

# 患者報告アウトカムに関する論文を読み活用するための 臨床家向けチェックリスト

## Clinician's Checklist for Reading and Using an Article About Patient-reported Outcomes

Albert W. Wu, MD, MPH, FACP; Anna N. Bradford, PhD, MSW, LCSW; Vic Velanovich, MD;  
Mirjam A.G. Sprangers, PhD; Michael Brundage, MD, FRCP, MSc; and Claire Snyder, PhD

訳 兼安貴子<sup>1,2)</sup> 鈴鴨よしみ<sup>3)</sup> 内藤真理子<sup>4)</sup> 宮崎貴久子<sup>5)</sup> 下妻晃二郎<sup>2,6)</sup>

### 要 旨

臨床家は自身の患者についての臨床的決定に際して根拠に基づく医療 (evidence-based medicine) の助けを必要とする。多くの健康問題において、治療目標は患者の機能が回復し、具合が良くなることを助けることである。患者の機能、ウェルビーイングや症状を測定するために、患者報告アウトカム (Patient-reported outcomes: PRO) 尺度と呼ばれる質問票がしばしば用いられる。一般に、臨床家は調査デザイン、評価尺度の開発、質問票による調査に関する訓練を受けていないため、PROを解釈し、臨床的エビデンスとして効果的に活用することは困難である。臨床家が自身の診療に関する情報を得るために、臨床的観点と患者の観点の双方から測定され

たアウトカムを理解し、活用できることがますます重要になっている。われわれの目的は、臨床家がPROを含む臨床研究論文を理解してその情報を意思決定に役立てることができるよう、「臨床家向けチェックリスト」を提供することである。このチェックリストでは、読者が研究論文を評価する際に考慮すべき重要な領域を箇条書きで示す。われわれは、臨床家がPROを用いた研究論文を読む際には以下の5つの要素を考慮することを提案する：研究デザインとPRO評価の戦略、PRO尺度の性能、結果の妥当性、結論の文脈、自身の患者集団への一般化可能性。PROは臨床研究や診療において、ますます重要な役割を果たすようになっており、この傾向は患者中心の医療 (patient-centeredness of care) を向上させる可能性がある。臨床家は患者のパートナー

原著者所属：Department of Health Policy and Management, Center for Health Services and Outcomes Research, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD (A.W.W.); Inova Fairfax Hospital Trauma Services, Falls Church, VA (A.N.B.); Division of General Surgery, University of South Florida, Tampa, FL (V.V.); Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands (M.A.G.S.); Division of Cancer Care and Epidemiology, Cancer Research Institute, Queen's University, and Cancer Centre of Southeastern Ontario, Kingston General Hospital, Kingston, ON, Canada (M.B.); Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD (C.S.)

原著者連絡先：Albert W. Wu, MD, MPH, FACP, Department of Health Policy and Management, Johns Hopkins University Bloomberg School of Public Health, 624 N Broadway, Rm 653, Baltimore, MD 21205 (awu@jhsph.edu) 宛

<sup>1)</sup>立命館大学 生命科学部 生命医科学科 <sup>2)</sup>立命館大学 総合科学技術研究機構 医療経済評価・意思決定支援ユニット (CHEERS) <sup>3)</sup>東北大学大学院医学系研究科障害科学専攻肢体不自由学分野 <sup>4)</sup>広島大学大学院医系科学研究科総合健康科学専攻口腔保健疫学研究室 <sup>5)</sup>京都大学大学院医学研究科社会健康医学系専攻 健康情報学分野 <sup>6)</sup>立命館大学 生命科学部

となり、患者の機能が回復し、具合が良くなることを支援するために、PROの使用方法を理解する必要がある。今回提案する臨床家向けチェックリストは、研究デザインが適切であり、測定方法が十分かつ適切に実施されたかどうかを判断し、結果の解釈や特定の患者集団への適用を支援することにより、

臨床家がPRO研究を系統的に評価することを支援するものである。

© 2014 Mayo Foundation for Medical Education and Research  
Mayo Clin Proc. 2014;89(5):653-661

患者は健康問題または健康維持に関する支援を求めて臨床家のもとを訪れる。健康問題には急性(例:新たに発症した肩の痛み)のものや慢性(例:糖尿病)のものがある。健康維持は予防と早期発見に重点を置いている。今日の多くの健康問題における治療目標は、患者の機能が回復し、具合が良くなることを助けることであり、機能やウェルビーイングの測定は患者中心ケアの重要な側面である<sup>1)</sup>。患者報告アウトカム(PRO)は質問票を用いた測定が可能であり、患者の観点から治療効果を評価する<sup>2)</sup>。米国食品医薬品局(FDA)はPROを次のように定義している。

PROとは、患者の回答について臨床家や他の誰の解釈も介さず、患者から直接得られる患者の健康状態に関するすべての報告である<sup>3)</sup>。

PROの例には、症状尺度(例:HIV Symptom Index<sup>4)</sup>)、機能状態(例:日常生活動作評価尺度<sup>5)</sup>)や健康関連QOL<sup>6)</sup>(例:36-item Short-Form Health Survey<sup>7)</sup>)がある。ランダム化臨床試験、コホート研究や効果比較研究にPROが含まれることが多くなっている<sup>8)</sup>。複数の治療選択肢の中で優れた治療法を示す際にPROが決定的な役割を果たすこともあれば<sup>9)</sup>、臨床的指標とPRO尺度の間で不一致が認められることもある(例:降圧薬が血圧を下げるとのみならず、副作用のために健康関連QOLを悪化させる場合)<sup>10,11)</sup>。いずれの場合においても、PROは治療の決定についての情報を提供する際に重要な役割を果たすことができる。

臨床家が自身の診療に関する情報を得るために、臨床的観点と患者の観点の双方から測定されたアウトカムを理解し、活用できることがますます重要と

なっている。しかし、患者ケアに関する情報を得るためにPROの結果を用いることに関心のある臨床家にとって、さまざまな障害が存在する。一般に、臨床家はその訓練期間中にPROの測定や解釈について教わる機会がないため、これらの方法にあまり馴染みがない。他のよく知らない検査と同様に、解釈や活用方法がわからないPRO尺度に対して臨床家は懐疑的または否定的となることがある。この分野での教育・訓練の欠如は、臨床家が臨床試験から得られたPROデータを自らの実地診療で活用することをさらに困難にする<sup>12)</sup>。その他の障害として、文献ではさまざまな種類の異なるPRO尺度や結論の報告様式が用いられており、この多様性が混乱に繋がり、臨床家によるPROの理解が損なわれる可能性がある<sup>13)</sup>。

1997年、Guyattら<sup>14)</sup>は、臨床家に向けて、健康関連QOL研究の活用方法についての影響力の大きな論文を書いた。それ以降、臨床研究、診療やPRO測定方法においていくつかの重要な変化が認められている。最初に、検討されるアウトカムは健康関連QOLから始まり、症状、機能面への影響、患者満足度等の明示的な領域まで拡大された。このPRO領域の増加は、考慮すべき質問票の増加に至った。2番目に、根拠に基づく医療(evidence-based medicine)の受容と適用は着実に増大しており、臨床家は確実な根拠のある研究方法の重要性をますます認識しつつある。しかし、臨床家にはPRO測定の科学的本質に関する教育が必要である。3番目に、収載医薬品数の増大に伴い、有効な治療を特定し、新薬に関連する症状や副作用を記録する目的でPROを測定する機会が増えている。4番目に、項目反応理論やコンピュータ適応型テスト等のPRO測定技術が進歩し、測定ツールの数が急増している<sup>15)</sup>。5

表1 PROを用いた研究を評価するための臨床家向けチェックリスト

1. PRO評価の戦略は適切であったか？	
a. PRO仮説は述べられているか？	PROに関する事前仮説が明示されている。
b. PRO評価は記述されているか？	用いたPRO尺度, 評価時期/被験者のフォローアップ
c. PROの内容は適切か？	研究者は患者が重要と考える患者の生活の側面を測定している。 PRO領域は疾患や治療の予想される影響と対応している。 PROのすべての重要な側面が含まれている。
2. PROは効果的に測定されたか？	
a. 信頼性や妥当性に関するエビデンスはあるか？	PRO尺度は意図どおりに機能したと考えられる。内的整合性, および/または再テスト信頼性, 構成概念妥当性のエビデンスが引用または十分に確立されている。
b. 欠測データの取り扱いが適切であったか？	全治療群の回答者が完了した質問票の数はすべての時点で同程度である。 欠測データの管理戦略についての記載がある。 死亡(が多い場合)の取り扱いに関するデータ解析計画がある。
3. 結果を信じるべきか？	
a. 内的妥当性	結論が確立されている。観察された効果は介入による可能性が高い。 治療以外の要因がPROに影響する場合, リスク調整が行われている。
4. 結果は臨床的文脈に置かれていたか？	
a. 結果の臨床的意味が説明されていたか？	PROに対する影響の大きさが記述されている。 PROスコアの観察された差の臨床的重要性が示されている。
b. 結果は私の患者ケアに役立つか？	生活の質と量の間の潜在的トレードオフを含む益と害が認識され, 調和されている。 結果に対して, 臨床家がどうすればよいか記述されている: 研究情報は臨床家が治療選択肢について患者とコミュニケーションする際に役立つ; 個々の患者に対する集団の結果の適用可能性。
5. 結果は私の患者に適用できるか？	
a. 臨床家の診療に対する外的妥当性	試験対象集団は臨床家が治療対象とする患者集団と十分類似している。

PRO = 患者報告アウトカム

番目に、近年では、効果比較研究に患者の観点を統合することが強調されており、患者中心アウトカム研究や患者中心ケアに焦点が当てられている<sup>16)</sup>。このことは、最近、米国において新たな Patient-Centered Outcomes Research Institute (患者中心アウトカム研究所) 設立のための医療関連法案が可決されたことから明確に示される<sup>17)</sup>。最後に、PROは電子カルテに組み込まれつつあり、臨床家がそれらを用いる機会が増大している<sup>8)</sup>。したがって、臨床家が、公表文献で用いられたPROの方法や報告されたPROの結論を評価できることがますます重要となってきた。

この「臨床家向けチェックリスト」の目的は、臨床家がPROを含む臨床研究論文の結果を自らの患者ケアに適用することを支援することである。われわれは、Guyattら<sup>14)</sup>の論文に基づき、臨床家がPROを含む研究調査をレビューすることを支援する方法と簡潔なチェックリストを提案する。併せて、その

ような論文の評価や活用方法の「具体例」も示す。

### 提案するチェックリスト

われわれは、PROを用いた公表研究を読む際には、以下の5つの要素を考慮することがきわめて重要であることを提案する：PRO評価の戦略と研究デザイン、PRO尺度の性能、結果の妥当性、結論の文脈、自身の患者への一般化可能性。これらは部分的にはGuyattらが公表したガイドライン<sup>14)</sup>に基づく。各要素の一般的な記述に続いて、臨床家が研究の適切性を判断するための「臨床家向けチェックリスト」中の設問を策定する(表1)。

#### PRO評価の戦略は適切であったか？

臨床研究の概念化やデザインにおいて重要な要素はPROを含む研究にも等しく当てはまる。リサー

チクエスチョン、研究デザイン、関心のある患者集団や主要アウトカムを特定し、PRO 選択の根拠を示すべきである。介入または治療が PRO のいくつかの面に影響を及ぼすことを示唆する既存のエビデンスはあるか？ その PRO 尺度は効果を評価できるのか？ 理想的には、PRO の開発やテストの段階で患者からの適切なインプットがあるべきで、そうでなければ、その PRO は患者にとって重要なものではなく、臨床家が重要であると考えられるものを反映した形で臨床家により開発されたものであろう。論文は主要および副次アウトカムと、これらが臨床、患者もしくは社会、いずれの観点から測定されるかを述べるべきである。PRO に関する仮説は事前に明示しなくてはならない。

初回評価やフォローアップ評価の時期を含む PRO 測定戦略について述べるべきである。この時期は、関心のある疾患状況における患者アウトカムの予想される経時的変化に関する知識と一致し、可能であれば、治療に関連した患者の状態の変化の時期に関する何らかの情報に基づくべきである。研究が治療前の QOL または症状重症度を評価し、仮説が要求する差を評価するために十分長いフォローアップ期間を設けることはきわめて重要である。PRO の内容は患者集団で観察される問題の程度や広がりに対応したものでなければならない。

これらの特性を評価するために、読者は最初に基本的な研究デザインが十分であり、研究により患者アウトカムに対する治療の影響を把握できる評価戦略であったかどうかを判断すべきである。症状や副作用のみを測定することへの圧力がしばしば存在するものの、患者の日常機能に対するこれらの症状の「影響範囲」を評価することが重要である。たとえば、第 II 相試験では症状に焦点が限定されるかもしれないが、第 III 相試験では患者の機能に対する治療効果をより包括的に評価すべきである。読者は PRO の重要な側面が省略されていないかどうかを、そのような省略は正しくない結論を導く可能性があるため、点検すべきである<sup>10)</sup>。

### PRO は効果的に測定されたか？

研究論文の検討に際して、読者は、研究仮説の検

証に用いた PRO 尺度が「有効に機能」したことを示唆する十分なエビデンスが引用されているかどうかを判断すべきである。患者の健康状態の効果的な測定は意味のある結論を導くうえで重要である。試験が群間差を識別、あるいは介入に関連した変化を特定するために、試験対象集団における PRO スコアの分布を観察しなくてはならない。方法の項では、臨床集団における PRO 尺度の内的整合性、再テスト信頼性、と構成概念妥当性に関するエビデンスを引用すべきである<sup>18,19)</sup>。質問票には「反応性がある」ことを示すエビデンス（例：患者の健康状態の経時的変化と相似するスコアの変化）を含めるべきである。さらに、PRO データの欠測がランダムな欠測である可能性は低いことから、著者らは欠測データの取り扱い計画の概要を述べるべきである。多くの死亡が予想される場合は、解析での死亡の取り扱い方法をあらかじめ計画すべきである。結果の項では欠測データの程度とパターンについて述べるべきである。上記の要素のいずれかがない場合、特に結論が治療効果や群間差がないことを示唆する場合には研究の結論を疑問視すべきである。

### 結果は信用できるか？

PRO の結果は明確に述べられなくてはならない。研究の内的妥当性、すなわち観察された効果が介入の結果である可能性が高いかどうかを立証すべきである。そのために、著者らはベースライン時の治療群の比較可能性を確立し、既知の交絡変数が測定されるようにすべきである。治療以外の要因が PRO スコアに影響を及ぼすことが知られている場合は、リスク調整のためのシステムを用いて、公正な群間比較を保証すべきである。治療効果の不均一性が予想される、重要な患者部分集団における結果は提示されるべきである<sup>20)</sup>。理想的には、これらの部分集団を事前に特定し、そうでない場合、結果は探索的とみなすべきである。

研究の内的妥当性を評価するために、読者は、観察された結果が他の要因ではなく介入による可能性が高いかどうか、リスク調整方針が成功裏に用いられたかどうか、そして最後に、効果が臨床的に妥当だと思えるかどうかを評価すべきである。

## 結果は臨床的文脈内に置かれているか？

PROの結果の臨床的意義について、変化が患者にもわかるほど十分大きいかどうか、治療の変更を促すものかどうか等を明示して考察すべきである。PROは、疾患や治療の正および負の影響や患者の日常機能に対する症状や副作用の「影響範囲」に関する包括的な情報を与える。介入が肯定的影響と否定的な影響の両方を有する場合、考察では益と害のバランスについて検討すべきである<sup>21)</sup>。このことは、生活の質と量の間トレードオフが存在する場合、たとえば治療により生存時間は延長するが、QOLは低下する(例：毒性の強い化学療法)場合等に重要である。これらの状況では、標準的賭け方や時間得失法のような健康関連QOLの選好に基づく尺度と評定尺度(EuroQol five-dimensional questionnaire<sup>22)</sup>, Quality of Well-being Scale<sup>23)</sup>)が有用である。これらの指標は病的状態と死亡の影響を単一のスコアに統合することができるからである。また、試験中の死亡が多い場合は、どの要因が他の要因に勝るかを患者に選択させることが重要である。試験のPROの結果を考えると、臨床家がどの治療選択肢を考慮すべきかが明らかな場合もあれば、明らかでない場合もあるが、著者らが推奨を述べることにより、試験の結論が翻訳され治療の変更につながる可能性が高まる。

読者は患者のQOLに対するその影響の大きさを特定し、患者ケアの変更を動機付けるほど十分大きいかどうかを判断すべきである。読者は研究の結論から示唆される益と害の潜在的なトレードオフについて考察すべきである。

## 自分の患者に当てはまるか？

臨床家が治療選択肢について患者と話し合おうとする場合、結論の外的妥当性は臨床家にとって重要である。読者は、研究がどの程度診療全般を模倣しており、結果が自分自身の患者集団に対して一般化可能かどうかを判断すべきである。理想的には、臨床家がこの作業を行うのを支援するために試験の著者はPROを含む研究結果の一般化可能性に触れるべきである<sup>13)</sup>。

## PRO 臨床家向けチェックリスト：

### 2 試験への適用例

例示として、公表論文2報にチェックリストを適用し、自分自身の診療のために臨床試験を評価し、その有用性を判断する際の使用方法を示す(表2と表3)。最初の例<sup>24)</sup>(表2)では、2種類の鼠径ヘルニア修復術を比較した試験を評価する。関心のある主要アウトカムはヘルニアの再発であった。副次アウトカム(疼痛、機能的状態)の測定にはPRO尺度、疼痛のビジュアルアナログスケール<sup>25)</sup>、と36-item Short-Form Health Survey<sup>7)</sup>を用いた。これらの尺度の内容、信頼性と妥当性のエビデンス、評価時期は適切と思われた。欠測データについての言及はなく、スコアの臨床的意味についての説明もなかったが、結果は妥当であると思われた。この試験によると、開腹術でのヘルニアの再発率は低かったものの、腔鏡下修復術の方が痛みは少なく、機能的状態の回復は早かった。しかし、欠測データが多く、有益な情報が含まれていた場合には、治療反応性が良好な健康な患者が含まれたために、両群の結果に偏りが生じた可能性があることに注意すべきである。

普遍的に合意された最良のアウトカムがないため、種々の機能的や心理的アウトカムについての詳細な検討は患者の価値を明快にし、決定を促すのに役立つ。この試験は、主要アウトカム(再発)に基づき開腹術の推奨を結論付けているが、個々の患者により害と益との評価は異なるかもしれない。この論文のエビデンスから、術後疼痛を有意に低下させ、通常活動により速やかに回復する、腹腔鏡手術を選択する者もいるであろう。傷病休暇日数が少ない個人にとっては、この情報が決定的となる可能性がある。

2番目の評価(表3)では、早期乳がん女性のためのグループエクササイズプログラムについてのランダム化比較試験<sup>26)</sup>を検討する。主要および副次アウトカム(QOL, 抑うつ, 感情)はどちらも以下のPRO尺度を用いて測定した: Functional Assessment of Cancer Therapy-General15, Beck Depression Inventory<sup>27)</sup>, Positive and Negative Affect Scale<sup>28)</sup>。この試験のデザインや実施は結論を十分に裏付けるものであり、この試験に組み入れられた女性の全体の機能とウェ

表 2 臨床家向けPROチェックリストを用いた試験<sup>24)</sup>の評価

1. PROの評価戦略は適切であったか?	
a. PRO仮説は述べられているか?	PRO仮説は明示的に述べられていない。含意：腹腔鏡下修復術は、開腹術よりも痛みが少なく、良好な機能的状態への回復が速やかである。
b. PRO評価は記述されているか?	痛み：VAS <sup>25)</sup> 、機能的状態または活動レベル：SF-36 <sup>7)</sup> PRO実施時期は一般に認められている術後回復の観点からはやや広めであったが、時点は適切と考えられる。腹腔鏡下ヘルニア修復術の根拠の一つは正常機能への速やかな回復と痛みの少なさであったことから、早期の測定は適切である。仮定した時点とマッチした測定時期と介入との関連：(1)ベースライン、(2)介入終了時、(3)6か月間フォローアップ時。
c. PROの内容は適切か?	尺度は痛み、機能的状態および活動レベルを測定した。必ずしも患者が重要と考えるすべての側面(例：睡眠、性的機能)を測定したわけではなかった。
2. PROは効果的に測定されたか?	
a. 信頼性や妥当性に関するエビデンスはあるか?	VAS <sup>25)</sup> およびSF-36 <sup>7)</sup> ともに、その信頼性は試験対象集団と類似の集団を含むさまざまな集団において広く検討されている。しかし、この試験での尺度の信頼性および妥当性についての報告はなかった。 両尺度は痛みおよび機能的状態の測定において十分に検証されている。VASは術後疼痛の評価のための妥当性および信頼性の高い尺度である。いずれも、鼠径部ヘルニアまたはその外科的修復における痛みまたは機能状態の測定に特化した尺度ではなかった。
b. 欠測データの取り扱いが適切であったか?	試験中のどの時点においても欠測データの割合についての報告はなかった。欠測データの取り扱い方法についての説明はなかった。死亡の取り扱いに関する規定はなかった。
3. 結果を信じるべきか?	
a. 内的妥当性	中止・脱落率が統計解析におよぼした影響は不明である。いくつかの比較は統計学的に有意ではなく、ベータ過誤は決定されていなかった。 多変量解析を用いて、層別変数(実施医療機関、ベースライン時の治療)、年齢、ベースライン時のアウトカムの値での調整を行った。
4. 結果は臨床的文脈に置かれていたか?	
a. 結果の臨床的意味が説明されていたか?	影響の大きさについての記述はあった。スコアの臨床的意味についての言及はなかった。介入の恩恵についての考察は行われた。開腹術群で疼痛が多く認められた以外に有害事象はなかったと報告された。
b. 結果は私の患者ケアに役立つか?	生活の質と量の間で特定されたトレードオフについての経済的評価は行われなかった。著者らは、読者が患者との意思決定を改善するために情報を用いる方法について提案している。
5. 結果は私の患者に適用できるか?	
a. 臨床家の診療に対する外的妥当性	この試験には男性患者のみが含まれ、一般集団よりもアフリカ系米国人の割合が高かった。自分自身の患者集団との関連性を評価する際には、これらの要因を考慮すべきである。

PRO = 患者報告アウトカム, SF-36 = 36-item Short-Form Health Survey, VAS = ビジュアルアナログスケール

ルビーイングへの影響はないことが示唆された。肩機能と歩行試験に基づく一部の結論からは運動の恩恵が示唆されたため、この研究は指導者付きのグループエクササイズを推奨した。しかし、クラスに通うための要件が被験者の介入への関心を減弱させた可能性を著者らは認めた。

## 考 察

PRO研究の質は、その結果として得られるPROデータの意思決定上の有用性に影響を及ぼす。これ

らのデータのエンドユーザーとして、臨床家はPROを用いた研究論文の有効活用に関するいくつかの困難に直面する。最初に、不適切な研究デザインと実施は結論を導き出すことを妨げる可能性がある。試験が重要なQOLの問題を見落とす可能性がある。尺度が十分機能しない場合がある。試験対象が実診療の多くの集団に一般化できない可能性がある。2番目に、試験対象集団についての限られた情報を提供すること等、方法と結果の提示は、臨床家を不利な状況に置くことがある。3番目に、PRO評価の実務にまつわるいくつかの問題は完全には解決

表3 臨床家向けPROチェックリストを用いた試験<sup>26)</sup>の評価

1. PROの評価戦略は適切であったか?	
a. PRO仮説は述べられているか?	仮説は以下のとおりであった:「12週間の指導者付きのグループエクササイズ...は早期乳がん治療中の女性のQOLを改善する可能性があり...恩恵は介入後6か月間維持される」。患者は介入(グループエクササイズあり)群または通常治療(グループエクササイズなし)群のいずれかにランダム化された。
b. PRO評価は記述されているか?	主要アウトカム評価としてFACT-General <sup>15)</sup> が示され、4つの下位尺度について述べられた。副次PRO評価はBDI <sup>27)</sup> およびPANAS <sup>28)</sup> であった。その他の身体活動の副次評価はBMI、12分間歩行試験、肩の可動性試験であった。FACTおよびその他のPRO尺度の開発および妥当性の検証に関する参考文献が示された。指標は一般的に「がん患者への使用に適切である」と述べられた。
c. PROの内容は適切か?	介入および仮説を考えると、データ収集時点(ベースライン、介入終了時、6か月後)は適切であった。 評価したアウトカムは簡潔な文献レビューにより裏付けられ、仮説に対し適切であった。
2. PROは効果的に測定されたか?	
a. 信頼性や妥当性に関するエビデンスはあるか?	この試験でのPROデータの信頼性および妥当性に関するデータは提示されなかった。これらの質問票の信頼性および妥当性はすでに確立されていた。
b. 欠測データの取り扱いが適切であったか?	割付および評価のフローチャートが提示された。介入群では対照群よりも追跡不能例が多かった。死亡は両群ともにまれな事象であった。 欠測データに対処する解析方法についての具体的な考察はなかった。
3. 結果を信じるべきか?	
a. 内的妥当性	著者らは主要仮説が支持されなかったことを認めている。結果の提示は運動と通常ケアに割り付けられた群の間で認められたアウトカムの違いに焦点が当てられた。主要エンドポイントを達成せず、どの特定のPROドメインが群間で異なると予想されるかについての規定がないことから、有意差( $P < .0001$ )および特定された傾向(例:12分間歩行、肩の可動性)の解釈は困難である。
4. 結果は臨床的文脈に置かれていたか?	
a. 結果の臨床的意味が説明されていたか? 結果は私の患者ケアに役立つか?	著者らは、運動プログラムのどの部分が恩恵と関連しているか、およびグループへの参加自体が有益であった可能性を判断することは困難であることを認めている。肩および歩行試験についての所見は身体的恩恵を裏付ける。
5. 結果は私の患者に適用できるか?	
a. 臨床家の診療に対する外的妥当性	先行研究では、運動が治療中および治療後の乳がんサバイバーのQOLの身体的および社会心理的側面を改善する可能性が示された。主要エンドポイントに関する統計学的に有意な結果がなく解釈は困難であるにもかかわらず、著者らはこの試験の所見のさまざまなポジティブな点を強調する。この試験ではグループによる運動プログラムを用いた。したがって、一般化可能性はグループエクササイズが実施可能な状況に限られる。「多数の」参加者がクラスに通う時間がなくクラスに出席できなかった。

BDI = Beck Depression Inventory, BMI:Body mass index, FACT = Functional Assessment of Cancer Therapy, PANAS = Positive And Negative Affect Scale, PRO = 患者報告アウトカム, QOL = Quality of life

していない。臨床家のためにスコアの意味を翻訳する能力は著しく進歩したものの、群間差や経時的変化に関するそれらの意味を解釈することは依然として難しい。われわれが直面する困難は、われわれの解釈を導くPROスコアからの蓄積データが臨床試験または観察研究からの群平均に基づく場合、個人のPROスコアにおける臨床的に意味のある変化を解釈し、そのスコアに基づいて治療計画を立案するための指針を適用する際に生じる。個人のスコアは

患者グループからの集成的なスコアよりも測定誤差が大きくなりやすい<sup>29)</sup>。

これらの欠点がときどきみられるものの、PROを用いた論文は、臨床家と患者がどの治療方針が最良のアウトカムをもたらす可能性があるかを決定する際に役立つ。本論文で提案する臨床家向けPROチェックリストは、臨床家がそのような研究を系統的に評価するのを支援し、結果を信じるに足る十分な厳格さをもって試験が実施されたかどうかを判断

し、結果を解釈して特定の患者集団に適用することに役立つ。チェックリストの目的は、臨床家が自分の患者と診療のために、PROを用いた研究の意義を評価するのを支援することである。試験結果が信用でき、PROエンドポイントが自身の患者集団にとって際立ったものであり、かつ結果の大きさが理解可能な場合、臨床家はPROの結果を他の臨床データや個々の患者の選好と統合する強力な立場にある。これにより、意思決定や、患者満足度、アウトカムが改善されるであろう。

Guyattら<sup>14)</sup>が公表した古典的な論文は直接実臨床家に向けたものであり、健康関連QOLに関する論文を読む際には以下の3つの問いを考慮することを提起している：「結果は妥当か?」「結果はどのようなものだったか?」「結果は私の患者のケアに役立つか?」。これらは、研究デザインや測定方法についてあまり強調していない。それ以降、臨床家は、患者との意思決定にQOLをより定期的に考慮に入れるようになった。FDAに代表される規制当局や学術団体は、PROを含む研究に標準化された厳格な方法を適用することの重要性を強調し、PROへの影響に関する事前仮説の組み立てから開始した<sup>3)</sup>。したがって、われわれの論文でもPROを用いた臨床研究のいくつかの方法論的側面を強調する。これらの研究は(1)仮説を明示的に述べ、(2)研究で用いるPROの開発とテスト段階での患者のインプットを記録し、(3)データ解析(特に欠測データが多い場合)においてバイアスを防ぐための方法について述べ、(4)PRO評価の結論の臨床的意味を説明すべきである。

いくつかの要因が、今後10年間での臨床研究におけるPROのより顕著な役割を予示している。最初に、将来の典型的な成人は50歳を超え、複数の慢性疾患を抱えて生活している<sup>30)</sup>。これらの個人に対して、患者中心のアウトカム研究は最適な水準の機能とウェルビーイングを達成するための最良の戦略を特定する。したがって、PROを用いた研究の公表数が増加する可能性が高い。また、システマティックレビューにおけるPROに対する注目も高まっている<sup>31,32)</sup>。2番目に、特に定性的および定量的方法の両方を用いたPRO測定における進歩と標準化の進展があった<sup>33-35)</sup>。患者報告アウトカム測定情報シ

テム(Patient-Reported Outcomes Measurement Information System: PROMIS)イニシアティブは、さまざまな疾患や健康状態にわたってPRO評価(例:痛み、疲労、身体機能、抑うつ、不安)を標準化し<sup>33)</sup>、さまざまな疾患状態での使用や解釈に適切な指標をもつための国家的努力の一つの例である。3番目に、診療録のデジタル化や患者報告データを収集できる患者ポータルの開発<sup>8)</sup>に伴い、ますます多くのPROデータが臨床家に提示されている<sup>8)</sup>。

臨床的意思決定におけるPROエビデンスの使用を促進するためには、多くのことを達成する必要がある。最初に、エビデンスの基礎を強化するために、研究者や資金提供者はより高品質のPRO研究の報告を求めなくてはならない。重要な最近の努力には、ランダム化臨床試験におけるPRO報告標準の開発や臨床試験報告に関する統合基準(Consolidated Standards of Reporting Trials: CONSORT)を拡張し、PROを含めるプロジェクトがある<sup>36,37)</sup>。現在、2009年に公表されたFDA文書のように、研究者による優れたPRO研究の実施を支援するためのいくつかのリソースが利用可能である<sup>38)</sup>。2番目に、健康測定尺度の選択のためのコンセンサスに基づく標準(Consensus-based Standards for the selection of health Measurement Instruments: COSMIN)イニシアティブ<sup>38)</sup>が編纂したようなコンセンサスの得られた用語の定義が必要である。3番目に、政策立案者は臨床試験やその他の形態の研究にPROを含めることへのサポートを継続しなければならない。最後に、医療研究のための従来のデータ情報源とPROとの間のリンクを探し求めなくてはならない<sup>8)</sup>。臨床的観点と患者の観点の双方から治療効果に関する情報を提供することは、最終的には意思決定プロセスを促すであろう。

## 結 論

臨床家は、PROについての情報を自らの診療に組み入れ、それを個々の患者のための治療法の決定に適用する方法に関する教育・訓練を必要とする。臨床家向けチェックリストのようなツールは、PROを用いた研究を利用する機会がもっとも多い人たちがよりアクセスしやすいものにするのに役立つかも

しれない。

## 謝 辞

本論文は国際 QOL 研究学会 (ISOQOL) のメンバーによるレビューを受け、ISOQOL 理事会による承認を受けた。本論文で述べたツールとメカニズムは ISOQOL が承認しているものではない。提示した尺度やプロジェクトは例示であり、網羅的な一覧ではない。本稿を読み、有益な改善点を指摘して頂いた 2 名の匿名の査読者にも謝意を表す。

**略語および頭辞語:** FDA=U.S. Food and Drug Administration, ISOQOL = International Society for Quality of Life Research, PRO = Patient reported outcome, QOL = Quality of life

## 文 献

- 1) Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: Committee on Quality of Health Care in America, Institute of Medicine; 2001.
- 2) Acquadro C, Berzon R, Dubois D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. Value Health. 2003;6(5):522-531.
- 3) Food and Drug Administration. Guidance for Industry, Patient- Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Silver Spring, MD: Office of Communications, Division of Drug Information Center for Drug Evaluation and Research Food and Drug Administration; 2009.
- 4) Justice AC, Holmes W, Gifford AL, et al. Development and validation of a self-completed HIV symptom index. J Clin Epidemiol. 2001;54(Suppl):S77-S90.
- 5) Katz S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. J Am Geriatric Soc. 1983;31(12):721-727.
- 6) Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med. 1993;118(8):622-629.
- 7) Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36), I: conceptual framework and item selection. Med Care. 1992;30(6):473-483.
- 8) Wu AW, Snyder C, Clancy CM, Steinwachs DM. Adding the patient perspective to comparative effectiveness research. Health Aff (Millwood). 2010;29(10):1863-1871.
- 9) Blazeby JM, Avery K, Sprangers M, Pikhart H, Fayers P, Donovan J. Health-related quality of life measurement in randomized clinical trials in surgical oncology. J Clin Oncol. 2006;24(19):3178-3186.
- 10) Contopoulos-Loannidis DG, Karvouni A, Kouri L, Loannidis JP. Reporting and interpretation of SF-36 outcomes in randomized trials: systematic review. BMJ. 2009;338:a3006.
- 11) Wilson IB, Cleary PD. Linking clinical variables with health related quality of life: a conceptual model of patient outcomes. JAMA. 1995;273(1):59-65.
- 12) Sprangers MA. Disregarding clinical trial-based patient-reported outcomes is unwarranted: five advances to substantiate the scientific stringency of quality-of-life measurement. Acta Oncol. 2010;49(2):155-163.
- 13) Brundage M, Bass B, Davidson J, et al. Patterns of reporting health-related quality of life outcomes in randomized clinical trials: implications for clinicians and quality of life researchers. Qual Life Res. 2011;20(5):653-664.
- 14) Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. Users' guides to the medical literature, XII: how to use articles about health-related quality of life. Evidence-Based Medicine Working Group. JAMA. 1997;277(15):1232-1237.
- 15) Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol. 1993;11(3):570-579.
- 16) Ahmed S, Berzon RA, Revicki DA, et al; International Society for Quality of Life Research. The use of patient-reported outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. Med Care. 2012;50(12):1060-1070.
- 17) 111th US Congress. The Patient Protection and Affordable Care Act (ACA). Pub L No. 111-148. March 23, 2010.
- 18) Haywood KL, Garratt AM, Fitzpatrick R. Quality of life in older people: a structured review of generic self-assessed health instruments. Qual Life Res. 2005;14(5):1651-1668.
- 19) Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22(8):1889-1905.
- 20) Weiss CO, Segal JB, Varadhan R. Assessing the applicability of trial evidence to a target sample in the presence of heterogeneity of treatment effect. Pharmacoepidemiol Drug Saf. 2012;21(Suppl 2):121-129.
- 21) EuroQol Group. EuroQolea new facility for the

- measurement of health-related quality of life. *Health Policy*. 1990;16(3):199-208.
- 22) Kaplan RM, Anderson JP, Wu AW, Mathews WC, Kozin F, Orenstein D. The Quality of Well-being Scale: applications in AIDS, cystic fibrosis, and arthritis. *Med Care*. 1989;27(3Suppl):S27-S43.
  - 23) Torrance GW. Utility approach to measuring health-related quality of life. *J Chronic Dis*. 1987;40(6):593-603.
  - 24) Neumayer L, Giobbie-Hurder A, Jonasson O, et al. Open mesh versus laparoscopic mesh repair of inguinal hernia. *N Engl J Med*. 2004;350:1819-1827.
  - 25) Agency for Health Care Policy and Research. *Acute Pain Management: Operative or Medical Procedures and Trauma (Clinical Practice Guideline)*. Bethesda, MD: Acute Pain Management Guideline Panel; 1992.
  - 26) Mutrie N, Campbell AM, Whyte F, et al. Benefits of supervised group exercise programme for women being treated for early stage breast cancer: pragmatic randomised controlled trial. *BMJ*. 2007;334:517.
  - 27) Beck A, Steer R, Brown G. *Manual for the Beck Depression Inventory-2*. San Antonio, TX: Psychological Corporation; 1996.
  - 28) Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol*. 1988;54(6):1063-1070.
  - 29) Donaldson G. Patient-reported outcomes and the mandate of measurement. *Qual Life Res*. 2008;17(10):1303-1313.
  - 30) Anderson G. Medicare and chronic conditions. *N Engl J Med*. 2005;353(3):305-309.
  - 31) Feeny DH, Eckstrom EN, Whitlock EP, Perdue LA. *A Primer for Systematic Reviewers and Others on the Measurement of Functional Status and Health-Related Quality of Life in Older Adults*. Agency for Healthcare Research and Quality; 2013.
  - 32) Johnston BC, Patrick DL, Busse JW, Schnunemann HJ, Agarwal A, Guyatt GH. Patient-reported outcomes in meta-analyses part 1: assessing risk of bias and combining outcomes. *Health Qual Life Outcomes*. 2013;11(1):109.
  - 33) Cella D, Riley W, Stone A, et al; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63(11):1179-1194.
  - 34) Liu H, Cella D, Gershon R, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemiol*. 2010;63(11):1169-1178.
  - 35) Rothrock NE, Hays RD, Spritzer K, Yount SE, Riley W, Cella D. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2010;63(11):1195-1204.
  - 36) Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD; CONSORT PRO Group. Reporting of patient reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA*. 2013;309(8):814-822.
  - 37) Brundage M, Blazeby J, Revicki D, et al. Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards. *Qual Life Res*. 2013;22(6):1161-1175.
  - 38) Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-549.

Wu AW, Bradford AN, Velanovich V, et al. Clinician's checklist for reading and using an article about patient-reported outcomes. *Mayo Clin Proc*. 2014 May;89(5):653-61. doi:10.1016/j.mayocp.2014.01.017. の日本語訳である。翻訳と掲載にあたり、Copyright Clearance Center (License No: 5304631355547) の許諾を得た。

本成果物は厚生労働科学研究費補助金「関連学会の取組と連携した PRO ガイドラインの作成 (研究代表者: 立命館大学 下妻晃二郎, 課題番号: 20AC1003)」の支援により作成した。