

がんのランダム化比較試験におけるQOLおよび 患者報告アウトカム評価項目の解析についての国際標準 —SISAQOLコンソーシアムの推奨—

International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomized controlled trials: recommendations of the SISAQOL Consortium

Corneel Coens*, Madeline Pe*, Amylou C Dueck, Jeff Sloan, Ethan Basch, Melanie Calvert, Alicyn Campbell, Charles Cleeland, Kim Cocks, Laurence Collette, Nancy Devlin, Lien Dorme, Hans-Henning Flechtner, Carolyn Gotay, Ingolf Griebesch, Mogens Groenvold, Madeleine King, Paul G Kluetz, Michael Koller, Daniel C Malone, Francesca Martinelli, Sandra A Mitchell, Jamme Z Musoro, Daniel O'Connor, Kathy Oliver, Elisabeth Piauxt-Louis, Martine Piccart, Chantal Quinten, Jaap C Reijneveld, Christoph Schürmann, Ashley Wilder Smith, Katherine M Soltys, Martin J B Taphoorn, Galina Velikova, Andrew Bottomley
(Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium)

訳 星野絵里¹⁾ 堺琴美¹⁾ 兼安貴子^{1,2)} 下妻晃二郎^{1,3)}

¹⁾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) ²⁾立命館大学 生命科学部 生命医科学科 ³⁾立命館大学 生命科学部

Lancet Oncol 2020; 21: e83-96

*共同筆頭著者

European Organisation for Research and Treatment of Cancer, Brussels, Belgium (C Coens MSc, M Pe PhD, L Collette PhD, L Dorme MSc, F Martinelli MSc, J Z Musoro PhD, A Bottomley PhD); Alliance Statistics and Data Center, Mayo Clinic, Scottsdale, AZ, USA (A C Dueck PhD); Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN, USA (J Sloan PhD); Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA (E Basch MD); Centre for Patient Reported Outcomes Research, Institute of Applied Health Research and National Institute for Health Research Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK (Prof M Calvert PhD); Patient Relevant Evidence, San Francisco, CA, USA (A Campbell MPH); Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, TX, USA (Prof C Cleeland PhD); Adelphi Values, Bollington, Cheshire, UK (K Cocks PhD); Centre for Health Policy, School of Population and Global Health, University of Melbourne, Melbourne, VIC, Australia (N Devlin PhD); Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Magdeburg, Magdeburg, Germany (Prof H-H Flechtner MD); School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada (Prof C Gotay PhD); Boehringer Ingelheim International GmbH, Ingelheim, Germany (I Griebesch PhD); Department of Public Health, Bispebjerg Hospital and University of Copenhagen, Copenhagen, Denmark (Prof M Groenvold MD); School of Psychology, University of Sydney, Sydney, NSW, Australia (Prof M King PhD); Office of Hematology and Oncology Products, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA (P G Kluetz MD); Center for Clinical Studies, University Hospital Regensburg, Regensburg, Germany (M Koller PhD); College of Pharmacy, University of Arizona, Tucson, AZ, USA (Prof D C Malone PhD); Outcomes Research Branch, Healthcare Delivery Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA (S A Mitchell PhD, A W Smith PhD); Medicines and Healthcare products Regulatory Agency, London, UK (D O'Connor MBChB); International Brain Tumour Alliance, Surrey, UK (K Oliver BA); Genentech, Roche Group, San Francisco, CA, USA (E Piauxt-Louis PharmD); Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium (M Piccart MD); European Centre for Disease

Methods Section, Stockholm, Sweden (C Quinten MSc); Department of Neurology and Brain Tumor Center, VU University Medical Center, Amsterdam, Netherlands (J C Reijneveld MD); Institute for Quality and Efficiency in Health Care, Cologne, Germany (C Schürmann PhD); Health Canada, Ottawa, ON, Canada (K M Soltys PhD); Department of Neurology, Leiden University Medical Center, Leiden (M J B Taphoorn MD); Department of Neurology, Haaglanden Medical Center, The Hague, Netherlands (M J B Taphoorn); Leeds Institute of Cancer and Pathology, University of Leeds, St James's Hospital, Leeds, UK (Prof G Velikova PhD); and International Society for Quality of Life Research, Milwaukee, WI, USA (Prof G Velikova)

連絡先: Dr Madeline Pe, European Organisation for Research and Treatment of Cancer, 1200 Brussels, Belgium

madeline.pe@eortc.org

付録はオンライン版を参照のこと。SISAQOL Consortiumの詳細については、<https://www.sisaqol-imi.org/>を参照のこと。

がんのランダム化比較試験 (RCT) では、治療のリスク、ベネフィット、および忍容性についての情報を提供するために、症状、機能、およびその他の健康関連 QOL の側面等の患者報告アウトカム (PRO) を評価することが多くなっている。しかし、専門家の見解および批判的な文献レビューからは、がんの RCT では最適な PRO の解析についてのコンセンサスが存在しないため、結果の解釈が困難であることが示された。PRO の解析のための推奨を確立するために、PRO および QOL 評価項目データの解析における国際標準策定 (Setting International Standards in Analyzing Patient-

Reported Outcomes and Quality of Life Endpoints Data) コンソーシアムが設立された。以下の 4 つが優先課題とされた：適切な統計手法に合致した研究目的の分類法の開発、PRO の解析のための適切な統計手法の特定、欠測データに関連する統計用語の標準化、欠測データの適切な取扱い方法の決定。このポリシーレビューでは、推奨の基礎となる批判的な文献レビューおよび多様な国際的ステークホルダーとの構造化された共同プロセスを通して開発された PRO の解析のための推奨を提示し、これらの推奨に関する進行中の議論についても考察する。

緒言

がんの臨床試験において患者報告アウトカム (PRO) を使用することにより、患者の声をがん治療のリスク・ベネフィット評価に組み込むことができる。また、それは患者、医療提供者、医療費支払者、および規制当局による意思決定を促進する¹⁻³⁾。現在、PRO はがんの臨床試験において頻繁に収集されているが、系統的レビューからのエビデンスでは、標準化されたコンセンサスがなく、PRO データの解析および解釈方法に関するガイドラインは不明確であることが示された^{4,6)}。このため、PRO の調査結果から導かれた結論を評価することは困難である⁷⁾。試験実施計画書における PRO 報告を改善するための推奨 (標準的な試験実施計画書項目：介入試験のための推奨-PRO 拡張版 [Standard Protocol Items: Recommendations for Interventional Trials-PRO extension]：SPIRIT-PRO)⁸⁾、および公表における PRO 報告を改善するための推奨 (臨床試験報告に関する統合基準-PRO 拡張版 [Consolidated Standards of Reporting Trials Statement-PRO extension]：CONSORT-PRO)⁹⁾ は存在するものの、特にがん治療のための資源がさらに制限され、治療費が増大する段階においては、患者の安全性、治療選択、および方針決定に対して、信頼性の高い情報を確実に提供するために、優れた方法論的実践により PRO の調査結果を入手し、試験を通して一貫した評価を行うことが重要である¹⁰⁾。このニーズに応えるた

めに、PRO および QOL 評価項目データの解析における国際標準策定 (Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data: SISAQOL) コンソーシアムが設立された⁷⁾。SISAQOL コンソーシアムはグローバルなマルチ・ステークホルダー・コンソーシアムであり、PRO 専門家、統計家、規制当局、ならびに国際学会、業界、がん研究機関、および患者団体の代表者から構成される。このポリシーレビューでは、がんのランダム化比較試験 (RCT) における PRO の解析のための一連のコンセンサス推奨を提示し、以下の 4 つの重要な優先課題を取り上げる¹¹⁾：適切な統計手法に合致した研究目的の分類法の開発、特定の PRO 研究目的について検討するための適切な統計手法の特定、欠測データに関連する統計用語の標準化、欠測データの適切な取扱い方法の決定。

推奨の開発

専門家およびマルチ・ステークホルダー・パネルの選択

SISAQOL による推奨に至った過程の概要を図 1 に示す。本論文の著者 2 名 (AB および CCo) が、がんの RCT で PRO に経験を有する専門家およびステークホルダーを招待した。その目標は国際的なマルチ・ステークホルダー・コンソーシアムを設立することであった。SISAQOL が異なる専門分野を代表する広範な国際的グループとなるように、専門家

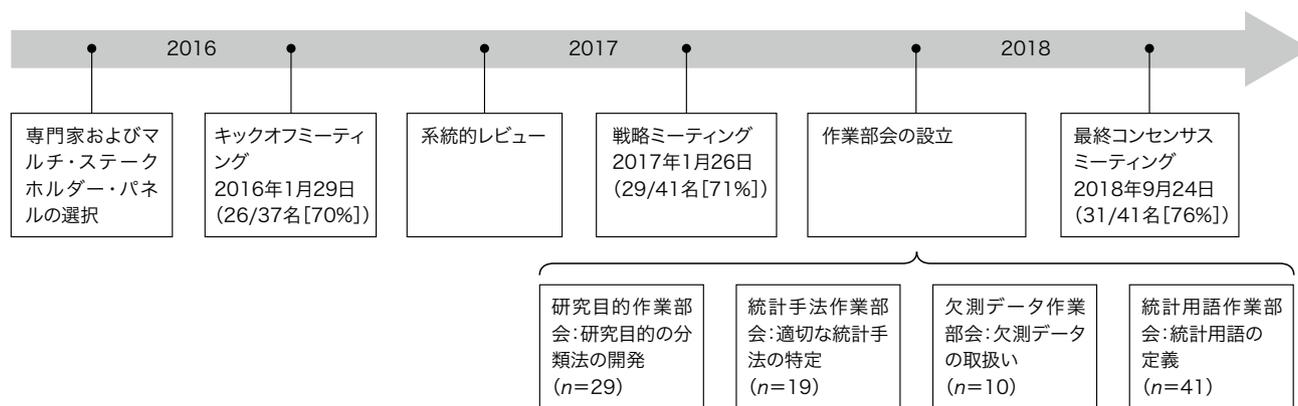


図1 SISAQOL 推奨の開発の概要

招待されたメンバーのうちでミーティングに出席したコンソーシアムメンバー数を示す。欠席者には完全なミーティング報告書が送付され、コメントおよび提案を追加することができた。SISAQOLコンソーシアムが報告書最終版を承認した。n=作業部会のメンバー数、SISAQOL=PROおよびQOL評価項目データの解析における国際標準策定 (Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data)

に相談し、同僚を推薦するように依頼した。代表者を募るために、年2回開催の欧州がん研究・治療機構 (European Organisation for Research and Treatment of Cancer: EORTC) QOL グループミーティングや国際学会 (例: 国際 QOL 研究学会 [International Society for Quality of Life Research], 国際医薬経済・アウトカム研究学会 [International Society for Pharmacoeconomics and Outcomes Research], 米国臨床腫瘍学会 [American Society of Clinical Oncology], 欧州臨床腫瘍学会 [European Society for Medical Oncology]) 等の主要イベントやミーティングの場で構想を説明した。要請があった場合は、EORTC と国際学会の間で了解書を締結した。これらの推奨の開発のすべての段階において招待された専門家の専門性およびプロフィールを付録 (p 1) に示す。

専門家の見解および系統的レビュー

2016年1月29日のSISAQOLキックオフミーティングには26名の専門家とステークホルダーが出席し、がんのRCTにおけるPROの解析の課題について議論を行った。国際標準が存在せず、この作業は早急に必要とされているとの合意に達した⁷⁾。異なるがん種のRCTにおけるPROの解析の現状を調査した系統的レビューの結果はこの見解を裏付けるものであった⁴⁻⁶⁾。以下の4つの重要な調査結果が強調された: 特定のPRO仮説がないこと、さまざまな解析方法の使用、PRO調査結果の臨床的意義に

対処できないこと、欠測データの無視。これらの調査結果は、試験実施計画書にPROを含めること¹²⁾ および公表論文でのPRO報告¹³⁻¹⁷⁾ について評価した系統的レビューの結果とも一致していた。

戦略ミーティング

2017年1月26日の戦略ミーティングには29名の専門家とステークホルダーが出席した。収集されたエビデンスに基づき、がんのRCTにおけるPROの解析のための国際標準が存在しないことが合意された。以下の核心的問題が特定された: 現行のPRO目的と仮説は広範囲に及び、PROの解析のための情報として価値がない傾向がある。そのため、コンソーシアムは以下の4つの重要な優先課題に焦点を置くことに同意した: 適切な統計手法に合致した研究目的の分類法の開発、特定のPRO研究目的について検討するための適切な統計手法の特定、欠測データに関連する統計用語の標準化、欠測データの適切な取扱い方法の決定。

作業部会

合意された優先課題に基づき、以下の4つの作業部会が設けられた: 研究目的、統計手法、統計用語標準化 (当初は、欠測データの定義と評価に焦点を置く)、欠測データの取扱い¹¹⁾。各作業部会は特定の目標と方法 (付録 pp 2-3) を設定した。各作業部会からの最終成果物をSISAQOL推奨のために提案する声明として用いた。

研究目的作業部会

系統的レビューの結果、がんのRCTにおいて十分定義されたPROの研究仮説がないことが一貫して示された^{5,6,12,15,17}。十分定義されたPROの仮説は試験目的に明確に沿っており、PROデータから推定する必要があるものの明確な理解を提供することにより、適切な解析に関する決定に情報を提供すべきである。研究目的作業部会のメンバーは、使用する統計手法に情報を提供することができるPRO研究目的のためのフレームワーク（すなわち、PRO研究目的の分類法）の開発、および重要なPRO目的のための標準化された定義作りを担当した。議論を通じて、最初のフレームワークが作成された。このフレームワークを研究目的作業部会の全メンバーに回付し、さらなる改善を依頼した。作業部会メンバー内で調査を行い、重要なPRO研究目的の定義（改善、増悪、安定状態）の標準化を行った（付録 pp 4-12）。

統計手法作業部会

系統的レビューの調査結果からは、PROデータの解析のための適切な統計手法についてのコンセンサスがないことが示された^{4,6}。さらに、PROの解析に関連するすべての臨床、試験デザイン、および解析上の問題に対処できる単一の解析手法はなかった。より科学的に妥当な方法の選択を行うためには、PROの解析のための統計手法の評価基準を設けることが必須であることが合意された¹¹。

文献検索および専門家の討議を通じて、19の統計学的基準の一覧を作成した。統計手法作業部会のメンバー内で調査を行い、提案された各統計学的基準を、がんの臨床試験におけるPROの解析において「必須 (essential)」「望ましい (desirable)」、または「必須ではない (non-essential)」のいずれかに分類した。その他の基準を収集するために自由回答形式の質問も含めた。個々の懸念がすべて取り上げられるまで、調査結果についての議論を行い、一連の基準を更新した（付録 pp 13-15）。

統計手法作業部会は、合意された一連の統計学的基準を用いて、転移性乳癌の系統的レビューにおいて特定された統計手法の最初の一覧を評価した⁵。統計手法の評価についての報告書案を統計手法作業部会のメンバーに回付し、レビューを依頼した（付

録 pp 16-26）。作業部会メンバーにより表明された個々の懸念がすべて取り上げられるまで、各PRO目的のために推奨される方法について議論し、修正を行った。

統計用語標準化作業部会

患者は（事前に定めた）疾患進行、死亡、忍容できない毒性、患者または臨床医の判断等のさまざまな理由により試験を中止・脱落することから、欠測PROデータはがんの臨床試験における現在進行中の課題である¹⁸⁻²⁰。多数の患者がPRO評価を完了できない場合、PRO推定値にバイアスが生じる可能性があるため、この問題の範囲を評価するために、試験における欠測データの割合を標準化された方法で評価すべきである²¹。しかし、欠測データの定義は不明確で、曖昧である。たとえば、試験実施計画書の検査スケジュールどおりの投与が期待できない場合、患者が疾患進行により試験を中止・脱落した後の観察されなかった評価が真に欠測データかどうかは明らかではない。したがって、この作業部会の目標は、欠測データの定義と欠測データの報告方法を標準化し、PRO試験対象集団（同意が得られ、PROデータ収集への参加に適格であったすべての患者）およびPROの解析対象集団（主要PRO解析に含まれる患者）との関係を明確化することであった。最初の一連の欠測データの定義および算出方法は転移性乳癌のRCTの系統的レビューから抽出した⁵。査読付きの公表論文の探索的な文献検索を行い、欠測データの他の定義および欠測データの割合の算出方法を特定した。コンソーシアムメンバーはこれらの定義を標準化するための調査に回答した（付録 pp 27-29）。調査結果について議論を行い、本作業部会からの個々の懸念がすべて取り上げられるまで、繰り返し修正を行った。

欠測データ作業部会

欠測データ作業部会の作業内容は、シミュレーション研究に基づき、欠測データの受け入れ可能な割合についての閾値（例：欠測データがどの割合で認められた場合に過剰であるとみなすか）を設定することが可能かどうかを特定し、PROの未完了理由を特定するための標準化された症例報告書を開発し、欠測データを取り扱うための全般的方針を推奨し、感度分析のためのさまざまな欠測データの設定

用のマクロの検証を行うことであった。

モンテカルロシミュレーションを行い、通常の RCT において欠測データの割合が増加した場合、バイアスおよび検出力にどのような影響が生じるかを評価した。シミュレーション結果は、今後の欠測データの閾値に関する推奨の基礎とすることを計画した²²⁾。

可能性のある PRO 未完了理由を備えた標準化された症例報告書の開発を試みるために、7つの異なる臨床試験ネットワークから既存の症例報告書テンプレートを収集した (例: Alliance for Clinical Trials in Oncology の症例報告書は既に公表されていた²³⁾)。27 の PRO 未完了理由からなる最初の一覧を収集した。コンソーシアムの全メンバー内で調査を行い、各未完了理由を標準症例報告書に含めるべきかどうか、患者の健康と関連があるか、データ品質に影響を及ぼすかを示した (付録 pp 30-31)。

SISAQOL 推奨ミーティング

2018 年 9 月 24 日の SISAQOL 推奨ミーティングには 31 名の専門家とステークホルダーが出席した。このミーティングの目的は、異なる作業部会により提案された声明を採択することであった。ミーティングは各作業部会を代表する以下の 4 セッションに分けられた: 研究目的の分類法, 統計手法の推奨, 欠測データに関連する用語の標準化, 欠測データの取扱い。

各声明について、出席者は賛成、反対、または棄権に投票した。声明に同意した投票者が 2/3 以上であった場合は、提案された声明を採択し、声明に同意した投票者が半数未満であった場合は、声明を却下した。声明が合意または却下の基準に合致しない場合、もしくはコンソーシアム内によるさらなる協議が必要であるとの合意が得られた場合は、当該声明を延期または継続協議とした。声明が以前の声明の採択を条件とし、以前の声明が採択されなかった場合は、当該声明を撤回した。特定の声明について棄権または投票しなかった出席者は総投票数には含まなかった。

SISAQOL 推奨およびそれらの検討項目を図 2 に示す。この表はこれらの推奨の簡潔な概要を示す。採択されなかった声明を採択されなかった理由とと

もに付録 (pp 35-36) に示す。

SISAQOL の推奨

推奨ミーティングでは 43 件の声明が提示され、そのうち 32 件 (74%) が採択, 8 件 (19%) が延期, 1 件 (2%) が却下, 2 件 (5%) が撤回された。提案されたすべての声明についての投票結果を付録 (pp 37-40) に示す。

研究目的の分類法

研究目的作業部会による提案された声明はすべて採択された (9/9 [100%])。がんの RCT のための PRO 研究目的の分類法が推奨された。このフレームワークは適切な統計手法に合致し、十分定義された PRO 評価を目的とした開発を支援するためのものである。このフレームワークの概要を以下の表に示す。

PRO 評価の目的の開発に際して、PRO ドメインおよび関心のあるタイムフレームは事前に定めるべきであるとコンソーシアムは結論付けた^{8,34)}。本質的に、各 PRO ドメインについて、以下の 4 つの重要な特性を事前に考慮する必要がある: 広範な PRO の研究目的 (治療効果および臨床的ベネフィット [検証的]), 記述的な患者の観点 [探索的または記述的]), 優越性, 同等性, または非劣性からなる群間での PRO の目的, 治療群または対照群のための治療群内での PRO の仮定 (増悪, 安定状態, 改善, または全体的効果等), 患者内または治療内での PRO の目的 (イベント発生までの時間, 時点 t でのイベントの程度, 時点 t でのレスポンドの割合, 全体的な PRO スコアの推移, 奏効パターンまたはプロファイル)。

推奨された声明 1~5 における各特性の検討項目を図 2 に示す。改善, 安定状態, 増悪, および全体的効果についての標準化された定義の推奨が採択された (推奨された声明 6~9 [図 2])。改善, 安定状態, および増悪の推奨される定義を図 3 に示す。

推奨される統計手法

このセクションで提案された声明のほとんどが採択された (6/7 [86%])。PRO の解析のための適切

推奨された声明	検討項目
セクション1：研究目的の分類法	
<p>RS 1 関心のある各PROドメインまたは項目についての広範なPRO研究目的を明示する：</p> <ul style="list-style-type: none"> ・治療効果または臨床的ベネフィット ・探索的または記述的な患者の観点 	<p>治療効果または臨床的ベネフィット 正式な治療群間の比較に関する結論を得るためにPROドメインを用いる場合は、以下の検証的ための規則に従う：各PROドメインについて事前の仮説が必要であり、次に試験終了時に統計学的な仮説検定を行う²⁴⁾。複数のPROドメイン、または1つのPROドメインの複数の評価時点に関心がある場合、多重検定の調整が必要である。十分定められた事前のPRO仮説の構成要素の詳細はRS 2～5に示す。</p> <p>探索的または記述的な患者の観点 試験中の患者の観点を記述するため、またはPROデータを探索し、将来の試験の情報源としてその調査結果を用いるためにPROドメインを用いる場合は、以下の記述的または探索的のための規則に従う：当該PROドメインについての事前の仮説は不要である。しかし、比較に関する結論を導き出すため、もしくは治療効果または臨床的ベネフィットを裏付けるために、これらのアウトカムを用いることはできない。調査結果は記述的(推定値の要約[信頼区間の有無は問わず]は行うが、統計学的検定は行わない)または探索的(仮説選択はデータドリブンであり、統計学的検定を行うこともある)のいずれかとして報告すべきであるが、臨床的ベネフィットまたは治療効果を示すエビデンスの基礎としてこの選択肢を用いてはならない²⁴⁾。</p> <p>上記2つのPRO目的は重要であり、互いに補い合うものであり²⁵⁾、同時に試験に含めることができる。ただし、試験実施計画書には、どのPROドメインを用いて治療効果または臨床的ベネフィットのエビデンスを提供するのか、患者の視点を記述するのか、あるいは探索的であるのかを明記すべきである。</p>
<p>RS 2 関心のある各PROドメインまたは項目について用いる治療群間比較を明示する：</p> <ul style="list-style-type: none"> ・優越性 ・同等性または非劣性 	<p>優越性評価のデザインおよび解析方法は同等性または非劣性の方法とは異なる^{24,26)}。治療間の差を評価することを目的とした統計学的検定(優越性検定)のp値が非有意であったことは、両治療群が類似している(同等性)または劣らない(非劣性)ことのエビデンスとして用いるべきではない。</p> <p>優越性 優越性評価のためのPROの目的は、事前に指定したPROドメインについて、治療群が対照群よりも臨床的に意味のある効果量の分だけ優れていることを示すことを意図する。臨床的に意味のある治療の差を示すための効果量については、試験実施計画書で事前に定めるべきである。治療効果がないとの仮説を棄却するために偏りがなく、十分な検出力を備えた検定が可能となるように、試験をデザインすべきである^{24,27,28)}。</p> <p>同等性または非劣性 同等性または非劣性評価のためのPROの目的は、事前に指定したPROドメインについて、治療群が対照群と類似している(同等性)または対照群よりも事前に指定した臨床的に意味のあるマージンの分だけ劣らない(非劣性)ことを示すことを意図する。これらのマージンは試験実施計画書において事前に指定することが重要である。治療効果が同等ではないまたは劣るとの仮説を棄却するために偏りがなく、十分な検出力を備えた検定が可能となるように、試験をデザインすべきである²⁷⁾。</p> <p>効果量(優越性)およびマージン(同等性または非劣性)の選択はPRO評価尺度および臨床的文脈に適合したものであり、臨床的および統計学的な根拠を示すべきである²⁷⁾。試験にはこれらの治療群間のPRO目的の組合せを含むことができる。ただし、試験実施計画書には、どのPROドメインまたは項目について優越性、同等性、または非劣性検定を行うかを明示すべきである。</p>
<p>RS 3-5 患者内または治療群内のPROの目的を試験実施計画書に明示する。妥当な患者内または治療群内のPROの目的には以下のものがある：</p> <ul style="list-style-type: none"> ・改善 <ul style="list-style-type: none"> ○改善までの時間 ○時点tでの改善の程度 ○時点tで改善を示したレスポンドの割合 ・増悪 <ul style="list-style-type: none"> ○増悪までの時間 ○時点tでの増悪の程度 ○時点tで増悪を示したレスポンドの割合 ・安定状態 <ul style="list-style-type: none"> ○安定状態(終了)までの時間 ○時点tで安定状態を示したレスポンドの割合 ・全体的効果 <ul style="list-style-type: none"> ○全体的なPROスコアの経時的推移 ○奏効パターンまたはプロファイル 	<p>治療群内のPROの仮定：改善, 増悪, 安定状態, または全体的効果 治療群内で増悪, 安定状態, または改善のいずれかが予想されるかの選択は先行文献, 専門家の知識, または早期のフェーズの試験結果に基づくべきである。治療群内の関心が効果が特定の方向ではなく、全体的効果である可能性もある(特定のPROドメインについての各患者におけるすべての利用可能なスコアの推移の要約)。しかし、全体的効果については、事前の治療群内での仮定がないため、導き出される結論は頑健性が低い可能性があることに注意する必要がある。</p> <p>どの治療群内のPRO仮定を用いるかを決定する際には、特定のPROドメインについての患者の観察されたベースライン値を考慮に入れるべきである。この決定は、当該PROドメインにおける臨床的に意味のある変化の評価の実施可能性への情報提供に役立つ。</p> <p>患者内または治療群内のPRO目的：イベント発生までの時間, 時点tでのイベントの程度, 時点tでのレスポンドの割合, 全体的なPROスコアの経時的推移, もしくは奏効パターンまたはプロファイル さまざまな患者内または治療群内のPRO評価項目が可能である。しかし、これらはしばしば無視され、誤って同義語として解釈される。たとえば、治療中の最初の増悪までの時間を評価するPRO評価項目は6週時点での増悪の程度を評価する評価項目とは等しくない。実際、これらのPRO評価項目では異なる解析方法が用いられ、異なる結論が生じる可能性がある。評価項目に応じて、PROドメインのための臨床的に意味のある閾値は患者レベル(例：患者内：患者をレスポンドまたはレスポンド以外に分類)の場合もあれば、群レベル(例：群内：群内での平均変化)の場合もある²⁹⁾。</p> <p>患者内のPRO目的 主な関心は、さらなる解析を行う前に、どの患者が臨床的に意味のある反応を示したかを特定することである。個人レベルで臨床的に意味のある閾値(レスポンドの定義)を定め、それを用いて、どの患者が臨床的に意味のある変化を示したか、または示さなかったかを特定する。この目的はイベント発生までの時間やレスポンドの割合等の評価項目と関連する。</p> <p>治療群内のPRO目的 主な関心は、平均的にみて特定の群が臨床的に意味のある変化を示したかどうかを評価することである。群レベルで臨床的に意味のある閾値を指定し、それを用いて、どの群が臨床的に意味のある変化を示したか、または示さなかったかを特定する。この目的は変化の程度等の評価項目と関連する。</p> <p>これらのPRO目的についての具体的な定義はRS 6～9に示す。</p>

推奨された声明	検討項目
<p>RS 6 改善は、ベースラインからの変化が事前に定めた改善の閾値レベルに達した場合と定義される(ベースライン後の改善)。フォローアップ評価の結果が改善の閾値以上の場合、改善は維持される(決定的改善)。フォローアップ評価の結果が改善の閾値を下回った時点で改善は終了する(一過性の改善【図3】)。</p>	<p>改善までの時間 臨床的に意味のある患者内レベルでの改善を事前に定め、臨床的に意味のある改善が認められるまでの時間を評価する。この事前に定めた改善の閾値を上回るまたは下回るスコアの変動は考慮しない。</p> <p>時点tでの改善の程度 臨床的に意味のある治療群内の改善を事前に定め、事前に定めた臨床的に意味のある時点での平均値または中央値の改善(および対応するCI)を評価する。観察されたスコアの変動を考慮に入れる。</p> <p>時点tで改善を示したレスポンスの割合 臨床的に意味のある患者内レベルでの改善を事前に定め、事前に定めた臨床的に意味のある時点で改善を示した患者数を評価する。この事前に定めた改善の閾値を上回るまたは下回るスコアの変動は考慮しない。</p>
<p>RS 7 増悪は、ベースラインからの変化が事前に定めた増悪の閾値レベルに達した場合と定義される(ベースライン後の増悪)。フォローアップ評価の結果が増悪の閾値以下の場合、この増悪は維持される(決定的増悪)。フォローアップ評価の結果が増悪の閾値を上回った時点で増悪は終了する(図3)。</p>	<p>増悪までの時間 臨床的に意味のある患者内レベルでの増悪を事前に定め、臨床的に意味のある増悪が認められるまでの時間を評価する。この事前に定めた増悪の閾値を上回るまたは下回るスコアの変動は考慮しない。</p> <p>時点tでの増悪の程度 臨床的に意味のある治療群内の増悪を事前に定め、事前に定めた臨床的に意味のある時点での平均値または中央値の改善(および対応するCI)を評価する。観察されたスコアの変動を考慮に入れる。</p> <p>時点tで増悪を示したレスポンスの割合 臨床的に意味のある患者内レベルでの増悪を事前に定め、事前に定めた臨床的に意味のある時点で増悪を示した患者数を評価する。この事前に定めた増悪の閾値を上回るまたは下回るスコアの変動は考慮しない。</p>
<p>RS 8 安定状態は、ベースラインからの変化がない、またはベースラインからの変化が事前に定めたベースラインの-margin内にあることと定義される。フォローアップ評価の結果がベースライン時の事前に定めた-margin内にある場合、この安定状態は維持される。フォローアップ評価の結果が事前に定めたベースライン時の-margin外となり、改善または増悪の閾値に達した時点で安定状態は終了する。</p> <p>状況によっては、関連のあるPRO目的が安定状態の定義に改善を含むことがある(少なくとも安定)。この場合、フォローアップ評価の結果が増悪の閾値に達しない限り、依然として安定状態であると結論付けられると定義される(図3)。</p>	<p>安定状態の現在の定義は3つの可能なカテゴリー(改善、増悪、安定状態)の区別を意味することから、コンソーシアムメンバー間での不一致(議論中)が生じた。しかし、状況によっては、改善と安定状態の間のカテゴリー、または増悪と安定状態の間のカテゴリー(合計5カテゴリー)が存在することがある。これらの追加の2カテゴリーは安定状態と改善または増悪の間の誤差-marginとして用いる、あるいは意味のあるカテゴリー(例:部分的改善または部分的増悪)として含めることができる。</p> <p>安定状態(終了)までの時間 安定状態までの時間については、臨床的に意味のある患者内の安定状態のレベルを事前に定め、臨床的に意味のある安定状態が認められるまでの時間を評価する。安定状態に達する前に増悪が生じることが予想される場合、この評価項目は有用である可能性がある。安定状態(終了)までの時間については、安定状態が終了するまでの時間、もしくは臨床的に意味のある改善または増悪が認められるまでの時間を評価する。</p> <p>時点tで安定状態を示したレスポンスの割合 臨床的に意味のある患者内レベルでの安定状態を事前に定め、事前に定めた臨床的に意味のある時点で安定状態を示した患者数を評価する。この事前に定めた増悪の閾値を上回るまたは下回るスコアの変動は考慮しない。</p> <p>時点tでの安定状態の程度 増悪または改善と異なり、安定状態は時点tでの安定状態の程度を評価するPROの目的設定はおこなわない。安定の基準を満たす2名の患者を比較する場合、片方の患者が他の患者よりも安定しているとみなすような順位付けを行うことはできない。定義上、安定状態の閾値内にある値を区別することはノイズ(意味のある変化を表さないランダムな変動)とみなされる。</p>
<p>RS 9 全体的効果は、特定のPROドメインまたは項目についての各患者のすべての利用可能なスコアの推移の要約と定義される。</p>	<p>治療効果または臨床的ベネフィットのPROの目的とともに、全体的効果の評価項目を用いることができるかどうかについては、コンソーシアムメンバー間で不一致が生じた(議論中)。推奨は、治療効果または臨床的ベネフィットのPROの目的と並行して全体的効果を用いることができる、であった。この種の評価項目では改善、増悪、および安定状態と比べて情報が失われることから(例:全体的なPROスコアの経時的推移は効果の方向および時期をとらえない)、全体的効果の評価項目を使用する際には注意を要する。たとえば、全体的なPROスコアの経時的推移は効果の方向および時期をとらえない。</p> <p>全体的なPROスコアの経時的推移 目標は、ある期間中のすべての利用可能なスコアを特定のPROドメインについての患者ごとの単一のデータポイントに要約することである。関心のあるタイムフレームは事前に定めるべきである。次に、その結果としてのアウトカムを2群の比較に用いることができる。全体的なPROスコアの経時的推移をとらえるために、平均値または中央値、最小値および最大値、曲線下面積など複数の要約尺度が存在する^{30,31)}。これらの要約尺度は、研究目的に応じて、ベースライン時のスコアを含む場合も含まない場合もある。また、これらの値の解釈を助けるために、臨床的に意味のある閾値は事前に定めるべきである。しかし、すべての利用可能なデータを1つのスコアに要約することにより情報が失われ、特定の時点での臨床的に意味のある変化が不明瞭になる可能性がある³¹⁾。したがって、全体的なPROスコアの解釈を裏付けるために、全体的なPROスコアの経時的推移の解析および提示には常に、事前に定めた期間(全体的なPRO尺度に含まれる期間)全体でのPROの経時的推移の提示も含むべきである。推奨される要約尺度は本文書には含まないが、将来の作業の一部である。</p> <p>奏効パターンまたはプロファイル 目標は、奏効の経時的推移を記述することである。これらの値の解釈を助けるために、臨床的に意味のある閾値も事前に定めるべきである。期限内に正確なプロファイルを事前に定めることは常に容易ではないため、この患者内または治療群内のPROの研究目的は、治療効果または臨床的ベネフィットのエビデンスではなく、記述的または探索的目的とともに使用することが推奨される。</p>

推奨された声明	検討項目
セクション2：統計手法の推奨	
<p>RS 10 PROデータの解析に必須の統計学的属性には以下のものがある：</p> <ul style="list-style-type: none"> ・2群間での統計学的検定の実施 ・臨床的に意味のある結果の作成 <p>極めて望ましい統計学的属性には以下のものがある：</p> <ul style="list-style-type: none"> ・共変量の調整(ベースライン時のPROスコアを含む) ・もっとも制約の少ない欠測データの取扱い ・クラスターデータの取扱い(反復評価) 	<p>この声明の開発の経緯についての詳細を、検討した他の統計学的属性の一覧とともに付録(pp 13-15)に示す。</p> <p>2群間での統計学的検定の実施 これらの推奨の現在の適用範囲はランダム化比較対照試験であり、群間での統計学的差の検定はランダム化比較対照試験の主な目標である³²⁾。</p> <p>臨床的に意味のある結果の作成 選択された統計手法は、非統計家にも容易に解釈可能で、有用な臨床的意思決定を導き、臨床実践に影響を及ぼす結果が得られる手法でなくてはならない。結果が統計学的に有意であることは、必ずしもその結果が臨床的に意味があることを示すわけではない³³⁾。したがって、統計手法は、差の統計学的検定に加えて、PRO尺度と直接結び付けることができ、治療効果の大きさ、確実性、および方向についての推定値を与えることができる手法でなくてはならない。この基準は、PROの解析においては、ノンパラメトリックな方法よりもパラメトリックな方法が望ましいことを意味する。パラメトリックな方法は分布についての仮定に依存することから、特に必要症例数が少ない場合にこれらの仮定からの逸脱を検討するために、ノンパラメトリックな方法を用いた感度分析を行うことが推奨される^{34,35)}。</p> <p>ベースライン時の共変量の調整(ベースライン時のPROスコアを含む) ベースライン時の共変量が関心のあるアウトカムと相関する場合、解析の効率を改善し、共変量からの条件付きバイアスを避けるために、そのような共変量の調整を行うことが推奨される^{36,37)}。たとえば、ベースライン時のPROスコアはしばしばフォローアップ時のPROスコアと相関する³⁸⁾。したがって、ベースライン時の共変量を取り込むことができる解析手法を用いることが重要である。その他の共変量には、人口統計学的変数(例：年齢、性別)、疾患特性(例：疾患部位、病期)、他の関連する変数(例：国)等がある。</p> <p>もっとも制約の少ない欠測データの取扱い 欠測の確率が関心のあるアウトカムと関連する場合、検出力の低下のみならず、推定値の潜在的バイアスを生じる可能性がある³⁹⁾。欠測データはがんの臨床試験のPROデータの解析ではほぼ常に生じ、もっとも制約の多い仮定は、欠測データの確率が関心のあるPROドメインまたは項目と関連する可能性はほとんどないとするものである⁴⁰⁾。</p> <p>クラスターデータの取扱い(反復評価) がんの臨床試験では、関心のあるPROドメインまたは項目の変化をとらえるために、PRO評価はしばしば反復して行われる。この種のデータの解析では各患者内でのPRO評価のクラスターリングと測定の時間的順序を考慮に入れる必要がある⁴¹⁾。</p>
<p>RS 11 イベント発生までの時間のアウトカム(改善、安定状態、または増悪)の評価では、ログランク検定ではなく、Cox比例ハザード検定を用いることが推奨される。</p>	<p>合意された一連の基準に基づく統計手法の評価方法の詳細を付録(pp 16-26)に示す。</p> <p>Cox比例ハザード検定を用いる場合は、比例ハザード性の仮定を確認すべきである⁴²⁾。この仮定が満たされない場合は、ログランクまたはCox非比例ハザードモデルを用いた感度分析を行い、調査結果の頑健性を評価することが推奨される。また、time-to-event(イベント発生までの時間)解析の一般的仮定(特に、打ち切りがイベント時間に依存しないこと)が成立しなくてはならない⁴³⁾。</p>
<p>RS 12 時点tでのイベントの程度(改善または増悪)の評価(試験デザインがベースラインと2回以上のフォローアップの場合)では、評価した他の統計手法ではなく、線形混合モデル(離散時間)を用いることが推奨される。</p>	<p>合意された一連の基準に基づく統計手法の評価方法の詳細を付録(pp 16-26)に示す。</p> <p>線形混合モデル(連続時間)、パターン混合モデル(pattern mixture model)、および縦断的なジョイントモデルは一連の基準を満たすものの、事前に必要な仮定(例：時間とアウトカム変数の間の関係に関する仮定)が少ないことから、線形混合モデル(離散時間)が推奨された。</p> <p>解析方針としては、線形混合モデルをデータに当てはめた後、特定の時点tでのテスト推定値を得る。試験のフォローアップ評価回数が少ない場合、この方法は適切である。線形混合モデルの一般的仮定が成立する。たとえば、ランダムな欠測(missing at random)の仮定が満たされなくてはならない。すなわち、線形混合モデルは、欠測データが既知および観察された要因に依存する場合、観察されたであろう治療効果の不偏推定値を与える⁴⁴⁾。</p>
<p>RS 13 時点tでのイベントの程度(改善または増悪)の評価(デザインがベースラインと1回のフォローアップのみの場合)では、ANOVA、ANCOVA、t検定、およびWilcoxon順位検定ではなく、線形回帰を用いることが推奨される。</p>	<p>合意された一連の基準に基づく統計手法の評価方法の詳細を付録(pp 16-26)に示す。</p> <p>多くの統計プログラム(例：SAS)⁴⁵⁾が線形回帰には完全例解析を用いていることから、この推奨される解析には注意を必要とする。そのような解析から得られた推定値は、欠測データが完全にランダムな欠測(missing completely at random)の場合にのみ妥当な推測を与える。</p>
<p>RS 14 SISAQOLの推奨では要約尺度を考慮すべきである。</p>	<p>当初の声明での目標は全体的なPROスコアの経時的推移の評価方法を推奨することであった。この文脈で、要約尺度とは、特定の期間における患者あたりのPROドメインの反復評価を単一のアウトカム(例：曲線下面積、全体平均、最小値および最大値)に併合することと定義される。提案された推奨としては、要約尺度を用いる場合は、アウトカムの群間比較を行うための線形回帰が推奨される。</p> <p>PROの解析でよく用いられるものの、概して2段階手順のための推奨としてみられる可能性があることから、この提案を推奨することはおおむね躊躇された⁴⁶⁾。さらに、データをプールの1つの値に要約した場合、情報が失われ、それがPRO調査結果の解釈可能性に影響を及ぼす可能性がある。</p> <p>文脈によっては、要約尺度はPROデータを理解する上で有用となる可能性があり、SISAQOLの推奨において考慮すべきであることが合意された。ただし、今後の作業では、どの要約尺度を推奨するかの評価、およびこれらのデータに対してもっとも適切な解析方法の特定を行うべきである。</p>

推奨された声明	検討項目
RS 15 奏効の経時的推移を記述するためには、反復測定ANOVA(時間×群の交互作用)よりも、線形混合モデル(オムニバス検定、離散時間変数、時間×群の交互作用)が推奨される。	<p>合意された一連の基準に基づく統計手法の評価方法の詳細を付録(pp 16-26)に示す。</p> <p>この方法の焦点は、時間×群の交互作用のp値を解釈することではなく、モデルを当てはめた後に、その結果として得られたパラメータを解釈することである。しかし、これらのプロファイルの事後的な記述は横断的には報告されるが、縦断的には報告されない。すなわち、あらゆる評価時点は平均値とCIをもつ。したがって、標本の(平均的な)縦断的プロファイルではなく、各時点の平均的なアウトカムについて解釈が行われる。</p> <p>個々の縦断的プロファイルに関心がある場合、より複雑なモデルが利用可能である。たとえば、時間を連続量として扱い、一次、二次、および三次の多項式の項を用いて時間曲線を近似することができる。しかし、これらのモデルの多くは特定の仮定に依存し、その結果、推定値またはグラフの解釈が困難となる可能性がある。どの時間曲線がもっとも適切な判断は容易ではなく、理想的には過去のデータからの情報を活用すべきである。</p>
セクション3：欠測データに関連する統計用語の標準化	
RS 16 欠測データとは、特定の研究目的またはestimandの解析に対して意味があると考えられるが、収集されなかったデータである。	<p>文献では欠測データの報告および取扱いの重要性について多くの注意が払われているが¹³⁾、何を欠測データとみなすかは不明確である。欠測データとは、欠測の理由にかかわらず測定されなかったPRO評価^{40,47)}、利用可能であると予想されたPRO評価の未完了²¹⁾、またはもし観察された場合は解析に対して意味があったと考えられる欠測値^{48,49)}を指す。</p> <p>医薬品規制調和国際会議(ICH)E9ガイドラインの定義を採用することは、解析に対して意味があると考えられるデータのみがPRO調査結果に寄与することを意味する。この枠組み内での欠測PROデータは、必要症例数の減少(情報のない欠測[non-informative missing]データ)、治療効果の推定値の歪曲(情報のある欠測[informative missing]データ)または両者により、PRO調査結果の解釈可能性に影響を及ぼす可能性がある。</p>
RS 17 解析に対して意味があるとは、特定の研究目的またはestimandに基づくPRO解析対象集団を指す。	<p>PRO試験対象集団とPROの解析対象集団を区別する必要がある。PRO試験対象集団とは、同意が得られ、PROデータ収集への参加に適格であったすべての患者と定義される。理想的には、PRO試験対象集団はITT集団と同一であるが、この条件は常に必要または実施可能とは限らない。ITT集団から逸脱した理由、および特定の部分集団からPROを全く収集しない理由の根拠を試験実施計画書に記載すべきである。PRO試験対象集団はITT集団の部分集団であり、同意または適格性、もしくは両者の理由でPROアウトカムが全く収集できなかった患者を除外したものである。PRO試験対象集団の患者は、患者のフォローアップ状態または観察にかかわらず、試験開始時に特定可能とすべきである。したがって、PRO試験対象集団はintention-to-collect(収集を意図した)PRO集団である。</p> <p>PROの解析対象集団は、主要PROの解析に含まれる患者を指し、PRO試験対象集団と可能な限り近づけるべきである。PROは同一患者において経時的に繰り返し評価されるため、一部の予定された評価が観察されない場合は注意を要する⁴⁸⁾。解析手法に応じて、計画した評価の対象から一部の患者を除外することは、これらの患者を意図したPROの解析対象集団からまとめて除外することを意味する可能性がある。PROの解析対象集団は定義されたPROの解析との関係でのみ存在する。いくつかの主要PROの解析が計画されている場合、各解析は各々のPROの解析対象集団に対応し、それらの集団は互いに異なる場合もあれば、異なる場合もある。</p>
RS 18 死亡した患者はPRO試験対象集団の一部であったが、これらの患者からのPRO評価はそれ以上期待できない。	<p>死亡後のPRO評価からは意味のある値が得られないため、これらの観察は期待すべきではない^{21,49)}。死亡後の評価はPRO推定値に寄与しないため、解析に対して意味はなく、したがって、欠測とはみなされない。</p>
RS 19-20 可変分母割合を報告すべきである。この割合は、指定された時点でPRO評価中の患者のうちで妥当なPRO評価を提出した患者の割合と定義される。完了割合の用語は、可変分母割合を用いた割合を表すために用いるべきである。	<p>PRO評価中の患者数は、その時点でなおPRO評価の提供が期待できる患者を指す。一方、PRO評価中ではない患者とは、その時点以降はもはやPRO評価の提供が期待できない患者と定義される。</p> <p>死亡後のPRO評価はPRO評価中ではないとみなし、完了割合の分母(PRO評価中の患者数)には含まないように標準化することが合意された。これは、死亡後の観察されなかった評価は欠測データとはみなさないことを示唆する。</p> <p>その他の理由(PRO評価中の患者数におけるPRO試験実施計画書外、患者の同意撤回、追跡不能等)を標準化すべきかどうかについては、さらなる議論を要する(付録pp 35-36)。</p>
RS 21-22 固定分母割合を報告すべきである。この割合は、PRO試験対象集団(同意が得られ、PROデータ収集への参加に適格であったすべての患者)のうちで、指定された時点で妥当なPRO評価を提出したPRO評価中の患者の割合として定義される。 利用可能なデータの割合の用語は、固定分母割合を用いた割合を表すために用いるべきである。	<p>利用可能なデータの割合(固定分母割合)は、生存者バイアス(可変分母割合には反映されない)およびPRO推定値のための観察データに寄与する患者数についての質問に対応するために必要であった。</p>
RS 23 百分率に加えて、(両方の割合について)あらゆる時点での分子および分母の絶対数を報告すべきである。	<p>CONSORTのチャートは欠測データとなった理由を報告する際に有用である可能性が提案された。理由についての3つの広範なカテゴリー(死亡、試験実施計画書で事前に指定した理由、試験実施計画書で事前に指定しなかった理由)を設けることが提案された。このアイデアを発展させるためのさらなる作業が必要である。</p>

推奨された声明	検討項目
セクション4：欠測データの一般的取扱い	
RS 24 臨床試験を行う際には、PROが欠測となった理由を探索することが重要である。	シミュレーション研究の結果、PRO調査結果に対する欠測データの割合の影響は欠測データとなった理由(例：informative, non-informative, または両者の組合せ)に依存することが示された。したがって、欠測データとなった理由を収集することは、PRO調査結果の頑健性に対する欠測データの割合の影響を評価する上で重要である。
RS 25 欠測データは、臨床試験およびPROデザインの戦略、ならびにトレーニングまたはモニタリングのアプローチにより、前向きに最小化すべきである。	全患者の完全な評価から導き出された頑健な治療比較の可能性を回復する解析手法はない ⁴⁸⁾ 。したがって、慎重なデザインおよび計画を通してPRO評価の欠測を防ぐことを、欠測PROデータの取扱いにおける1次戦略とすべきである ⁴⁹⁾ 。詳細情報はMercieca-Bebber et al. ⁵⁰⁾ を参照のこと。
RS 26 欠測PROデータの取扱いに必要なデータの収集について統計解析計画書に記載することが推奨される(欠測データとなった理由、および解釈または補完のため補助データ)。	慎重な計画および収集方針を採用したとしても、欠測データの発生は避けられない。欠測データについては、解析中に検証不能な仮定を置く必要がある ⁵¹⁾ 。欠測データとなった理由や補助データの収集は、主要解析および感度分析におけるこれらの患者の取扱い方法を正当化する上で役立つ可能性がある ^{18,51)} 。
RS 27 主要な統計解析方法：項目レベルおよび評価尺度レベルでの欠測データの取扱い方法を試験実施計画書または統計解析計画書で事前に定めること。	統計解析方法の選択と同様に、欠測データの取扱い方法が異なると、結果が異なる可能性がある ⁵²⁾ 。したがって、主要解析で用いる欠測データの取扱い方法を事前に文書化することが重要である ⁸⁾ 。
RS 28 主要な統計解析方法：評価尺度内の項目レベルの欠測データは、評価尺度の開発中に開発されたスコア化アルゴリズム(利用可能な場合)に従って取り扱うべきである。	欠測項目の取扱い方法についての一般的な推奨はあるものの ⁵³⁾ 、欠測項目の取扱い方法を標準化するために、PRO尺度はスコア化アルゴリズムとともに開発される。主要解析ではこれらのスコア化アルゴリズムを用いるべきであり、欠測項目を取り扱うその他の方法は感度分析の一部として含めることができる。PROのための正式なスコア化アルゴリズムの変更が生じた場合、それを受けて開発者が更新したガイドラインに従うべきである。
RS 29 主要な統計解析方法：群および時点ごとに欠測データの理由および割合についての批判的な評価を行うべきである。	欠測データには多くの可能性のある理由が存在する(例：患者の同意撤回、患者の転居)。理由および欠測データの量に応じて、欠測データの取扱い方法は異なることがある ^{18,51)} 。
RS 30 主要な統計解析方法：統計手法作業部会が定めた方法を用いて、すべての利用可能なデータを使用すること。	欠測データを無視する必要がある方法や完全データを有する患者のみを用いて解析を行う方法(例：完全例解析)は推奨されない ⁵¹⁾ 。欠測データについて、完全例解析よりも弱い仮定を置くことから、すべての利用可能なデータの利用が可能なが推奨される ⁵⁴⁾ 。
RS 31 主要な統計解析方法：臨床試験の文脈内で正当化されない限り、明示的な補完は推奨されない。	明示的な単純な補完方法(last observation carried forward法等)は、患者特性の違いにかかわらず、定数を用いて欠測値を補完するため、推定値のばらつきが過小推定される ⁵⁴⁾ 。定数を補完すると、ばらつきが小さくなるため、 <i>p</i> 値が小さくなる ⁵⁵⁾ 。
RS 32 感度分析は試験実施計画書または統計解析計画書において事前に定めるべきである。さまざまな仮定の下での欠測データの影響を評価するためには、2種類以上の欠測データの取扱い方法を用いることが推奨される。	欠測データの取扱いでは、欠測値とアウトカムとの関係について、検証不能な仮定を置くことを必要とする。結論の頑健性を検証するために、欠測データについての異なる一連の仮定を用いた感度分析が必要である ⁵⁶⁾ 。主要解析と一致する結果は、試験の結論に対して欠測データが重要な影響を及ぼさなかったことをある程度保証する。しかし、感度分析の結果が矛盾する場合、試験の結論に対する欠測データの補完を検討すべきである ⁵¹⁾ 。 追加の感度分析の事前指定、解析、および報告を行うための試験研究者の作業量が増大することから、不一致が生じた。

図2 SISAQOLが推奨する声明およびそれらの検討項目

ITT=intention to treat, PRO=患者報告アウトカム, RS=推奨された声明, SISAQOL=患者報告アウトカムおよびQOL評価項目データの解析における国際標準策定(Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data)

な統計手法を定める一連の「必須」および「極めて望ましい」統計学的基準が推奨された。統計手法が必須の基準を満たさない場合、その手法はPROの解析のための適切な手法としては推奨しなかった。

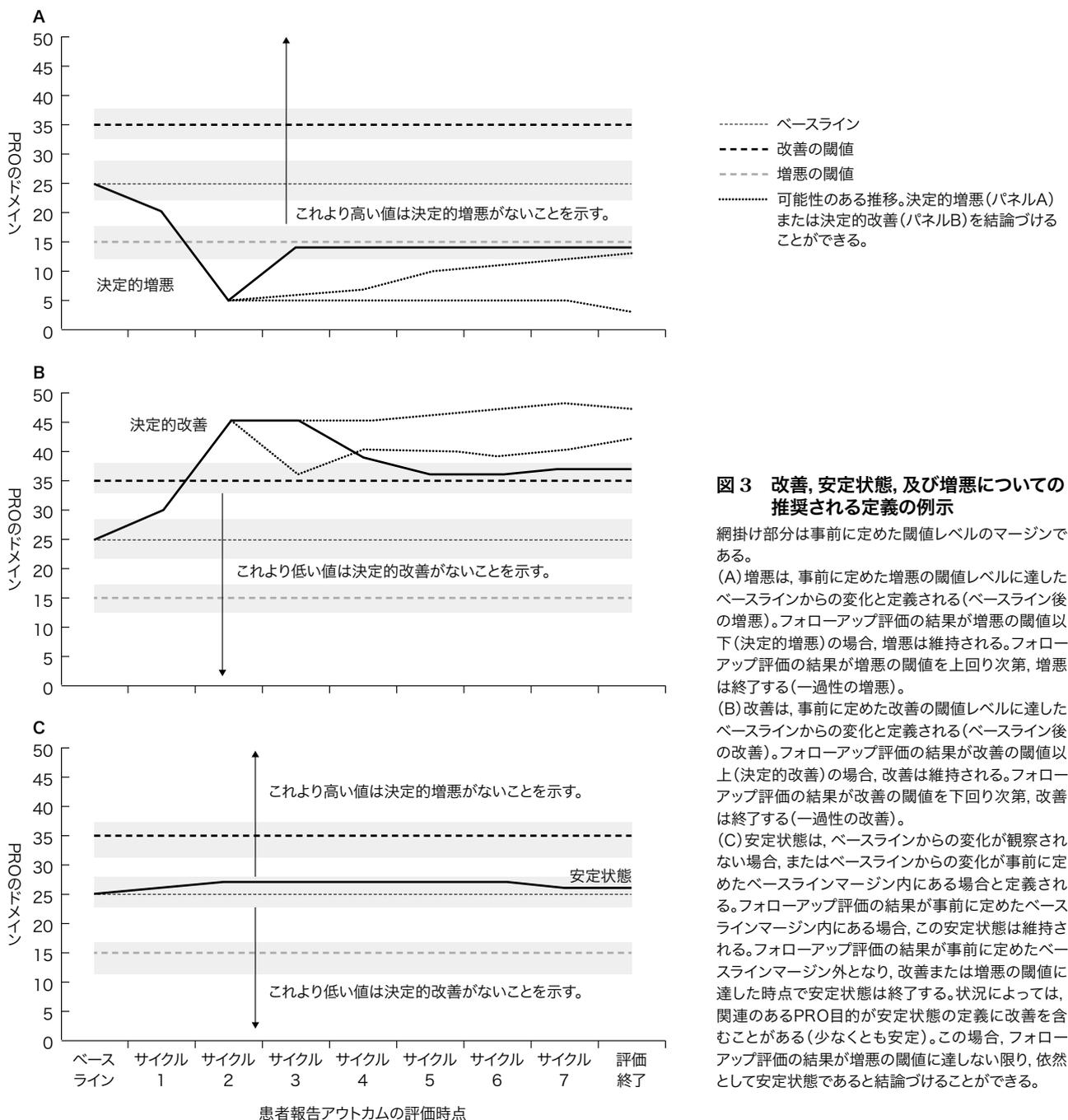
以下の2つの必須の統計学的特性が特定された：比較検定を行う能力(統計学的有意性)および解釈可能な治療効果の推定値を生じる能力(臨床的意義)。極めて望ましい基準は、共変量(ベースライ

ン時のPROスコア等)による調整能力、最小限の制約での欠測データの取扱い、およびクラスターデータの取扱い(反復評価)であった。これらの基準についての詳細情報は図2の推奨された声明10に示す。2つ以上の統計手法が「必須」および「極めて望ましい」の基準に等しく当てはまった場合は、より単純な手法を優先した。特定の目的に対しては、より複雑なモデル(例：パターン混合モデル [pattern

表 推奨される主要な統計手法に合致した研究目的の分類法の概要

	治療効果または臨床的ベネフィット(検証的目的)		記述的な患者の観点(探索的または記述的目的)
	優越性(治療群間の目的)	同等性または非劣性(治療群間の目的)	
改善			
改善までの時間	Cox比例ハザード, および治療群間差についての事前に定めた効果量	Cox 比例ハザード, および治療群間差についての事前に定めた同等性マージンまたは事前に定めた非劣性マージン	探索的:Cox比例ハザード, 記述的:改善までの時間の中央値, 特定の時点での改善の確率, またはHRおよびCI
時点tでの改善を示した患者の割合	ロジスティック混合モデル, (Cochrane) Mantel-Haenszel検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要	ロジスティック混合モデル, (Cochrane) Mantel-Haenszel 検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要	探索的:ロジスティック混合モデル, (Cochrane) Mantel-Haenszel検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要, 記述的:時点tでのレスポンスの割合, もしくはORまたはRRおよびCI
時点tでの改善の程度	線形混合モデル(離散時間), および治療群間差についての事前に定めた効果量	線形混合モデル(離散時間), および治療群間差についての事前に定めた同等性マージンまたは事前に定めた非劣性マージン	探索的:線形混合モデル(離散時間), 記述的:ベースライン時および時点tでの平均的な大きさおよびCI, または時点tでの改善の程度の平均およびCI
安定状態			
安定状態(終了)までの時間	Cox比例ハザード, および治療群間差についての事前に定めた効果量	Cox 比例ハザード, および治療群間差についての事前に定めた同等性マージンまたは事前に定めた非劣性マージン	探索的:Cox比例ハザード, 記述的:安定状態(終了)までの時間の中央値, 特定の時点での安定状態(終了)の確率, またはHRおよびCI
時点tでの安定状態を示した患者の割合	ロジスティック混合モデル, (Cochrane) Mantel-Haenszel検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要	ロジスティック混合モデル, (Cochrane) Mantel-Haenszel 検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要	探索的:ロジスティック混合モデル, (Cochrane) Mantel-Haenszel検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要, 記述的:時点tでのレスポンスの割合, もしくはORまたはRRおよびCI
時点tでの安定状態の程度	NA*	NA*	NA*
増悪			
増悪までの時間	Cox比例ハザード, および治療群間差についての事前に定めた効果量	Cox 比例ハザード, および治療群間差についての事前に定めた同等性マージンまたは事前に定めた非劣性マージン	探索的:Cox比例ハザード, 記述的:増悪までの時間の中央値, 特定の時点での増悪の確率, またはHRおよびCI
時点tでの増悪を示した患者の割合	ロジスティック混合モデル, (Cochrane) Mantel-Haenszel検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要	ロジスティック混合モデル, (Cochrane) Mantel-Haenszel 検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要	探索的:ロジスティック混合モデル, (Cochrane) Mantel-Haenszel検定, または単純ロジスティックモデルを推奨するかについてはさらなる議論が必要, 記述的:時点tでのレスポンスの割合, もしくはORまたはRRおよびCI
時点tでの増悪の程度	線形混合モデル(離散時間), および治療群間差についての事前に定めた効果量	線形混合モデル(離散時間), および治療群間差についての事前に定めた同等性マージンまたは事前に定めた非劣性マージン	探索的:線形混合モデル(離散時間), 記述的:ベースライン時および時点tでの平均的な大きさおよびCI, または時点tでの増悪の程度の平均およびCI
全体的効果			
全体的なPROスコアの推移	さらなる議論が必要	さらなる議論が必要	さらなる議論が必要
奏効パターンまたはプロファイル	NA [†]	NA [†]	探索的:線形混合モデル(離散時間または連続時間), 記述的:ベースライン時およびタイムフレーム内のすべての時点での平均的な大きさおよびCI, またはタイムフレーム内のすべての時点での平均変化およびCI, または平均プロファイルの推移およびCI

治療群間の優越性の目的のために推奨される統計手法を最初に概念化した。しかし、これらの手法は非劣性または同等性の目的に外挿できる可能性があり(ただし、適切なマージンを事前に定めること)、探索的な目的に外挿できる可能性がある(ただし、そのような調査結果を臨床的ベネフィットまたは治療効果のエビデンスの基礎として用いてはならない[図2])。記述統計量は、統計手法作業部会による研究目的を備えた適切な統計手法の評価に関する作業に基づく(付録 pp 19-27)。HR=ハザード比, NA=該当せず, OR=オッズ比, PRO=患者報告アウトカム, RR=リスク比。*安定の基準を満たす2名の患者を比較する場合、片方の患者が他の患者よりも安定しているとみなすような順位付けを行うことはできない。定義上、安定状態の閾値内にある値を区別することはノイズ(意味のある変化を表さないランダムな変動)とみなされる。†期限内に正確なプロファイルを事前に定めることは常に容易とは限らないため、奏効パターンまたはプロファイルは、治療効果または臨床的ベネフィットのエビデンスではなく、記述的または探索的目的とともに使用することが推奨される。



mixture model]) を推奨する利点があるものの、この利点はしばしば強力で検証不能な仮定を必要とし、統計家以外の者には容易に解釈できない結果を生じ得る。主要PROの解析に対しては、実施可能性、有用性、解釈可能性、および統計学的正確性のバランスが不可欠であると判断された。しかし、感度分析では、主要結果の頑健性を検証するためのより複雑なモデルを使うことができる。

合意された一連の統計学的基準および選択基準に基づき、各PRO目的のための統計手法が推奨され、以下の2種類の統計手法が推奨された：time-to-event (イベント発生までの時間) なPRO目的の場合はCox 比例ハザード、時点 t でのイベントの程度、および奏効パターンまたはプロファイルの場合は線形混合モデル (推奨された声明 11, 12 および 15 [図 2])。例外的に、PRO デザインがベースライン時と

1回のフォローアップ評価のみを要する場合は、適切な統計手法として線形回帰が推奨された（推奨された声明13 [図2]）。

特に、PROの解釈に対する臨床的意義が必須の基準であることが合意されたため、ノンパラメトリックな方法よりもパラメトリックな方法が推奨された。しかし、パラメトリックな方法には限界があり、もっとも重要なことに、この方法は分布の仮定に依存する³⁵⁾。この限界に対処するには、感度分析ではノンパラメトリックな方法を用い、これらの仮定からの逸脱を調べることが推奨される³⁵⁾。

レスポンドの割合についての縦断的データを評価するための適切な統計手法については合意に至らず、さらなる議論を要する。また、PROデータの推移についての推奨される要約尺度（例：最小値および最大値、曲線下面積、全体平均）についても合意に達しなかったが、要約尺度はSISAQOLの将来の作業の一部であることが認識された（推奨された声明14 [図2]）。順序データを連続データとして解析することの適切性についてはさらなる検討を要する。この問題についての議論は統計学的近似、モデルの複雑度、および解釈の容易さについての問題を中心に展開した。

欠測データに関連する標準化された用語

このセクションで提案された声明11件中8件（73%）が採択された。欠測PROデータの定義についての推奨が提案された：欠測PROデータの定義は、特定の研究目的の解析に対して意味があると考えられるが、収集されなかったデータとする（推奨された声明16および17 [図2]）^{48,49)}。この定義は、科学的質問によっては、観察されなかった評価がすべて欠測データとみなされる訳ではないことを意味する（例：死亡後の観察されなかった評価、PRO目的が治療中の患者に重点を置く場合に治療中止後の観察されなかった評価、PROの目的を達成した後の観察されなかった評価）。しかし、解析手法に応じて、観察されなかった評価はすべて黙示的に欠測データと同様に扱われることがある⁵⁷⁾。推奨される各方法に特有の欠測データの取扱い方法についての推奨はSISAQOLコンソーシアムの作業の次のステップである。

このポリシーレビューでは、予想されるデータの参照セットとの関連で欠測データを区別することの重要性を強調している（推奨された声明19～22 [図2]）。議論の結果、以下の2つの定義が生じた：いわゆる利用可能なデータの割合の分母は固定であり、PRO試験対象集団（同意が得られ、ベースライン時にPROデータ収集への参加に適格であったすべての患者）における患者数で定義される。他方、完了割合の分母は可変であり、指定された時点でのPRO評価可能例（その時点でPRO評価が得られると予想されるすべての患者）における患者数で定義される。両割合の分子は、指定された時点で妥当なPRO評価を提出した患者数である。完了割合の分母は選択した研究上の問いに依存することに注意する（例：進行時までのPROのみを収集すべきか、あるいは進行後のPROも収集すべきか）。死亡した患者は、死亡後の評価時点での完了割合の分母から除外することが推奨された。しかし、これらの患者は利用可能なデータの割合（常にベースライン時の固定された患者集団を指す）の分母には含まれる（推奨された声明18 [図2]）。

欠測データ

このセクションで提案された声明の半数強（9/16 [56%]）が採択された。シミュレーション研究を行い、重大な欠測データを定義する閾値を設けることが可能かどうかを評価した²²⁾。閾値に関する合意には至らなかったが、シミュレーション研究の結果、PRO調査結果に対する欠測データの割合の影響は欠測データの種類（informative missing または non-informative missing データ）に依存することが示された。欠測データとなった理由を収集することがPRO調査結果の欠測データの影響の評価において重要であることが推奨された（推奨された声明24 [図2]）²⁰⁾。欠測データとなった理由を標準化された方法で収集するための症例報告書が必要であり、コンソーシアムによりさらなる開発が行われる予定である。既存の規制ガイドラインに一致した、欠測データの取扱い方法に関する一般的な推奨が提案された（推奨された声明25～30 [図2]）。

考 察

SISAQOLの目的は、がんのRCTにおけるPROの解析についての標準的なアプローチを促進するための一連の推奨を開発することである。批判的な文献レビュー、ならびに国際的な専門家およびステークホルダーとの協議を通して、SISAQOLは適切な統計手法に合致し、十分に定義されたPROの研究目的のフレームワークを提供する。time-to-event (イベント発生までの時間)アウトカムの適切な解析手法として、Cox 比例ハザードモデルが推奨された。時点 t でのイベントの程度、および奏効パターンまたはプロファイルの解析には、線形混合モデルが推奨された。欠測PROデータ、完了割合、および利用可能なデータの割合の標準化された定義についての推奨が、その標準化された計算および報告方法とともに提案された。欠測PROデータを取り扱うための一般的推奨もいくつか提案された。

がんの臨床試験から頑健なPROの結論を導き出すことは、標準化された研究目的および解析標準について合意に至って用いることだけではない。それは、プロセス開始時からの患者代表者の意味のある関与、質の高いデータ収集、および透明性の高い結果報告を備えた、熟考された試験計画およびデザインを必然的に伴う。この一連の推奨は臨床研究者、臨床試験医、および統計家がPRO試験の概念化およびデザイン、統計解析の品質、がんの臨床試験におけるPROの臨床的解釈を改善することに役立つと考えられる。SISAQOLは、がんの試験におけるPROについてのベストプラクティスに関する方法的推奨のツールボックスを加えており、SPIRIT-PRO⁸⁾、CONSORT-PRO⁹⁾、その他の関連ガイドライン^{56,58)}等も包含している。SPIRIT-PROおよびCONSORT-PROの推奨は、読者がPRO評価項目のデザイン、解析、および解釈の頑健性を判断できるように、試験実施計画書および最終報告書のための優れた質の高い報告に重点を置いているが、SISAQOLの推奨はPROのデザインおよび解析の品質の改善に重点を置いている。質の高い報告と優れた方法論は交換可能ではない。包括的な目標は臨床試験におけるPROの報告および方法論をともに改善することである。

安全かつ有効ながん治療に対する大きなニーズやがんの臨床試験に要する多額のコストおよび複雑さを考えると、規制当局、医療費支払者、臨床医、患者およびその家族による臨床的および医療に関する方針決定が頑健で科学的に正しい国際標準に基づいて行われ、限られた研究資源を無駄にしないことが重要である¹⁰⁾。

本研究の限界および今後の作業

PROの解析についての本標準にはいくつかの限界がある。1番目に、われわれはがんのRCTに焦点を置いた。多くの問題がその他の健康状態に一般化できる可能性があるものの、この一般化はさらなる精査を必要とする。もう1つの限界は、健康関連QOL (HRQL) を選好で重み付けした (preference-weighted) 尺度 (選好に基づく [preference-based] 尺度、または多属性効用 [multi-attribute utility] 尺度とも呼ばれる) に対するこれらの標準の妥当性にある。選好で重み付けした尺度には以下の2つの目的がある：(1つは) 特殊な種類のHRQL要約スコア [HRQLドメインを通常は一般集団の選好、時に患者の選好で重み付けしたもの] で表される効用スコア、(もう一つは) 質調整生存年 (quality-adjusted life-years) や費用効用分析における質による重み付け。本論文で報告した標準がこれらの目的に適用可能かどうかについては、さらなる議論が必要である。これらの標準をがんのRCTのために改良するには、さらなる作業を要する。1番目に、提案された声明のうち一部 (例：時点 t での患者の割合についての統計手法、要約尺度、欠測データに関するいくつかの問題) については合意に至らず、さらなる議論を必要とする。2番目に、研究目的の分類法は将来のがんの臨床試験に適用し、PRO評価をエンドポイントとした試験計画時の目的に適合するかどうかを評価し、必要に応じて、さらなる改訂を行う必要がある。3番目に、各PRO評価のために選択された統計手法は主に、系統的レビューでみられた、PROの解析で通常用いられる統計手法に基づくものであった。コンソーシアムメンバーは他の手法を提案する機会を有し、PROの解析のために適切な統計手法があった可能性はある。それにもかかわらず、

評価した一連の統計手法は長年検証され、科学的に厳密であり、ほとんどの場合について適用可能である。4番目に、欠測データの取扱い方法等の推奨される各方法についての統計学的なベストプラクティスは合意に達する必要がある。5番目に、特定のPRO目的について検討するために適切な要約尺度についての合意も必要である。特定された限界への対処に加えて、今後のステップでは、PROの解析に対して意味のある対象集団および中間事象の特定を含めるべきである。最後に、これらの推奨と最近提唱されているestimandsのフレームワーク²⁴⁾との関係についても検討する必要がある。

結 論

がんのRCTでは、治療のリスク、ベネフィット、安全性、および忍容性についての有益なエビデンスを提供するために、症状、機能、およびその他のHRQL評価項目などのPROデータを評価する機会が増えている。PRO調査結果は患者、医療提供者、医療費支払者、および規制当局の意思決定者に情報を提供する。これらの理由から、PRO調査結果は頑健で、複数試験を通して一貫して導き出され、意味のある結果を生じることが不可欠である。今回のSISAQOL推奨は、がんのRCTにおけるPROの解析のための国際的コンセンサスに基づく標準策定に向けた重要な第1段階である。

検索式および選択基準

本ポリシーレビューのための参考文献はPubMed検索により特定した。検索用語には、“(“patient reported outcome analysis”) OR (“quality of life analysis”)” AND “cancer” AND “clinical trials” (“患者報告アウトカム解析”) OR (“QOL解析”)” AND “がん” AND “臨床試験”を用いた。データの制約は含めなかった。著者自身のファイルの検索によっても論文を特定した。英語の公表論文のみをレビューした。最終の参考文献一覧は、オリジナリティー、および本ポリシーレビューの広範なスコープとの関連性に基づき作成した。

貢献者 本稿は、2016年1月29日にベルギーのブリュッセルで開催されたSISAQOLキックオフミーティングの出席者により概念化された。著者全員が個々の作業部会の作業に貢献した。著者全員が2018年9月24日開催のSISAQOLコンセンサスマーティングにおいて討議を行い、本作業を最終化した。著者全員が本論文をレビューし、改訂に貢献した。著者全員が本稿の最終案を承認した。

利益関係の宣言 MCは本研究以外で、Astellas社、Takeda社、Glaukos社、Merck社、Daiichi Sankyo社、および患者中心アウトカム研究所 (Patient-Centered Outcomes Research Institute) から謝金を受領し、Health Data Research UK、Innovate UK、英国国立衛生研究所 (National Institute for Health Research: NIHR)、Macmillan、およびUCB Pharma社から助成を受領したことを報告する。ACは2019年1月31日までGenentech社の従業員であったことを報告する。KCは本研究以外で、欧州がん研究・治療機構 (European Organization for Research and Treatment of Cancer: EORTC) から助成を受領し、Bristol-Myers Squibb社、Endomag社、Celgene社、およびAmgen社から謝金を受領したことを報告する。IGはBoehringer Ingelheim International社の従業員であり、同社は使途制約のない研究助成金をEORTCに提供した。MKoは本研究以外で、EORTC、Biofrontera社、およびKomitee Forschung Naturmedizinからの助成を受領し、Janssen-Cilag社、Lily社、およびVerband Forschender Arzneimittelherstellerから謝金を受領したことを報告する。KOは本研究以外で、Bristol-Myers Squibb社、Roche社、Novocure社、Lilly社、Pfizer社、MagForce社、Novartis社、Medac社、Photonamic社、Northwest Biotherapeutics社、VBL Therapeutics社、AbbVie社、Elekta社、Apogenix社、およびBayer社から助成を受領したことを報告する。EP-LはGenentech社の従業員である。MPIは本研究以外で、Radius社の諮問委員会の委員であり、Radius社、Synthon社、およびServier社からInstitut Jules Bordetへの助成、AstraZeneca社、Lilly社、Merck Sharp & Dohme社、Novartis社、Pfizer社、およびRoche-Genentech社からInstitut Jules Bordetへの助成および謝金、Odonate社、Camel-IDS社、Crescendo Biologics社、Periphagen社、Huya社、Debiopharm社、PharmaMar社、G1 Therapeutics社、Menarini社、Seattle Genetics社、Immunomedics社、およびOncolytics社から謝金を受領したことを報告する。GVは本研究以外で、Roche社、Eisai社、Novartis社、およびPfizer社から謝金を受領し、英国政府NIHR、Breast Cancer NOW、およびEORTCから助成を受領したことを報告する。ABは本研究の遂行期間中にBoehringer Ingelheim International社およびGenentech社からEORTCへの助成ならびにEORTC研究助成を受領し、本研究以外でMerck社からの助成を受領し、EORTC QOLグループ理事会の委員であることを報告する。その他の著者は全員が開示すべき利益相反はない。

謝辞 欧州がん研究・治療機構 (European Organisation for Research and Treatment of Cancer) は Boehringer Ingelheim 社から本研究を開始するための用途制約のない研究助成金を受領し、Roche グループの一員である Genentech 社から継続的な資金提供を受けた。国際 QOL 研究学会 (International Society for Quality of Life Research) および国際医薬経済・アウトカム研究学会 (International Society for Pharmacoeconomics and Outcomes Research) の支援および最終稿のレビューに深謝する。欠測データに関する調査研究における Gina Mazza の支援に深謝する。本論文は個々の著者の見解を反映したものであり、米国食品医薬品局 (US Food and Drug Administration), 米国国立がん研究所 (US National Cancer Institute), 英国医薬品・医療製品規制庁 (Medicines and Healthcare Products Regulatory Agency), 医療品質・効率性研究機構 (Institute for Quality and Efficiency in Health Care), カナダ保健省 (Health Canada), 英国国民保健サービス (UK National Health Service), 英国国立衛生研究所 (National Institute for Health Research), または英国保健省 (UK Department of Health) の公式見解または方針を代表するものではない。本研究は米国国立衛生研究所 (US National Institutes of Health: NIH) からの資金提供は受けなかった。AWS と SAM は NIH の従業員である。その他の著者のうちで、NIH の完全または部分的な資金提供を受けた者、NIH の従業員、または NIH 助成を受領した者はいなかった。

参考文献

- Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials.gov (2007–13). *Contemp Clin Trials* 2015; 43: 1–9.
- Basch E, Geoghegan C, Coons SJ, et al. Patient-reported outcomes in cancer drug development and US regulatory review: perspectives from industry, the Food and Drug Administration, and the patient. *JAMA Oncol* 2015; 1: 375–79.
- Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. *Lancet Oncol* 2018; 19: e267–74.
- Hamel J-F, Saulnier P, Pe M, et al. A systematic review of the quality of statistical methods employed for analysing quality of life data in cancer randomised controlled trials. *Eur J Cancer* 2017; 83: 166–76.
- Pe M, Dorme L, Coens C, et al. Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. *Lancet Oncol* 2018; 19: e459–69.
- Fiteni F, Anota A, Westeel V, Bonnetain F. Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review. *BMC Cancer* 2016; 16: 122.
- Bottomley A, Pe M, Sloan J, et al. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol* 2016; 17: e510–14.
- Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA* 2018; 319: 483–94.
- Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013; 309: 814–22.
- Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014; 383: 156–65.
- Bottomley A, Pe M, Sloan J, et al. Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. *Clin Trials* 2018; 15: 624–30.
- Kyte D, Duffy H, Fletcher B, et al. Systematic evaluation of the patient-reported outcome (PRO) content of clinical trial protocols. *PLoS One* 2014; 9: e110229.
- Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Qual Life Res* 2016; 25: 2977–83.
- Efficace F, Fayers P, Pusic A, et al. Quality of patient-reported outcome reporting across cancer randomized controlled trials according to the CONSORT patient-reported outcome extension: a pooled analysis of 557 trials. *Cancer* 2015; 121: 3335–42.
- 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: 653–64.
- Mercieca-Bebber R, Friedlander M, Calvert M, et al. A systematic evaluation of compliance and reporting of patient-reported outcome endpoints in ovarian cancer randomised controlled trials: implications for generalisability and clinical practice. *J Patient Rep Outcomes* 2017; 1: 5.
- Kyte D, Retzer A, Ahmed K, et al. Systematic evaluation of patient-reported outcome protocol content and reporting in cancer trials. *J Natl Cancer Inst* 2019. 111: 1170–78.
- Bell M, Fairclough D. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res* 2014; 23: 440–59.
- Fayers P, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. Chichester: John Wiley & Sons, 2013.
- Palmer MJ, Mercieca-Bebber R, King M, Calvert M, Richardson H, Brundage M. A systematic review and development of a classification framework for factors associated with missing patient-reported outcome data. *Clin Trials* 2018; 15: 95–106.
- Machin D, Weeden S. Suggestions for the presentation of quality of life data from clinical trials. *Stat Med* 2002; 17: 711–24.
- Mazza G, Pe M, Dorme L, et al. How much missing data is too much? Monte Carlo simulations to develop SISAQOL guidelines for missing data handling. 25th Annual Conference of the International Society for Quality of Life Research, Dublin, Ireland October 2018. *Qual Life Res* 2018; 27: ab208.1.
- Atherton PJ, Burger KN, Pederson LD, Kaggal S, Sloan JA. Patient-reported outcomes questionnaire compliance in Cancer

- Cooperative Group Trials (Alliance N0992). *Clin Trials* 2016;13: 612–20.
- 24) ICH Expert Working Group. ICH Harmonised Tripartite: guideline statistical principles for clinical trials E9. 1998. https://database.ich.org/sites/default/files/E9_Guideline.pdf (accessed Jan 7, 2020).
- 25) Tukey JW. We need both exploratory and confirmatory. *Am Stat* 1980; 34: 23–25.
- 26) Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *J Am Med Assoc* 2006; 295: 1152–60.
- 27) European Medicines Agency. Guideline on the choice of non-inferiority margin. 2008. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin_en.pdf (accessed Jan 7, 2020).
- 28) European Medicines Agency. Points to consider on switching between superiority and non-inferiority. 2000. https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-switching-between-superiority-non-inferiority_en.pdf (accessed Jan 7, 2020).
- 29) King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011; 11: 171–84.
- 30) Fairclough DL. Summary measures and statistics for comparison of quality of life in a clinical trial of cancer therapy. *Stat Med* 1997;16: 1197–209.
- 31) Curran D, Aaronson N, Standaert B, et al. Summary measures and statistics in the analysis of quality of life data: an example from an EORTC-NCIC-SAKK locally advanced breast cancer study. *Eur J Cancer* 2000; 36: 834–44.
- 32) National Cancer Institute. NCI dictionary of cancer terms. 2013. <https://www.cancer.gov/publications/dictionaries/cancer-terms> (accessed Jan 7, 2020).
- 33) Wasserstein RL, Nicole AL. The ASA statement on p-values: context, process and purpose. *Am Stat* 2016; 70: 129–33.
- 34) Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials. *JAMA* 2013; 309: 814–22.
- 35) Altman DG, Bland JM. Parametric v non-parametric methods for data analysis. *BMJ* 2009; 339: 170.
- 36) European Medicines Agency. Guideline on adjustment for baseline covariates in clinical trials. 2003. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf (accessed Jan 7, 2020).
- 37) Senn SJ. Covariate imbalance and random allocation in clinical trials. *Stat Med* 1989; 8: 467–75.
- 38) Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. *BMJ* 2002;323: 1123–24.
- 39) Fairclough DL, Peterson HF, Cella D, Bonomi P. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Stat Med* 1998; 17: 781–96.
- 40) Troxel AB, Fairclough DL, Curran D, Hahn EA. Statistical analysis of quality of life with missing data in cancer clinical trials. *Stat Med* 1998; 17: 653–66.
- 41) Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*, 2nd edn. Hoboken, NJ: John Wiley & Sons, 2011.
- 42) Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis—an introduction to concepts and methods. *Br J Cancer* 2003; 89: 431–36.
- 43) Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer* 2003; 89: 232–38.
- 44) Cnaan A, Laird NM, Slasor P. Tutorial in biostatistics: using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med* 1997; 16: 2349–80.
- 45) SAS. The REG procedure: missing values. https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_reg_sect026.htm (accessed Jan 7, 2020).
- 46) Bell ML, King MT, Fairclough DL. Bias in area under the curve for longitudinal clinical trials with missing patient reported outcome data. *SAGE Open* 2014; 4: 2.
- 47) Fairclough DL, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? *Stat Med* 2002; 17: 667–77.
- 48) Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012; 367: 1355–60.
- 49) ICH Expert Working Group. International council for harmonisation of technical requirements for pharmaceuticals for human use: estimands and sensitivity analysis in clinical trials E9(R1). 2017. https://database.ich.org/sites/default/files/E9R1_EWG_Draft_Guideline.pdf (accessed Jan 7, 2020).
- 50) Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. *BMJ Open* 2016; 6: e010938.
- 51) European Medicines Agency. Guideline on missing data in confirmatory clinical trials. 2011. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf (accessed Jan 7, 2020).
- 52) Fielding S, Fayers P, Ramsay CR. Analysing randomised controlled trials with missing data: choice of approach affects conclusions. *Contemp Clin Trials* 2012; 33: 461–69.
- 53) Fayers PM, Curran D, Machin D. Incomplete quality of life data in randomized trials: missing forms. *Stat Med* 1998; 17: 679–96.
- 54) Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med* 2013; 86: 343–58.
- 55) Liu G, Gould AL. Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *J Biopharm Stat* 2002; 12: 207–26.
- 56) US Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. <https://www.fda.gov/media/77832/download>(accessed Jan 7, 2020).
- 57) Kurland BF, Johnson LL, Egleston BL, Diehr PH. Longitudinal data with follow-up truncated by death: match the analysis method to research aims. *Stat Sci* 2010; 24: 211–22.
- 58) European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: the use of

patient-reported outcome (PRO) measures in oncology studies.
2016. https://www.ema.europa.eu/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf

(accessed Jan 7, 2020).

© 2020 Elsevier Ltd. All rights reserved.

Coens C, Pe M, Dueck AC, et al. International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. *Lancet Oncol.* 2020 Feb;21(2):e83-e96. doi: 10.1016/S1470-2045(19)30790-9. の日本語訳である。翻訳と掲載にあたり, Copyright Clearance Center (License No: 5379061257980) 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。