさや、特定のアウトカム状態を達成した患者の割合の差など）のような、治療のベネフィットの大きさ）の両方に関連

する。

対照的に、感覚や機能に関連していると考えられる測定法の中には、通常的生活において患者にとっての意味（重要性）が自明ではなく、十分に評価されていないものも多く、たとえば、多くの認知機能検査や院内での運動検査などが挙げられる。感覚または機能に関する1つ以上の特定の側面の改善が意図されたベネフィットである疾患において、測定法が意義のあるものとして解釈できないと、治療のベネフィットを示すことができないことになる。患者がどのように感じているか、または機能しているかについての意義のある側面を説明する、または推測するために用いることができる評価法が、これらの疾患に対する治療の評価法に必要である。こうした考察にはしばしば、評価特性を説明するため、また例として患者の機能が用いられる。しかし、重要なポイントや属性は、感覚、機能の両方の評価に適用できる。

本報告書では、生存ではなく、患者がどのように感じているか、あるいは機能しているかの評価に焦点を当てている。アウトカム評価としての生存（多くの場合、生存期間）は、生存の質としてみなすことができる感覚や機能とは明らかに異なる。多様な感覚や機能とは異なり、（原因に起因する死亡に限定されないときは）死亡には判断するための明確な定義があり、容易に意味が理解できる。非致死的なアウトカム、たとえば重度の不可逆的な身体障害を非常に望ましくないと思なすような疾患においてさえ、死亡の意味は明確である（死亡についての詳細な考察は、J.H. Powers, D.L. Patrick, M.K. Walton 他による未発表の報告書を参照）。

患者評価と研究のエンドポイントの関連性

第Ⅲ相試験の主たる目的は、エンドポイントで明確に規定された患者評価に関する研究データの解析を通じて、患者アウトカムにおける治療関連の差を検証することである。エンドポイントに用いられる患者評価であるアウトカム評価は、患者の健康状態のある側面を表すことを目的とした評価またはスコア（カテゴリカルまたは連続）を提供するツールによる測定のことである。すべてではないが、第Ⅲ相

臨床試験の大部分は主要エンドポイントの根拠として、何らかの臨床評価を用いている。アウトカム評価の測定特性には、内容的妥当性、信頼性、変化を検出する能力、および解釈可能性—研究結果と治療のベネフィットとの関係を解釈または理解する能力が含まれる。これらの特性は研究の成功に大きく影響する。

エンドポイントを定義することは、研究期間中に規定された1回以上の時点における患者評価のための特定の方法を同定すること、群間比較のための統計解析方法を規定することが含まれる。他に定義されているエンドポイントの要素とは別で、評価自体はエンドポイントではない。エンドポイントの他の側面（たとえば評価回数、評価時点、統計解析方法）も研究結果の解釈に影響を与えるため、この点を認識することが重要である。

医療によく用いられる一部の患者評価は、アウトカム評価として使用することに適しているが、一部は適していない。患者の健康状態は、臨床においてさまざまな理由・方法で複数回に分けて評価される（診断、予後の推定、効果のモニタリングなど）。特定の目的には特定の評価が適していることが多い。たとえば、多くの場合、実質的な医学的経験は、治療前に特定の新たな臨床的徴候や症状の有無が、特定の疾患の診断にどのように役立つかを理解することにつながっている。こうした臨床的徴候の有無は、患者がその疾患の既知の影響を受けるリスクがあることを示している。

対照的に、診断・治療後の治療のベネフィットは疾患の重要な影響を低下させること、または回避することである。しかし、多くの疾患では、介入実施後に特定の診断徴候を取り除くことが、患者にとって疾患の有害な転帰を回避することにつながることを立証するには、しばしば医学的経験が不足している。したがって、この診断徴候はエンドポイントとして優れた評価法とはいえないのである。

意図する治療のベネフィットの同定

前述したように、治療のベネフィットとは患者の日常生活において、患者がどのように感じ、または機能しているか（または生存しているか）にとって

意義のある側面に対するベネフィットのある効果である。多くの疾患では、単一の症状が、患者の標準的な生活の一部である複数の関連する活動を障害したり、妨げたりすることがある。この障害は患者にとって重要である。これらの密接に関連した影響を受ける活動は、グループとして、患者が生活の中でどのように機能するかの一側面として考えることができる。このグループ分けを同定することは、疾患によって悪影響を受ける人の生活の一側面を一元的に抽象化または概念化することである。

腕の動作に影響を与える疾患、たとえば多発性硬化症やパーキンソン病は、身支度、食事、排泄のような腕の動きに依存している多くの動作を実行する患者の能力を直接的に低下させる可能性がある。機能的な能力に影響を受けるこのグループは、影響を受けるさまざまな動作を繰り返しリストしなくしていいように、「上肢に依存する個々の動作」という用語（または開発者が選択した同様の表現）によって概念的に識別される場合がある。患者の生活にとって本質的に意義のある側面にベネフィットのある影響を与える治療は、明らかに治療のベネフィットを提供している。

もう1つの例として、下肢脱力の影響を受けている患者は、バス停から会社まで歩くこと、ショッピングモールや食料品店を歩くこと、または自宅内を歩くことに問題を抱えている可能性がある。これらの動作や多くの類似した動作は、典型的な健常人の生活に共通するものである。これらの動作の抽象的な特徴は「歩行パフォーマンス」と呼ぶことができるかもしれない。個々の動作を実行できないことは、その動作ができない個人にとって意味がある。下肢筋力を増強することにより、バス停から職場までの歩行能力を向上させる治療は、歩行に依存する同様の動作にもベネフィットのある効果が期待される。個人の生活の側面における活動を改善する治療は、意義あるベネフィット、すなわち歩行に関する治療のベネフィットであると考えられる。

一部では、疾患の重要な影響は、これらの例よりも少ない抽象度（abstraction）で命名することができる。たとえば、一部の患者では、疼痛は転移性乳がんまたは前立腺がんの重要で有害な特徴である。疼痛緩和（疼痛の軽減）は、重要な治療のベネフィッ

トであると考えられる。

多くの疾患では、疾患によって影響を受ける関連する動作（機能的な能力）をいくつかのグループに分けることができる。これらの機能的な能力は、グループ間よりもグループ内の方で共通性が高い場合がある。たとえば、「歩行動作」「手や腕に依存する動作」および「認知に基づく動作」である。各グループは、疾患によって影響を受ける健康にとって明確で意義のある側面の抽象化された概念を、同定できる。健康に関するこれらの別々の側面のうち、1つまたは複数の側面で、動作を行う能力を向上させることは治療のベネフィットとなるだろう。

意図された治療のベネフィットである「健康にとって意義のある側面」は、研究のエンドポイントで使用するアウトカム評価の開発期間中に明確なコミュニケーションを促進する表現で示されるべきである。健康にとって意義のある側面に用いる表現として同定されたフレーズは、アウトカム評価開発期間に関与する人たちや臨床試験に使用するアウトカム評価を検討する人たちが明確に理解できるものであるべきである。たとえば、有効な治療法によってもたらされる意図するベネフィットに関しては、研究者、研究デザイン開発者および規制当局の職員が確実に理解できるものである。

アウトカム評価とともに意図する治療のベネフィットを明確に提示することが、開発期間中は必要不可欠である。アウトカム評価の測定法と意図する健康にとって意味のある特定の側面との関連性を理解することが、測定法の実施基準にとって非常に重要な部分である (J.H. Powers, D.L. Patrick, M.K. Walton 他, 未発表の報告書)。さらに、規制当局の承認を得た治療法では、健康にとって意義のある側面を明確に同定することが、治療のベネフィットを説明するうえで極めて重要である。

医薬品の承認申請に治療のベネフィットを記載するための表現は、研究終了後に複数のエンドポイントについて立証されたすべてのベネフィットを検討した結果に基づいていることに留意すべきである。それにもかかわらず、記載内容は識別された健康にとって意義のある側面と研究のエンドポイントにおける各アウトカム評価との既知の関連性に基づいている。

疾患は複数の徴候が患者のさまざまな機能（または感覚）に影響を与えているため（たとえば、患者に悪影響を与える脚の筋力低下、認知機能障害、呼吸機能障害および疼痛）、意図した治療のベネフィットとして適切であろう複数の健康面が存在する可能性がある。健康にとって意義のある側面は、明確なアウトカム評価と臨床研究のエンドポイントによってそれぞれが示される必要がある。

場合によっては、健康のいくつかの別々の側面を組み合わせて、包括的な単一の健康概念にすることが適切であろう。たとえば、神経変性疾患の中には、日常生活動作のために上肢と下肢の両方を使用する能力が損なわれ、認知機能が損なわれるものがある。機能障害の別のグループは、治療のベネフィットを得られる可能性のある健康にとって別の意義のある側面を定義している。しかし、包括的な治療のベネフィットが意図され、「疾患 X の障害」に対する効果と名づけられることもある。それにもかかわらず、特定の機能障害、たとえば、上肢を使った個々の動作の障害、歩行の障害、認知機能障害を同定することは、意図されたベネフィットを適切に特定し、健康にとって意義のある側面のいくつかの構成要素を評価するために、複数のアウトカム評価が必要かどうかを判断するために必要である。

複数のアウトカム評価を単一の複合エンドポイントに統合する場合、そのエンドポイントは慎重に設定されるべきである。そのエンドポイント構造は、個々の構成要素の評価と、複合した結果にそれぞれがどのように寄与したかを分析できるようにすべきである。これにより、治療に関連することを示しているのはどのベネフィットであるかの理解を促すであろう。

多面的な疾患に対する治療法を開発するときに留意すべき重要な原則は、健康に関連するさまざまな側面は患者にとって重要度が異なるということである。ある側面は、他の側面よりも患者にとって大きな意味を持ち、より重要であったりする。さらに、すべての疾患の徴候が特定の治療法からベネフィットを受けることができるわけではない。したがって、治療法の開発者は、疾患に関連したさまざまな健康にとって意義のある側面を考慮し、その治療法がベネフィットをもたらす可能性があり、臨床試験での

評価を正当化するのに十分な重要性を有するものを研究のために選択すべきである。

関心のある概念 (COI) の測定: 患者がどのように感じているか, 機能しているかを評価するための実用的アプローチ

臨床試験では、アウトカム評価が研究のエンドポイントで用いられる測定値を得るための具体的な方法である。結果を解析し、選択された健康にとって意義のある側面に対して治療のベネフィットが生じたかどうかに関する結論を導く。研究のエンドポイントの評価を選択する際には、その評価が疾患に対する治療のベネフィットを直接示しているか、または（間接的に示している）治療のベネフィットとの関連を示すことができるかどうかを留意することが必要である。多くの場合、患者報告アウトカム (PRO) ツールは、健康にとって意義のある側面を直接測定するために使用される。しかし、多くの障害では、適切な PRO ツールが開発されていないか、あるいは、たとえば自分自身で報告できない人のように、関連する日常の機能を直接測定することが困難である^{22,23,25}。

このような状況では、特定の機能的能力の評価法に関心のある研究者は、代わりに、健康にとって意義のある側面の動作をより単純な（より狭く定義された）身体能力に分けている。健康にとって意義のある側面の動作を行う際には、これらの能力に依存していると考えられている。これらの能力の1つは、健康にとって意義のある側面の中の多くの動作と明確に関連し、かつ、より容易に測定されると仮説を立てられるかもしれない。たとえば、歩行障害における下肢の筋力、またはある種の認知障害における短期記憶である。概念的に簡略化された身体動作は、治療を研究するための実用的な測定ツールを作成する機会として選択することができる。概念的に簡略化された身体動作は、測定法のための関心のある概念 (COI) と呼ばれる (図1)。

健康にとって意義のある側面としての疼痛の例では、疼痛という経験の全体像には、複数の概念的要因または特性がある。これらには、強度、期間、頻度および質 (例: 鋭痛, 灼熱痛または電撃痛) が含

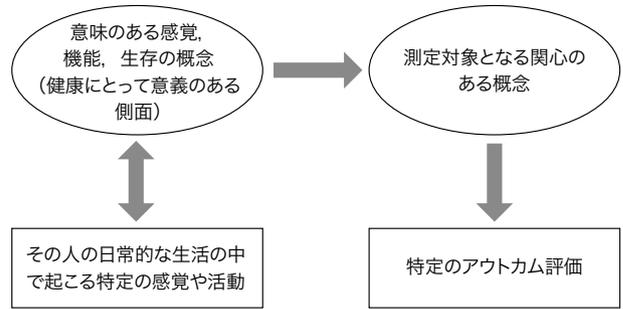


図1 概念とアウトカム評価の関連性

臨床試験を計画する際に研究者は、患者の日常生活の中で起こりうる関連する感覚や活動の中から、疾患の悪影響を受けており、介入によってベネフィットを示すことが期待されるものを選択する。感覚または機能のグループは、健康にとって意味のある概念化された特定の側面として同定される。感覚または機能の選択されたグループを直接測定することを計画しない場合には、測定可能であると考えられ、実質的に健康にとって意義のある側面に影響を与える関心のある概念を策定する。関心のある概念の測定は、具体的かつ明確に定義された臨床アウトカム評価として用いられる。健康にとって意義のある側面と関心のある概念の選択の間には、健康にとって意義のある側面の情報を提供し、かつ測定するための実際的な手段 (特定の臨床アウトカム評価) を持つと考えられる測定のための関心のある概念を持つために、研究者がそれらを改良させていくことによって、相互作用が生じる可能性がある。場合によっては、疾患によって影響を受けた感覚または機能を直接測定することを意図しており、測定のための関心のある概念は、健康にとって意義のある側面と同一である。

まれる。異なる障害から生じる疼痛の経験は、これらの特性が異なる。その結果、疼痛が健康にとって意義のある側面である場合には、研究者は異なる疼痛特性を COI として選択する可能性がある。特定の障害では、測定する疼痛属性として疼痛強度が選択されるかもしれないが、別の障害では、測定のための COI として疼痛頻度が選択されるかもしれない。

すべての研究において、実際の測定に対して対象とする COI がなければならないことを認識することが重要である。さらに、COI は健康にとって意義のある側面である可能性もあれば、そうでない可能性もある。PRO がアウトカム評価である場合、測定のための COI は通常、健康にとって意義のある側面、すなわち COI は健康にとって意義のある側面と同一である。これは健康にとって意義のある側面が、妥当なツールを用いて直接測定可能である場合は効果的である。また、患者の感覚のように健康にとって意義のある側面を直接評価できるのが患者のみである場合にも必要である。

それにもかかわらず、多くの研究では、測定のための COI は健康にとって意義のある側面と同じではない。さらに、選択された健康にとって意義のある側面に対して、治療のベネフィットが生じたかどうかを理解するのに有用な複数の COI が存在することが多い。研究者はエンドポイントの基準を設定するために、一測定に対して COI を 1 つ選択しなければならない (図 2)。このような状況において、COI を明確に同定することは、アウトカム評価の開発を助け、それが意図された使用と解釈に対して妥当性があることを確立するのに役立つであろう。

測定方法は、測定される COI を表すことを意図したカテゴリ評価または連続スコアを生成する手順であり、すなわち COI を具体化した表現である。たとえば、疼痛強度は複数の方法で測定することができる。測定ツールでは、患者に 1) 現在の疼痛の強度を表すスコアを 0～10 の間で選択してもらい、2) 過去 1 ヶ月間の平均的な疼痛の強度をスコア化してもらい、3) 毎日使用した鎮痛薬の数を日誌に記入することを求めることがある。COI を測定するための明確で具体的に定義されたこれらの方法は必ずしも同等ではない。

これらの方法は、疼痛の経験に関して明確に異なる側面を測定しているため、あるいは異なる測定特性を持っているため、臨床試験で異なる結果をもたらす可能性がある。たとえば、信頼性に関して、1 ヶ月間の想起期間での疼痛の評価と現在の疼痛のスコアを比較すると異なる可能性がある。測定の解釈は本質的に明確でない可能性がある; たとえば、錠剤数が、各錠剤により得られた疼痛緩和期間に関連していることもあれば、疼痛の強度が患者個人の閾値をどれくらいの頻度で超えたかに関連することもある。測定特性の違いは、特定の使用状況に対する評価の適合性にも影響しうる (「使用場面」セクションを参照)。

アウトカム評価の解釈方法を理解したら、測定特性の評価を開始することができる。評価が妥当で、使用目的に適していることを保証するために、臨床試験のアウトカム評価を開発している場合や選択している場合に、このステップは必要不可欠である。アウトカム評価の測定特性を評価するための測定特性および原則、またその適合性の確認は、本考察の

範囲を超えているが、他の論文では考察されている^{21-23, 25-28)}。

アウトカム評価の特性は、1) 治療効果を検出する、2) 効果を治療のベネフィットとして解釈するための能力にとって必要不可欠である。アウトカム評価と健康にとって意義のある側面との関連性は、臨床試験のエンドポイントにアウトカム評価を用いる前に理解しておくべきである。新しいアウトカム評価が、治療のベネフィットを示すために適していることを保証するための重要なステップの 1 つが、アウトカム評価が健康にとって意義のある側面と明確に関連しているとする仮説を確認することである。アウトカム評価の違いは、健康にとって意義のある側面の変化を反映したものでなければならない。COI が健康にとって意義のある側面と同一である場合であっても、(COI の具体的な実施形態としての) アウトカム評価と健康にとって意義のある側面との関連性を検討すべきである。

使用場面: 臨床試験の重要な要因

臨床研究は、特定の集団を対象に、臨床試験のエンドポイントでアウトカム評価がどのように使用されるかの詳細を含めた一連の手順を定めた研究デザインで実施される。これらの研究デザインの特性は、まとめて使用場面 (Context of Use: COU) と呼ばれる。COU は、アウトカム評価の設定および使用方法について示すものである。COU の重要性は、FDA によって強調されている²⁹⁾。

COU は新しいアウトカム評価の開発を開始するとき、または既存のアウトカム評価法の評価を開始するときに明確に提示すべきである^{22,23)}。アウトカム評価の測定特性は、COU で定義される要因に影響されうる。COU の包括的な記述は、試験の特定の用途に適したアウトカム評価が開発されることを確実にするのに役立つ。

研究デザインおよび (または) 他の研究要素の変更は、アウトカム評価の性能特性に実質的に影響を与える可能性があることに留意すべきである^{30,31)}。変更した COU にアウトカム評価を適用するときは、改訂した COU について明確に記載すべきである。これらの変更は試験結果の解釈を変えることさえあ

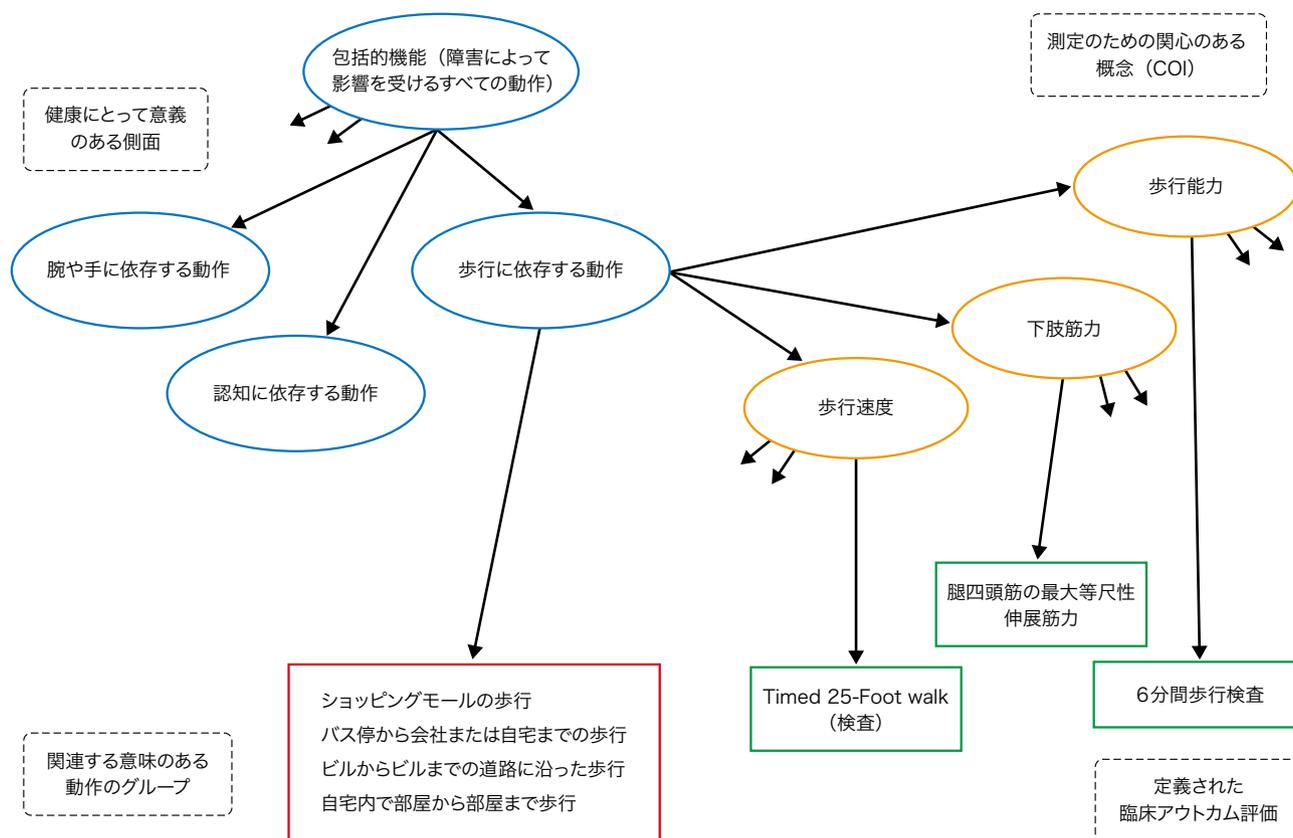


図2 意味のある機能と間接的なアウトカム評価測定との関連性

患者にとって意味のある関連動作のグループ (赤色の長方形。意味のある特定の動作) が、ある介入によって改善する場合には、患者に対する治療のベネフィットとして同定される。グループ分けの抽象的な概念 (青色の楕円形。歩行に依存する機能) と同一である健康にとって意義のある側面に名称が付けられている。この特定のタイプの機能は、疾患を持つ患者が治療のベネフィットを望む可能性のある感覚または機能の意味のある領域 (青色の楕円形) の1つに過ぎない。個々の健康にとって意義のある側面とは、疾患によって影響を受けるすべての意味のある感覚や機能の包括的な範囲の特定の部分であり、そのうちのどれかが治療のベネフィットを意図しており、臨床試験で評価される可能性がある。この例では、歩行に依存する機能が、研究されるべき潜在的な治療のベネフィットとして選択されている。選択されていない健康上の側面と関連する特定の意味のある動作は図示されていない。同定された動作の評価法は、定義されたアウトカム評価 (例: PRO。図中になし) において直接評価されることがある。この場合の測定に関心のある概念は、同定された健康にとって意義のある側面と同じであろう (図1参照)。それよりもむしろ、選択された健康にとって意義のある側面は、意味のある活動を行ううえで重要であると考えられる、狭義の関心のある概念 (COI, オレンジ色の楕円) に分解してもよい。手順は、COI (定義されたCOA, 緑色の枠内) の測定を表す特定の臨床アウトカム評価 (COA) ツールとして考案することができる。この例の中では、パフォーマンスアウトカム (Performance Outcome: PerfO) ツールが、それぞれのCOIツールに対して示されているが、どのCOIに対しても他のPerfOや他の種類のCOAを開発できるであろう (図3参照)。COIとして同定された身体活動を測定するために具体化されたPerfOの手順は、患者の標準的な普段どおりの生活の一部ではないタスクである。COAによって、手順を実施する際に観察された質または量に対するスコアが提供され、研究のエンドポイントを設定するために用いられる。人の日常生活に対するスコアやスコアの変化の意味は、本質的に正確には知られていないが、意義のある機能的活動を反映していると仮定されている。患者に対する特定のスコア (または変化) の実際の意味は、PerfO手順の記述だけでは知ることができず、ツール開発の過程で評価されるべきである。破線の枠内には、図1に示されているように、該当領域で図示されている概念の種類または動作のカテゴリの名前が付けられている。楕円形は特定の概念に名称をつけている: 健康にとって意義のある側面 (青色) またはCOI (オレンジ色) のいずれかである。実線の枠内は、選択されたCOIを測定するために、意味のある (赤色の枠内グループ) 日常的な生活における特定の機能的動作、またはそれぞれがCOA (緑色の枠内グループ) として開発されているかもしれない特定の手順に名称を付けている。概念またはCOAにつながない短い矢印は、その他の可能性が存在することを表しているが、図中には示されていない。COI 関心のある概念; PRO 患者報告アウトカム。

る。新しい COU において、アウトカム評価特性を再評価することが望ましい。

十分に定義された COU を作成するための重要な構成要素には、以下のようなものがある。

1. 研究対象の疾患: 研究対象の疾患は、診断基準などを用いて明確に定義されていることが望ましい。疾患が患者に及ぼす重要な影響を十分に記載し、記載された健康にとって意義のある側面が疾患に関連していることを確認する必要がある。また、研究参加者が研究登録時に罹患しているか、疾患を発症するリスクがあるのか、つまりその治療法が疾患を治療するのか、予防するのかを見極めることも重要である。
2. 患者のサブ集団: 多くの疾患は患者間で不均質であり、患者個人の中では時間の経過とともに変化する。対象集団として意図された精確な疾患サブ集団を特定する必要がある。1) 疾患の表現型 (またはサブタイプ), 2) 患者に特異的な疾患特性 (重症度, 期間, 体の特定部位の関与), 3) 対象となる患者の人口統計学的特性 (年齢, 人種, 民族), 4) 過去の治療歴, 5) 併存疾患, または 6) 試験の選択規準または除外規準として使用される可能性のあるその他の要因など, これらの患者を識別するために幅広い要因が重要であると考えられる。
3. 文化, 言語, またはその他の地域に関連した要因: 多くのアウトカム評価では, アウトカム評価の結果の主要な情報を得るための方法を, 患者への質問に依存している。このように, 研究の対象集団全体で複数の言語が使用されている場合, 異なる各言語で質問が同じ意味を持つかどうかを確認することが重要である。たとえばいくつかの質問は, 患者がどのような要因を考慮して回答しているかの違いなど, 文化的な差異を生じさせる可能性がある。さらに, COI と健康にとって意義のある側面との関連性に影響を与える文化的影響がある可能性がある。この関連性は, アウトカム評価が使用される文化的状況によって異なる可能性がある。研究結果の解釈可能性を確保するためには, 以下の要因を考慮する必要がある。このテーマに関する詳細な情報は, 翻訳と言語バリデーション

に関する ISPOR の 2 つの PRO タスクフォース報告書を参照されたい^{32,33)}。

4. 標準的に付随するケア: 臨床試験では, 標準的な臨床診療の一部であり, その治療が疾患の経過に影響を及ぼす可能性がある場合には, 既存の治療法の一部を試験的介入と併用することが一般的に認められている。アウトカム評価が新たな臨床研究で併用療法の使用が可能になる前の時点で開発された場合は, アウトカム評価の測定特性に影響を与えることがあるため, 予想とは異なる可能性がある。研究の実実施設間の測定特性の不均衡は, 併用療法に関して試験で認められている選択肢のみで, これらの特性を評価する場合や, たとえば, 世界各国で実施されている研究において, 実施設間で治療が異なる場合に生じる可能性がある。したがって, 標準治療の意図する選択肢は COU の一部として特定されるべきであり, 新規または既存のアウトカム評価を開発する際には, これらの問題を念頭に置いておくべきである。
5. エンドポイントの位置付け: エンドポイントの位置付けには 2 つの側面がある。それは, アウトカム評価を用いたエンドポイントが, (研究の解析計画に示されている) 研究の目的の中でどこに該当するかと, エンドポイントが達成されることで規制上どのような役割を果たすことができるかを示すことである。アウトカム評価はある状況では適切であっても, 他の状況では適切でない可能性がある。たとえば, ある使用場面では市販のための承認を決定する裏づけとなり, 特定の疾患において主要な有効性の主張に対する適切なエビデンスを提供する。しかし, 異なる COU においては有効性の補足的な表示主張に対してのみ適切であるかもしれない。アウトカム評価の中には, 有効性の主張を裏付けるのに適切ではないものもある。したがって, エンドポイントの目的および規制上の役割を明確にすることが, COU の重要な要素である。
6. エンドポイント内での使用方法: アウトカム評価はエンドポイントの要素の 1 つである。アウトカム評価は, 異なる臨床試験においてさまざまな方法でエンドポイントやデータ解析に組み入れられる可能性がある。研究期間中にアウトカム評価を

測定した回数とエンドポイントに使用した回数には差があるかもしれない。また、研究期間中の評価の時点(タイミング)に違いがある場合もある。また、エンドポイントの解析方法としては、1) 研究期間中の特定の時点におけるグループのアウトカム評価の平均値、2) 特定の時点で特定のアウトカム評価基準を満たしてレスポnderとなった患者の割合、3) 研究期間中の複数の時点を用いた反復測定、4) イベント発生までの時間など、いくつかの選択肢がある。エンドポイントでアウトカム評価を用いる方法が異なれば、研究期間中の患者の経験の要約が異なってくる。これらの異なる用途は、健康にとって意義のある側面との異なる関係に変換することができる。アウトカム評価の測定数およびタイミングは、解析方法とともにCOUの一部として規定されるべきである。

7. 測定する場面: アウトカム評価は、自宅、外来クリニック、病院の入院患者など、さまざま場面で行うことができる。これらの異なる場面によって、実際に得られる測定は変化し、アウトカム評価の測定特性を変化させることがある。したがって、測定する場面はCOUの一部として明記すべきである。
8. アウトカム評価の実施方法: アウトカム評価は1種類以上の方法や実施方法でデザインすることが可能である³⁴⁾。1) 誰がアウトカム評価を実施するか、たとえば、自分で実施、そうでなければ、訓練を受けていない人物、または訓練を受けた専門家が実施、2) どのようにアウトカム評価を実施するか、たとえば、視覚による vs. 音声による、対面で vs. 電話で、および電子的に vs. 非電子的に、といった選択肢がある。COUの方法および実施方法を明記することが非常に重要である。

実施方法の違いが、得られたデータに影響することがあり、アウトカム評価の測定特性を変えることがある。これは特に1つの試験内で複数のアウトカム評価が用いられているが、アウトカム評価の測定特性の同等性が示されていない場合に問題となる。別の方法や実施方法が導入されている場合は、同等性を証明する必要がある。このテーマに関する詳細な情報は、測定法の同等性および混合研究法のデー

タ収集に関するISPORの2件のタスクフォース報告書を参照されたい^{34,35)}。

アウトカム評価の属性

COUはアウトカム評価を使用する場面および方法を説明しているが、アウトカム評価を定義しては「いない」。アウトカム評価の慎重かつ完全な定義には、測定を実施する手順を正確に記述することが含まれる。臨床試験において評価法を評価、開発および使用するためには詳細が必要である。その定義は測定特性に影響し、どのように測定特性を評価するかに関する特徴的な属性(図3で解説)を決定する。

重要なアウトカム評価属性

属性1: アウトカム評価は患者の積極的な関与または評価者の判定に依存するか?

臨床評価は患者または別の人、たとえば観察結果を統合し、その結果を評価に変換するため、また結果(測定)を記録(報告)するために、訓練を受けた医療専門家、配偶者、介護者または教師に依存する。臨床評価は、1) 患者の意思の変動(連続した患者内または患者間で評価の差を生じさせることがある)、2) 評価者の判定に対する依存により、患者の真の臨床状態とは無関係に変動の影響を受けやすい評価である。臨床試験において患者アウトカムを評価するために用いる臨床評価は臨床アウトカム評価(clinical outcome assessment: COA)と呼ばれる。

臨床評価は、両方の変動源の原因の影響を受けやすい。一部の患者評価は、(物理的に、または電子的に)記録を作成するため、または評価あるいはスコアの基礎となる動作を行うために患者の積極的な関与を必要とする。記録を作成したり、動作を行ったりするプロセスは、評価に参加する患者の意思や動機の影響を受ける。動機のレベルや評価プロセスに対する注意力は、患者内でも時間の経過とともに変わる可能性があり、患者間でも系統的に異なる可能性がある。このような変動は、関心のある基礎となる医学的状態(症状または機能的な能力)の違いとは無関係に、評価やスコアの違いをもたらす可能性がある。こうした変動はアウトカム評価の測定特性に潜在的に影響を与える。

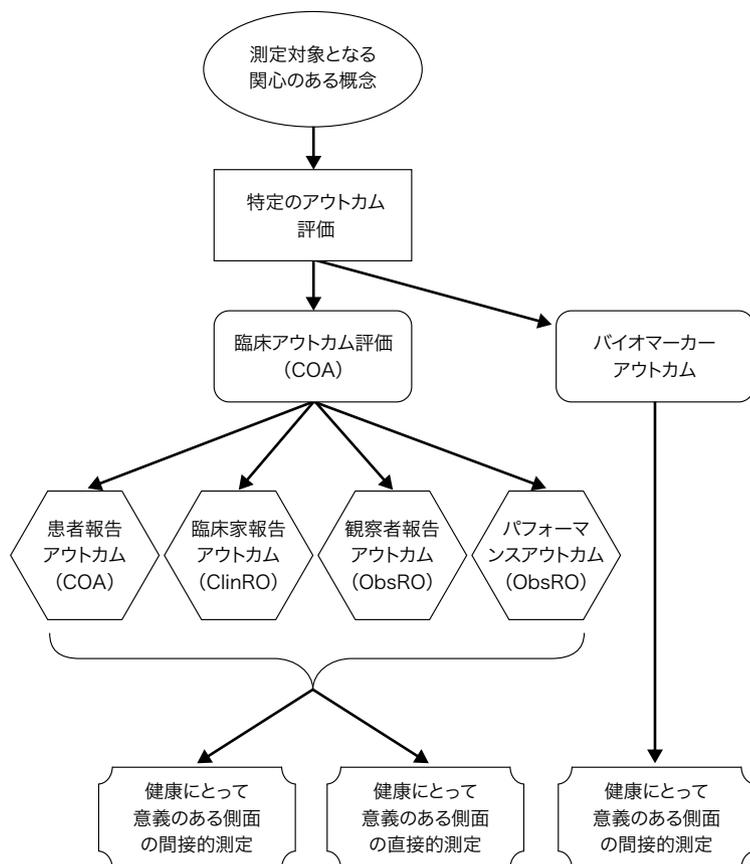


図3 アウトカム評価の属性

関心のある概念の測定を具体化するために、特定のアウトカム評価を選択または作成する。アウトカム評価には2つの主要なタイプがある：臨床アウトカム評価とバイオマーカーである。臨床アウトカム評価は、その判断が報告される測定に影響を与える人のタイプを同定する属性を持つ。臨床アウトカム評価は、患者、臨床家、または臨床家以外の観察者の判断に影響されることがある；また、臨床アウトカム評価は患者が行ったタスクの記録だけを行い、判定していないものもある（パフォーマンスアウトカム）。臨床アウトカム評価は、潜在的な治療のベネフィットとして選択された意味のある感覚や機能を直接報告する場合もあれば、それらの感覚や機能に関して間接的に有益であると考えられる測定を報告する場合もある（図1参照）。バイオマーカーは、健康にとって意義のある側面を間接的にしか測定できない。

評価者に依存する測定には、患者（自己観察）であれ、判断力を用いて評価を行う他の人であれ、変動性が内在している。患者以外の評価者は、1) 質問に対する患者の回答、2) タスクを実行する患者の観察、または3) 評価を決定するために複数回実施した観察結果の相対的な重み付けをもとに、判断を行う可能性がある。評価者の判断は、過去の経験や評価者のバイアスに影響され、患者の医学的状態の違いとは無関係に、評価者間のスコアリング差が生じる可能性がある。評価者の判断が最終的な評価にどのように影響するかの大きな格差によって、COAの妥当性を損ねたり、信頼性を低下させたりしうる。

患者の医学的状態と無関係な測定には患者または評価者の影響がある可能性があるが、COAは、対象COUの中で十分に定義され、信頼性の高いものであることを保証したうえで開発することは可能である。このようなCOAは、患者の医学的状態に関する情報量が多く、非常に広い範囲の治療のベネフィットについて患者にとって意味のある治療効果

を明らかにすることができる。

対照的に、バイオマーカー評価は患者の動機や評価者の判断の影響をほとんど、またはまったく受けない評価法である（バイオマーカーの定義に関するワーキンググループは代用エンドポイントに用いられるバイオマーカーにかかわる問題について検討し、その考察でバイオマーカーを定義した³⁶⁾。臨床評価に焦点を当てた）。今回の考察では、バイオマーカーと臨床評価をより確実に区別するために、バイオマーカーの特徴付けがワーキンググループのものとは異なっている。本報告書で用いたこの区別は、エンドポイントに使用する臨床評価を開発するときに生じる幅広いCOAのタイプおよび複雑性に対処する場合に重要である。例としては、標準化された方法で測定された血中または尿中のタンパク質レベル、または磁気共鳴イメージングで可視化された病理学的病変の自動定量的なサイズ測定が挙げられる。

多くのバイオマーカーは、診療や臨床試験のプロトコルで「臨床検査値」と呼ばれるものの中に含

まれていることに注意が必要である。それにもかかわらず、体液中に物理的に存在する物質のこのような検査室での測定は、患者の動機や評価者の影響のばらつきの影響を受けないため、ここで説明した枠組みで定義されているような臨床評価ではない。このフレームワークにおける「臨床評価」という用語の「臨床」は、これらの評価をバイオマーカーと区別するために、「患者に関わる、または患者に関与する」という一般的な言葉の意味から狭められている。

属性 2: 評価者は誰か？どのようなタイプの評価者が判断を下して測定を行うのか？

評価者は、評価を実施している間に情報を入手し、測定として記録（報告）される評価をつけるために、聞いたこと、または観察したことに対して判定を行う人である。臨床試験において患者評価には多種多様な人が関与する可能性があるが、この COA に関する考察では、われわれは評価者として次の 3 つのカテゴリを使用する。1) 患者、2) 臨床家（患者評価に対して専門的な判定によって評価する人と定義する、例：研究者や看護師）および 3) 臨床家以外の観察者、である。異なる背景や経験、訓練によって、人は異なる視点およびスキルを持っている可能性が高いため、彼らの判定は異なる可能性があり、評価に大きな影響を与えることがある。臨床アウトカム評価には、評価者の判断を伴わず、患者の積極的な関与に依存する、つまりバイオマーカーではない COA とされる第 4 のカテゴリがある (図 3)。

適切に定義、開発、そして評価された場合に、COA のこれら 4 つのカテゴリのいずれもが、十分に定義された信頼性の高い COA として受け入れることが可能であり、臨床試験のエンドポイントとして用いるのに適したものとなる。どのタイプの COA が特定の COI にとってより有利であるかは、意図する使用場面に強く影響されるため、新規 COA の開発や既存の COA の評価を行う際には、慎重に検討する必要がある。

PRO 評価は、患者が評価者となる評価法である。PRO 評価は、質問に対する患者の直接の回答に依存する。これらの回答は、紙によって、電子化された質問票を介して、あるいは患者の観察や報告が、インタビュー側側の解釈（判断）なしに、話したと

おりに正確に記録される面接など、さまざまな方法で記録される。患者が直接報告することで、幅広い感覚や機能を捉えること、ならびに患者がどのように感じているか、たとえば、疼痛や気分の落ち込みについて直接測定結果を提供することが可能であるため、PRO の COA を開発するうえで関心が高まっている。このテーマに関する FDA のガイダンス²⁴⁾では、これらのツールに規制当局が関心を持っていることを取り上げている。

臨床家報告アウトカム (ClinRO) 評価は、適切な専門的訓練を受けた研究者チームのメンバーが評価者となる評価である。臨床家は、専門知識と判断を患者の観察、または患者との会話に適用し、COA の定義に従った評価を行う。たとえば、ClinRO の評価では、臨床家が面接での質問に対する患者の反応を解釈したり、患者の行動の質を判断したり、身体所見を判断したりすることが求められることがある。

ClinRO 評価とは、評価を決定する人が適切に判断をくだすためには、特定の専門的なトレーニングを受けなければならない COA である。ClinRO のアウトカム評価はすべて COA であるが、すべての COA が ClinRO というわけではないことに留意すべきである。その区別は臨床 (属性 1 参照) と臨床家 (特定の種類の報告者) との違いである。

観察者報告アウトカム (Observer Reported Outcome: ObsRO) 評価は、患者以外の人を観察を行い、評価・記録するが、特別な専門的訓練を必要としない評価法である。それにもかかわらず、点数は観察者の視点の影響を受ける。これらには、仲間、たとえば、親、配偶者、または患者の介護者によって行われるのが最適な COA に含まれる。観察者である評価者は、何を観察し、点数をつけるために何を判定するかをしばしば教えられるが、医療専門家の訓練は必要ではない。観察者評価者は、この狭義の意味で使われる用語で、ClinRO 評価を行う際に患者と対話し、観察しなければならない臨床専門家と対照的な意味で使われている。

パフォーマンスアウトカム (Performance outcomes: PerfOs) とは、患者の評価は行われるが、評価者の判断が測定値に影響を与えない COA である。PerfO は評価を決めるための判断に依存しない特定の方法

で、定量化される定義されたタスクに対する患者のパフォーマンスに基づいている。PerfOsには6分間歩行検査や一定時間内にキーと絵記号数をマッチさせた数を調査するシンボルマッチのようなツールが含まれる。臨床家や観察者がPerfO評価のタスクを指示・モニタリングしているが、パフォーマンスを定量化するための判断は行わない。これらの評価は、患者の動機または意思がタスクのパフォーマンスに参与しているため、依然としてCOAとして分類される。

PerfsOsは一般的に医療機関で評価されるが、スコアは評価者の判断に依存しないため、しばしば良好な信頼性を持つようにデザインすることが可能であり、治療のベネフィットに関する身体機能活動が、健康にとって意義のある側面である障害を扱う施設共同臨床試験の多くにとって魅力的なものとなる。しかし、PerfOsは通常、意図した意義のある機能的活動を測定するものではなく、その代わりに、意義のある機能の情報提供に役立つと考えられるCOIを測定するようにデザインされている(図2)。

属性3:アウトカム評価は、患者がどのように感じているか、または機能しているかという意義のある側面を直接測定しているか?

健康にとって意義のある側面に関する直接的なエビデンスを提供するCOAツールは、直接測定型COAと呼ばれる。いくつかの研究では、エンドポイントのCOAが健康にとって意義のある側面を直接評価しているため、観察された治療効果は本質的に意義あるベネフィットを示していると解釈できる。すべてではないが、多くのPROツールは、このように解釈可能性を持つように意図されている。

間接測定型COAは、関連する測定可能なCOIを評価するものである(図3)。間接評価型の解釈は、意図する健康にとって意義のある側面に関して本質的に明確ではない。アウトカム評価と健康にとって意義のある側面との間に十分に定義された関係があることで、エンドポイントに対する良好な効果を健康にとって意義のある側面に対する好ましい効果として(すなわち、治療のベネフィットとして)解釈することが可能になる。

COAが健康にとって意義のある側面の直接評価

型か間接評価型かを認識することは、COIの測定と健康にとって意義のある側面との関係を立証するエビデンスが必要かどうかを判断するうえで重要である。

間接的であることの段階的な質を認識することが重要である。一部のCOAは、患者の日常生活において、どのように患者が機能しているかに近いが同一ではないCOIを表している能力または動作を測定する。例として、日常生活動作³⁷⁾または視力検査をシミュレーションした医療機関でのパフォーマンスを測定するツールが含まれる。これらのCOAは、健康にとって意義のある側面そのものを評価するのに近い測定である。その他のCOAは、日常生活における患者の機能とは実質的に異なるCOIを測定する(例:仰臥位の大腿四頭筋等尺性筋力)。

間接的なCOAの変化や差の解釈は、COAの測定と健康にとって意義のある側面との関係を明確に定義する追加のエビデンスに依存する(JH Powers, DL Patrick, MK Walton, et al, unpublished reportにある測定法の原則に関する考察を参照、未発表の報告書)。間接測定型COAについては、COIと健康にとって意義のある側面との間の間接性の程度が、優れた測定基準に従ってCOAを開発し評価する過程で得られるべきエビデンスの量と種類の指針となる(JH Powers, DL Patrick, MK Walton, et al, 未発表の報告書)。関心のある概念のCOA測定と健康にとって意義のある側面との関係を確立することは、日常生活の中で患者の機能と密接に似ているCOAの場合には、患者の日常生活とは実質的に異なるCOAの場合よりも、一般的にはより簡単である。

この考察は、実際に健康にとって意義のある側面から一般的に非常に遠いバイオマーカーにも関連している。治療のベネフィットを示すためにバイオマーカーを使用することを裏付けるために、バイオマーカーと感覚や機能の意義のある側面との関係を確立するにはかなりの量のエビデンスが必要である。しかし、十分なエビデンスが得られる場合には、バイオマーカーは非常に価値があり、たとえば、多くの高血圧治療では血圧、一部の糖尿病治療ではヘモグロビンA1c値など、代替エンドポイントとして使用することができる。

使用場面が測定特性に与える影響

特定の臨床試験に対する COA の適合性は、その試験で規定されているエンドポイントに用いられている測定特性に依存する。前述したように、これらの測定特性は具体的な COU や評価を実施するとき用いられた方法といった複数の要因によって影響されうる。COU は、COI のアウトカム評価測定が、意図された健康にとって意義のある側面と十分に関連しているかどうか、また、信頼性や変化を検出する能力などのアウトカム評価の測定特性に影響を与える可能性がある。

単一疾患の範囲内であっても、COU の要素によって、観察された治療効果の解釈可能性が変化する。COA の持つ性能のうちのほんの一部である治療効果は、疾患の軽度の病期においては患者にとって臨床的な価値があるかもしれないが、同じ疾患の重度の病期では患者にとって重要ではないかもしれない。

変化を検出するエンドポイントの能力は、異なるデザイン特性を持つ臨床研究では異なる可能性がある。ある患者サブ集団に対して良好な性能を示す COA は、疾患集団全体の別の部分では許容できない天井効果または床効果がある場合がある。緩やかな進行性の悪化が自然な疾患経過であり、患者の衰えを軽減する（疾患の悪化を軽減する）という治療のベネフィットを示すことを目的とした研究では、1 年間の試験で治療効果を検出できれば COA の信頼性は十分であると考えられる。しかし、1 年間の試験と同じ規模の 3 ヶ月間の研究では、治療効果を示すのは適切でない可能性がある。

同様に、ある使用場面に適した COA は、意図する健康にとって意義のある側面に変化がなくても、別の疾患を研究した場合には妥当ではないかもしれない。たとえば、ある疾患では健康にとって意義のある側面に密接に関連している COI (COA によって測定される) は、別の疾患ではあまり関連していないかもしれない。それは、別の疾患の他の特徴が、健康にとって意義のある側面に対する COI の影響力を減少させるからである。

COA を同定する

意図された治療のベネフィットを明示的に同定するのと同様に、COA は名称と説明の両方で明確に同定される必要がある。COA の測定特性および臨床試験に対する適合性は特定の COU によって異なることがあるため、COA を開発・評価するときには COU も明確に記述すべきである。

PerFo 評価のために実施するタスク、または ClinRO 評価のために意図する COI を表しているような適切な名称をつけるべきである。さらに、COA を新しい場面 (COU) 用に改訂する場合は、バージョン (版) を明確にすべきである。さまざまな研究者が異なる研究で COA を使用しているため、手順が時間の経過とともに変化し、長年使用されてきた COA が存在する。研究報告書には変更点が記載されていないことがよくある; たとえば、COA のバージョン (版) や詳細が記載されていなかった³⁸⁾。このように、同じ COA を使用しているように見える研究でも、実際には明らかに異なるツールを使用している可能性がある。COA を評価または使用するうえでのこうした混乱を避けるために、明確で説明的な新たな COA の名称を用いること、既存の評価法のバージョン (版) を同定することが重要である。

健康にとって意義のある側面を直接測定する評価の場合、これは測定手順の記述から明確になるかもしれない。残念ながら、意図する健康にとって意義のある側面に関する本質的な明晰さは、間接的な評価ではめったに当てはまらない。したがって、意図された健康にとって意義のある側面は、それが暗黙のうちに明確であると思われるかどうかにかかわらず、すべての評価において明示的に同定されるべきである。

最後に、重要な臨床研究のエンドポイントに計画されている既存の COA を評価し、健康にとって意義のある側面との関係を確認するのに十分な情報があるかどうかを判断することが必要不可欠である。名称それ自体は、COA が健康にとって意義のある側面の適切な測定法であるという結論を裏づけとはならない。たとえば、「疾患 X 障害スケール」という名前の COA だけでは、疾患 X やその他の疾患の「障害」を適切に測定できるとは限らない。推奨さ

れるアプローチは以下の通りである：既存の COA において、健康にとって意義のある側面との関係が十分に定義されているかどうかを判断するために、エビデンスを評価すること。新しい COA では、実際に測定された内容を反映するように名前を付け、意図した健康にとって意義のある側面とは別の名称とすること。

結 論

医学的介入のベネフィットは、一般的に、有効性のエンドポイントに良好な効果を示す介入に関する比較臨床試験で示される。臨床試験の成功は、多くの要因、その中でも研究のエンドポイントを定義するために適切なアウトカム評価が利用できるかどうか依存する。

疾患におけるベネフィットを示すためには、これまでに存在していない患者評価を開発するか、または特定の疾患の研究のために十分に定義され、信頼性があり、解釈可能であることが示されていない評価を慎重に再評価する必要がある。1) 既存の治療法よりも新しい治療法の優位性を示すことができる改良されたエンドポイント、または 2) これまで評価されていなかった疾患の側面に対する治療効果を評価するための新しいエンドポイントをサポートするために、新たな評価法の開発も必要とされることがある。

エンドポイントで示す効果は、患者にとって意味のある効果、すなわち、患者がどのように感じ、機能し、または生存するかに正の効果があると解釈される必要がある。研究のエンドポイントは、アウトカム評価（もっとも一般的な臨床評価）に基づいて構築される。これは、十分に特定された使用場面（COU）で使用された場合に、健康にとって意義のある側面を表すことを意図している。COA は、1) 健康にとって意義のある側面の直接的な測定、または 2) 健康にとって意義のある側面と強い関係があり、より容易に測定できると考えられる COI の測定、いずれかの関心のある概念を測定する。

COU におけるアウトカム評価のパフォーマンス特性を理解するための最初のステップは、アウトカム評価の属性、特に本報告書で考察している属性を

同定することである。すべてのアウトカム評価は、1) 患者の意思 / 動機または評価者の判断が評価またはスコアを決定するかどうか、2) 判断を行う評価者のタイプ（患者、臨床家、観察者）または評価に評価者の判断が入らない PerfO、そして最後に、3) 意図された健康にとって意義のある側面が直接的に評価されるか、間接的に評価されるかによって分類することができる。

これらの属性によってアウトカム評価を分類することより、研究者が COA の開発と評価を行うことを支援することができる。これらの取り組みの中で注意を払うべき要素としては、1) 評価を実施するための詳細な手順の特定、2) アウトカム評価が十分に定義され、信頼できるものであるかどうかの評価方法の決定、3) 特定の COU を開発している間にアウトカム評価法を改良する、などが挙げられる。COA が十分に定義され、信頼性があり、計画された臨床試験に適しているかどうかを評価するには、1) 新しい COA を適切な名称で明確かつ完全に同定するか、既存の COA を用いる場合はそのバージョン（版）を含めること、そして 2) 健康にとって意義のある側面、測定した COI および COU の詳細な説明を記述することを必要とする。

本稿に提示された概念および用語は、COA を明確に記述し、効果的に開発または評価する助けとなる。COA のこれらの特性の多くは、古くから存在する COA に関して明確に規定されてはいなかった可能性がある。このことは、こうした古くから存在する COA が無効であり、今後の使用に適していないことを意味するわけではない。むしろ研究者は、古くから存在する COA の特徴を完全に理解していなかった可能性があり、さらなる評価が必要であることを認識すべきである。研究者は COA を選択する過程で、このことを考慮すべきである。将来の臨床試験のために実質的な COA の開発や評価法が必要であると研究者が判断した場合には、研究デザインを最終化する前にこの取り組みをしっかりと開始すべきである。これらの推奨事項に従うことで、COA の改善やより優れた研究のエンドポイントの改善につながり、より有益で効率的な臨床試験が可能になる。

用語集

簡易な要約の定義 (詳細な説明は本文を参照すること。)

バイオマーカー: 患者の動機や意思, または評価者の判断に影響されない評価。一般的なバイオマーカーのタイプには, 血液の生化学的測定と X 線画像の定量的測定がある。

臨床評価: 1) 患者の意思のばらつき (患者内または患者間で個々の患者の評価の違いを引き起こす可能性がある), または 2) 評価者の判断への依存 (これは経験または視点の違いにより評価者間で異なる可能性がある) のために, 患者の真の臨床状態とは無関係にばらつきの影響を受けやすい評価。

臨床アウトカム評価 (Clinical Outcome Assessment: COA): 臨床試験において患者アウトカムの測定に用いられる臨床評価ツール。PRO, ClinRO, ObsRO および PerfO の 4 種類がある。

臨床家報告アウトカム (Clinician Reported Outcome: ClinRO): 多くの研究者チームが評価者である COA の一種。研究者の専門的な訓練は, どのような評価やスコアが報告されるかの判断に依存する。すべての ClinRO は COA であるが, すべての COA は ClinRO ではない。

測定に関心のある概念 (Concept of Interest: COI): アウトカム評価が測定することを意図している概念。COI は, 感覚または機能の選択された意義のある側面と同一であるかもしれない。しかし, COI とはしばしば, 患者の日常生活の中で本来意味のある感覚や機能ではないものを簡略化した形や構成要素のことである。それはすなわち, 健康にとって完全に意義のある側面ではないが, 健康にとって意義のある側面と十分かつ間接的に関連すると考えられている。

使用場面 (Context of Use: COU): 研究デザインの詳細, 研究内での COA の使用方法, 結果の解釈の記述。

直接的に意義のある COA: 患者の実際の感覚または機能を直接測定する COA。感覚または機能は患者にとって意味があり, 患者の日常 (標準的な) 生活の一部でなければならない。すべての COA は COI を測定することを目的としている。COI が意図する健康にとって意義のある側面と同一である場合のみ, COA は直接的に意味があるものとなる。

感覚, 機能, 生存: 疾患により悪影響を受ける可能性がある個人における普段どおりの (日常) 生活の側面。

間接的に意味のある COA: 意味があり, 患者の日常生活

の一部である感覚または機能を間接的に評価する COA。しかし, 間接的に意味のある COA は, 健康にとって意義のある側面と優れた関連性を持つことが意図されている。

健康にとって意義のある側面: 患者が気にかけている疾患の健康上の側面 (感覚, 機能または生存) であり, 患者はこの要素が 1) 悪化しない, 2) 改善する, または 3) 予防されることを望んでいる。

観察者報告アウトカム (Observer Reported Outcome: ObsRO): 特殊な専門トレーニングを必要としない患者以外の人が観察, 評価, 記録を行うことができる COA。それにもかかわらず, 評価は観察者の視点の影響を受ける。

アウトカム評価 (Outcome Assessment): 患者の医学的状態のある側面を表すことを目的とした評価またはスコア (カテゴリカルまたは連続量) を提供する測定ツール。適切なアウトカム評価には COA とバイオマーカーの両方が含まれていてもよい。

患者報告アウトカム (Patient Reported Outcome: PRO): 患者から直接報告される COA。健康状態に関する質問への患者の回答が, 他の誰かによって修正または解釈されることがなく記録される。

パフォーマンスアウトカム (Performance Outcome: PerfO): 特定の方法で定量化される定義されたタスクを実行することで患者を評価する COA。研究者チームのメンバーが PerfO タスクを実施し, 患者のパフォーマンスをモニターすることはできるが, 研究者はパフォーマンスを定量化するために判断を下さない。

研究のエンドポイント: 臨床試験で得られた結果を解釈し, その治療法が治療のベネフィットをもたらしたかどうかを決定するために, 解釈され, また臨床試験から得られたアウトカムの結果。研究のエンドポイントは, 特定のアウトカム評価で構成され, 研究期間中に規定された時点で測定され, 特定の統計手法に基づいて解析される。

治療のベネフィット: 患者が生活においてどのように感じているか, 機能しているか, または生存しているかという意義のある側面に対する良好な効果。疾患によって影響を受ける健康面への影響とは, 患者が影響を受けていることを気にしている感覚や機能の変化であり, 患者は 1) 悪化しない, 2) 改善する, または 3) 予防されることを好む。治療法によって影響を受ける感覚や機能の側面は, 患者の普段どおりの (日常) 生活の中で何が起こるかであるべきである。

謝 辞

私たちはこの原稿に対して重要なコメントをよせてくれた Food and Drug Administration Center for Drug Evaluation and Research の Elektra Papadopoulos and Ashley Slagle に感謝する。

ISPOR メンバーのコメントは、ISPOR の Best Practice を特徴付ける高品質のコンセンサスに基づいたアウトカム研究レポートに貢献した。ISPOR PRO レビューグループの以下の 40 人のメンバーによる、草稿に対する貴重な書面によるフィードバックに深く感謝する: Sarah Acaster, Michael Adena, Ethan Basch, Marc Berger, G.S. Bhatta- charyya, Monika Bullinger, Joe Capelleri, Lee Yee Chong, Susan Dallabrida, Slaveyko Djambazov, Celeste Elash, Francis Fatoye, João Guerra, Chad Gwaltney, Mike Hagan, Katarina Halling, Stacie Hudgens, Christine C. Huttin, Zhanna Jumadilova, Cicely Kerr, Bellinda King-Kallamanis, Kathy Lasch, Hsiao-Yi Lin, Yvonne Lis, Linda Gore Martin, Oren Meyers, Talya Miron-Shatz, Annabel Nixon, Cathy Anne Pinto, Antoine Regnault, Diana Rofail, Sarah Shingler, Ajit Singh, Lisa Strouss, Jun Su, Tara Symonds, Diane Turner-Bowker, Wendy J. Ungar, and Etta Vinik. また、4 回のプレゼンテーションの際に、口頭でコメントをいただいた方々のコメント、専門知識、見識に感謝する。

この出版物の内容は、必ずしも米国保健福祉省 (the Department of Health and Human Services) の見解や方針を反映しているわけではなく、商号、商品、または組織についての言及は、米国政府による承認を意味するものではない。この記事で表明された見解は著者の個人的な見解であり、欧州医薬品庁 (European Medicines Agency) またはその委員会や作業部会、国会機関の立場を代表、またはその立場を反映したものとして解釈または引用することはできない。

財政的支援の財源: このタスクフォースと原稿には、財政的支援がない。1 人の共著者は、彼の全般的な活動に対する財源についての開示を求めた (この原稿に特化したものではない)。JH Powers は、米国国立がん研究所 (the National Cancer Institute) と米国国立衛生研究所 (National Institutes of Health) (契約番号 HHSN261200800001E) からの連邦資金で全体的または部分的に資金提供を受けている。

参 考 文 献

- Johansson GM, Hager CK. Measurement properties of the Motor Evaluation Scale for Upper Extremity in Stroke patients (MESUPES). *Disabil Rehabil* 2012;34:288–94.
- Coster WJ, Haley SM, Ludlow LH, et al. Development of an applied cognition scale to measure rehabilitation outcomes. *Arch Phys Med Rehabil* 2004;85:2030–5.
- Burk K, Schulz SR, Schulz JB. Monitoring progression in Friedreich ataxia (FRDA): the use of clinical scales. *J Neurochem* 2013;126(Suppl. 1):118–24.
- Howard K, Berry P, Petrillo J, et al. Development of the Shortness of Breath with Daily Activities questionnaire (SOBDA). *Value Health* 2012;15:1042–50.
- Padua L, Evoli A, Aprile I, et al. Myasthenia gravis outcome measure: development and validation of a disease-specific self-administered questionnaire. *Neurol Sci* 2012;23:59–68.
- Merkel PA, Clements PJ, Reveille JD, et al. Current status of outcome measure development for clinical trials in systemic sclerosis: report from OMERACT 6. *J Rheumatol* 2003;30:1630–47.
- Brown B, McKenna SP, Solomon M, et al. The patient-reported impact of scars measure: development and validation. *Plast Reconstr Surg* 2010;125:1439–49.
- Marquis P, Lasch KE, Delgado-Herrera L, et al. Qualitative development of a patient-reported outcome symptom measure in diarrhea-predominant irritable bowel syndrome. *Clin Transl Gastroenterol* 2014;5:e59.
- Aabenhus R, Thorsen H, Siersma V, Brodersen J. The development and validation of a multidimensional sum-scaling questionnaire to measure patient-reported outcomes in acute respiratory tract infections in primary care: the acute respiratory tract infection questionnaire. *Value Health* 2013;16:987–92.
- Drobes DJ, Thomas SE. Assessing craving for alcohol. *Alcohol Res Health* 1999;23:179–86.
- Gold DA. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. *J Clin Exp Neuropsychol* 2012;34:11–34.
- Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 2007;6:1094–105.
- Paltamaa J, Sarasoja T, Leskinen E, et al. Measures of physical functioning predict self-reported performance in self-care, mobility, and domestic life in ambulatory persons with multiple sclerosis. *Arch Phys Med Rehabil* 2007;88:1649–57.
- Hobart J. Rating scales for neurologists. *J Neurol Neurosurg Psychiatry* 2003;74(Suppl. IV):iv22–6.
- Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure. *Arch Dermatol* 2004;140:1513–9.
- Terwee CB, Bouwmeester W, van Elsland SI, et al. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthritis Cartilage* 2011;19:620–33.
- Terwee CB, Mookkink LB, Steultjens MPM, Dekker J. Performance based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Rheumatology* 2006;45:890–902.
- Cano SJ, Hobart JC, Hart PE, et al. International Cooperative Ataxia Rating Scale (ICARS): appropriate for studies of Friedreich’s ataxia? *Mov Disord* 2005;20:1585–91.
- Spuls PI, Lecluse LLA, Poulsen MNF, et al. How good are clinical severity and outcome measures for psoriasis? Quantitative evaluation in a systematic review. *J Invest Dermatol* 2010;130:933–43.
- Croarkin E, Janoff J, Barnes C. Evidence-based rating of upper

- extremity motor function tests used for people following a stroke. *Phys Ther* 2004;84:62–74.
- 21) Rothman M, Burke L, Erickson P, et al. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health* 2009;12:1075–83.
 - 22) Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: part 1—eliciting concepts for a new PRO instrument. *Value Health* 2011;14:967–77.
 - 23) Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: part 2—assessing respondent understanding. *Value Health* 2011;14:978–88.
 - 24) US Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Available from: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm064981.htm> [Accessed July 20, 2014].
 - 25) Matza LS, Patrick D, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO Good Research Practices for the Assessment of Children and Adolescents Task Force. *Value Health* 2013;16:461–79.
 - 26) Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
 - 27) Hobart JC, Lamping DL, Thompson AJ. Evaluating neurological outcome measures: the bare essentials. *J Neurol Neurosurg Psychiatry* 1996;60:127–30.
 - 28) Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? *Injury* 2011;42:236–40.
 - 29) US Food and Drug Administration. Clinical Outcome Assessment Qualification Program. Available from: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm>. [Accessed May 13, 2015].
 - 30) Freeman J, Walters R, Ingram W, et al. Evaluating change in mobility in people with multiple sclerosis: relative responsiveness of four clinical measures. *Mult Scler* 2013;19:1632–9.
 - 31) Post B, Merkus MP, de Bie RMA, et al. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Mov Disord* 2005;20:1577–84.
 - 32) Wild D, Eremenco S, Mear I, et al. Multinational trials—recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient Reported Outcomes Translation & Linguistic Validation Good Research Practices Task Force Report. *Value Health* 2009;12:430–40.
 - 33) 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;8:94–104.
 - 34) 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;12:419–29.
 - 35) Eremenco S, Coons SJ, Paty J, et al. PRO data collection in clinical trials using mixed modes: report of the ISPOR PRO Mixed Modes Good Research Practices Task Force. *Value Health* 2014;17:501–16.
 - 36) Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89–95.
 - 37) Schmitter-Edgecombe M, McAlister C, Weakley A. Naturalistic assessment of everyday functioning in individuals with mild cognitive impairment: the day-out task. *Neuropsychology* 2012;5:631–41.
 - 38) Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. *Int J Stroke* 2009;4:200–5.

Walton MK, Powers JH 3rd, Hobart J, Patrick D, Marquis P, Vamvakas S, Isaac M, Molsen E, Cano S, Burke LB; International Society for Pharmacoeconomics and Outcomes Research Task Force for Clinical Outcomes Assessment. Clinical Outcome Assessments: Conceptual Foundation-Report of the ISPOR Clinical Outcomes Assessment - Emerging Good Practices for Outcomes Research Task Force. *Value Health*. 2015 Sep;18(6):741-52. doi: 10.1016/j.jval.2015.08.006. の日本語訳である。翻訳と掲載にあたり、Copyright Clearance Center (License No: 4665761032931) の許諾を得た。

本成果物は、国立研究開発法人日本医療研究開発機構（AMED）革新的がん医療実用化研究事業 支持 / 緩和治療領域臨床試験に関する各分野における方法論確立に関する研究（研究開発代表者：国立がん研究センター東病院 放射線治療科 全田 貞幹，課題番号：JP21ck0106499）の支援を受けた。