

MOS Short-Form 36-Item Health Survey : 新しい患者立脚型健康指標

福原 俊一*

I は じ め に

健康関連クォオリティ・オブ・ライフ (QOL) は、医療の裨益者の視点で捉えた主観的健康度と日常生活機能を定量化した指標と言えるが、近年、種々の臨床疫学的研究やヘルス・サービス研究において、多用されるようになっていく。特に、米国では、健康関連QOL (functional statusと呼ばれることもある) を医療評価研究のアウトカム指標のひとつとして用いることが、ほとんどルーチン化していると言っても良いほどである。このような背景には、根治が困難な慢性疾患中心の疾患分布、患者中心の医療や質の高い医療を求める社会のニーズ、そして医療裨益者の視点に立った新しい種類の医療評価研究の必要性の認識、等いくつかの要因が指摘されている。

MOS Short-Form 36-Item Health Survey (以下SF-36と略) は、いわゆる包括的・プロフィール型尺度に分類される健康関連QOL尺度のひとつである。SF-36は、米国では主観的な健康度・日常生活機能を構成する最も基本的な要素を測定するアウトカム指標としてスタンダードになっている。また、SF-36は医療評価研究だけでなく、医療保険など実際の医療サービスの分野でも利用されている。例えば、米国の高齢者向けの公的な保険制度であるメディケアは、その多くの部分を民間の保険会社 (殆どがマネジドケア方式をとっている) に委託しているが、委託を受ける側に課せられた契約条件のひとつ

として、全ての患者のQOL (SF-36) を測定したデータを報告することが義務づけられている程である。

SF-36の開発は、Medical Outcome Study (MOS)¹⁾と密接な関連がある。MOSは、1986年に米国で開始された大規模な医療評価研究である。米国で近年盛んに実施されるようになった本格的ないわゆるアウトカム研究の走りであったといえる。ボストン、シカゴ、ロサンゼルス等の大都市で診療する523人の臨床医 (異なる医療保険支払い制度のもとで診療) を無作為に抽出し、これらの医師を1986年2月の9日間の間を訪れた全ての成人の患者を対象に調査を行った。MOSでは、多くのアウトカム指標が活用されたが、このMOSの大きな特徴の1つは、主観的な健康度およびこれらの変化に伴う日常生活機能・社会生活機能の制限の程度を定量的に測定するための健康関連QOL指標を調査項目に用いたことであった。MOSはいくつかのパートで構成されているが、5つの主要慢性疾患の診断がついた2,349人を対象とした縦断的な観察研究のパートでは、約100項目からなるMOS-Long Formが用いられた。この結果から得られたデータをもとに、MOS-Long Formを36項目にまで項目数を削減したSF-36が開発されることになった。なおMOSの2万人以上を対象とした断面観察研究のパートでは、SF-36のうちの20項目からなるSF-20が用いられている。

SF-36を用いた調査・研究において最も汎用されているデータ収集方法は自己記入式の質問紙法であるが、視覚障害がある回答者などのための面接バージョンも作成されている。SF-

* 東京大学大学院医学系研究科講師

36は、36の質問項目からなり、回答選択肢は一部を除き5から6のLikert型の形式を取っている。標準的なスコアリング方法は、36項目それぞれの素点を8つの下位尺度にグルーピングして、0から100点のスケールに換算するようになっている。8つの下位尺度の日本語名、原版名、略語、およびそれぞれの下位尺度に含まれる質問項目の内容を表1に示す。

SF-36の歴史的な開発過程、概念モデル、それぞれの下位尺度の持つ意味とこれに含まれる質問項目の出典や内容、スコアリング方法、欧米

における種々の基礎的・臨床的な研究作業の結果などはマニュアルや文献を参照のこと²⁾。

II 日本語版の開発と標準化

SF-36を世界各国で翻訳し、文化的、計量心理学的にその国々での妥当性を検討する国際プロジェクトであるIQOLA (International Quality of Life Assessment) が1991年に5カ国により開始された。その後、日本からは著者が、7カ国目としてこれに参加し、現在約25カ国以上にその対象が拡大している。欧米文化

と大きく異なるアジアの文化にあって、欧米文化のもとで開発されたQOL尺度の内容が、日本人に理解しうるものか、日本人の健康観に適合するものであるか、計量心理学的な信頼性と妥当性の検討、国民標準値の測定、臨床的妥当性の検討、などを目的として作業を行った。

(1) SF-36日本語版の翻訳および文化的な側面を配慮した表現の修正 (cross-cultural adaptation)

国際プロジェクトIQOLA (International Quality of Life Assessment) では翻訳過程の標準化のために、ガイドライン³⁾が作成され、われわれも、これにもとづいてSF-36日本語版の作成を行った。多段階にわたる翻訳、逆翻訳、Linear Analogue Scale Analysis (LASA) 調査⁴⁾による個々の回答選択肢間の等距離性の定量的な検討、Bilingual・Biculturalの言語専門家や社会心理関係者などによる翻訳のquality check、また健常者および患者を対象としたフォーカス・グループ面接やパイロットテストによる翻訳された表

表1 SF-36 8つの下位尺度と各質問項目

下位尺度日本語名(項目数) (原版名: 略号)	質問項目の内容
身体機能(10) (physical functioning: PF)	・激しい活動をする ・適度の活動をする ・少し重いものを持ち上げる、運ぶ ・階段を数階上までのぼる ・階段を一階上までのぼる ・ひざまずく、かがむ ・1キロメートル以上歩く ・数百メートルくらい歩く ・百メートルくらい歩く ・自分で入浴・着替えをする
心の健康(5) (mental health: MH)	・かなり神経質であった ・どうにもならないくらい、気分がおちこんでいた ・落ちついていて穏やかな気分だった ・落ち込んで、ゆううつな気分だった ・楽しい気分だった
日常役割機能(身体)(4) (role-physical: RP)	・仕事・普段の活動時間をへらした ・仕事・普段の活動ができなかった ・仕事・普段の活動の内容によっては、できないものがあつた ・仕事や普段の活動をすることが難しかった
日常役割機能(精神)(3) (role-emotional: RE)	・仕事・普段の活動時間をへらした ・仕事・普段の活動時間が思ったほどできなかった ・仕事・普段の活動時間が集中してできなかった
体の痛み(2) (bodily pain: BP)	・体の痛みの程度 ・痛みによっていつもの仕事ができなくなった
全体的健康観(5) (general health perception: GH)	・現在の健康状態の評価 ・病気になりやすい ・人並みに健康である ・私の健康は悪くなるような気がする ・私の健康状態は非常に良い
活力(4) (vitality: VT)	・元気一杯だった ・活力にあふれていた ・疲れ果てていた ・疲れを感じた
社会生活機能(2) (social functioning: SF)	・家族・友人などとのつきあいが身体的あるいは心理的な理由でさまたげられた ・人とのつきあいをする時間が身体的あるいは心理的な理由でさまたげられた

現の理解度および内容に対する受容度などに関する最終的な検討,を行った。

われわれの作業では,2度にわたる日本語版の改変が行われ,最終的にSF-36日本語版のVersion1.2が完成した。これら一連の過程は,文献⁵⁾を参照のこと。

(2) SF-36日本語版の計量心理学的な検討

以上のようにして作成されたSF-36日本語版を,いくつかの異なる対象群(健康人,種々の慢性疾患患者,など)に配布し,その回答結果を用いて,以下の計量心理学的検討を行った⁶⁾⁸⁾。信頼性に関しては,テスト—再テスト法による尺度の安定性,再現性を検討した。内的整合性信頼性に関してはCronbach's α を測定し,尺度の信頼性の検証の一手段として用いた。集束性および弁別性妥当性は,8つの下位尺度仮説の妥当性を検証するものであり,これには,項目—下位尺度間の相関をNew MAP (multiple trait analysis program) ⁷⁾を用いて行った。

その結果,信頼性に関しては,テスト—再テスト法の結果,信頼性係数は0.78~0.86で,全てのサブ・スケールにおいて十分な安定性を示した。内的整合性信頼性では,Cronbach α 値は0.71~0.91で全てのサブ・スケールにおいて十分な値を示した。集束性・弁別性妥当性に関しては,8つのサブ・スケール仮説の成功率は,各サブ・スケールで集束性・弁別性ともに90~100%で良好であった。構成概念的妥当性に関しては,主成分分析の結果,諸外国と同様の2因子の結果が導き出された⁹⁾。なお選択された2因子の累積分散量は,total varianceの66%,reliable varianceの81.9%であった。しかし,下位尺度の1つである「精神状態の変化による日常役割機能制限」(RE: Role Emotional)の因子負荷量が身体的な因子と精神的な因子の両方に同等に負荷していたことは,欧米諸国と異なる結果であった。(ただし,若い女性で,高学歴の群では,欧米諸国と同様の負荷パターンを示した。)またGH, BP, VTの因子負荷量が,欧米よりも精神的な因子により負荷していた。

(3) 国民標準値の測定

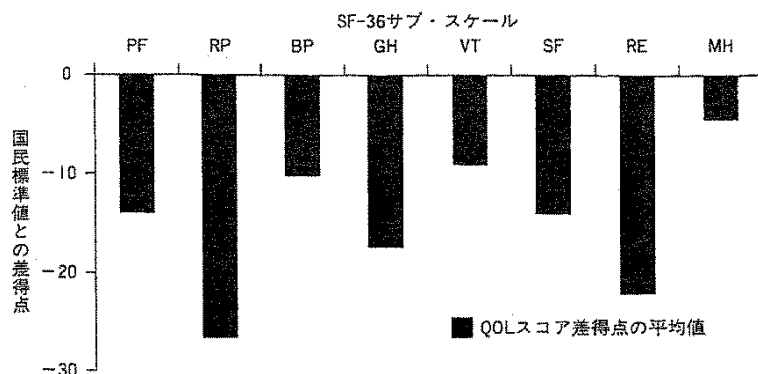
断面的観察研究においては,測定されたQOLスコアがどのような意味を持つものを把握するためには,いわゆる血液検査における正常値と同じような,基準となるスコアが存在することが必要になってくる。しかし,QOLなどの主観特性は文化的な相違の影響など複雑な要素の影響を受けやすいので,国際的な基準値をもうけることは困難である。そのため,SF-36が使用されるそれぞれの国で独自の標準的なスコアが有用となる。

日本におけるSF-36スコアの標準値を明らかにすべく,われわれは1995年10月に自己記入式質問紙表を用いた全国調査を行い,その平均と分布を明らかにすることを試みた。対象となる母集団は日本に在住する16歳以上の住民全てとした。標本抽出に関しては,4,500人に調査を行うことを前提に,一標本抽出地点あたり15人ずつとして300地点を全国に分散させることにした。全国300地点を比例配分して,日本全国の縮図となるような標本抽出を行ったわけである。まず,日本全国を47の地域層に分けた。すなわち,この47の地域層は,都市規模5レベル(12大都市,人口15万人以上の都市,5万人~15万人の都市,人口5万人未満の都市,郡部)と,地域10ブロック(北海道,東北,関東,北陸,東海,近畿,中国,四国,九州,沖縄)から構成されている。この47層に対し,満16歳以上の人口(平成2年の国勢調査による各層の人口データに基づき)で300地点を比例配分し,各層の標本地点数を算出した。このように算出された標本地点数に沿って,各地域層より標本地点を無作為に抽出した。(第一段階の無作為抽出)次に,各標本地点の住民基本台帳から15人の住民が無作為に抽出された。(第二段階の無作為抽出)最終的に全国より4,500人の16歳以上の住民が標本として抽出された。3,410の質問表が回収されたが,そのうち3,395例(75%)がスコアリングとして分析可能なものであった。

この調査では,回答者の属性(年齢,性別,職種,教育歴,婚姻状況,居住状況,年間収入など)や,主要な慢性疾患(計18カテゴリー)

の有無などについてもデータを収集した。この結果、例えば、「心の健康」(MH)を除いて、加齢に従いQOLが低下する傾向があること、慢性疾患の有無と数がほとんど全ての下位尺度と負の関連があることなどが分かっている⁹⁾。この詳細な結果については、SF-36日本語版マニュアルの内容の一部として近日出版する予定である。

図1 透析患者のSF-36尺度得点：国民標準値との比較



III 臨床応用研究に関して

前述のようにして開発され、validateされた日本語版SF-36は、さまざまな臨床研究や、population basedな研究に応用されている。

厚生科学研究のもと、約6,000人の透析患者を対象にした断面的観察研究が行われた。透析患者はSF-36の下位尺度全てにおいて国民標準値と比較して有意にQOLの低下を認められることが報告された。図1に結果の一部を示す¹⁰⁾。また、透析条件、透析方法の種類、原疾患、透析期間、などがQOLとどのような関連があるかについても検討がなされている¹¹⁾。この厚生科学研究の最終年には腎移植患者を対象に断面的観察研究を行っている。この調査の結果から、日本人の腎移植患者のQOLはこれまで欧米で報告されたSF-36を用いた同様の研究の結果とは異なるパターンで低下しており、特に全体的健康観(GH)の低下が著しいことが明らかになった¹²⁾。

また保存期慢性腎不全患者約600人を対象とした縦断的観察研究では、これらの患者が、活力や日常役割機能などのいくつかの下位尺度で、国民標準値の一年間の変化と比較して、有意に速いスピードでQOLの低下を認めていることが報告された。さらに、QOLの低下速度を説明するいくつかの因子のうち、腎性貧血の進行がQOL低下に最も強い影響を与えていることも示唆された¹³⁾。この結果に基づき、腎性貧血の治療が保存期慢性腎不全患者のQOLに与える効果を評価するための無作為化比較試験を実施

し、調査を完了するところである。

上記の腎臓患者に加え、C型慢性肝炎の患者¹⁴⁾、炎症性腸疾患の患者、パーキンソン病の患者、慢性閉塞性呼吸器疾患の患者¹⁵⁾、などいくつかの慢性疾患患者を対象としたQOL研究が実施され、その結果が報告されている。

IV 考 察

QOL尺度の翻訳および異文化間の調整作業は、単なる翻訳と異なり、多段階にわたる作業および多角的な観点からの検討を要する作業である。特に欧米とアジアというような文化が大きく異なる場合では、母国語への変換に際して特段の注意を要することが示唆された。また、尺度の翻訳・異文化間調整と計量心理学的な検証の2つの作業は、これまで独立した作業としてとらえられてきたが、本研究では最初に行った計量心理学的な検討作業の結果が、尺度の内容の翻訳に重要な改変を促した。これはこの2つの作業が、密接に関連し不可分なものであることを示唆している。

計量心理学的検討作業により、SF-36日本語版の信頼性と妥当性を確立することができた。特に因子分析により、日本人の健康を構成する概念が、少なくともSF-36日本語版によって測定された結果では、他国と同様に2因子であると考えられた。鬱スケールによる基準関連妥当性やSF-36を用いた臨床研究の結果などから、日本人も身体的な要素と精神的な要素をかなり

明確に区別してとらえていることは確かなようである。しかし、この2因子への各下位尺度の因子負荷のパターンが他の欧米諸国と全く同一ではなく、特に日常役割機能に関して、日本独特の因子負荷パターンが認められた。この傾向は日本だけに見られるのか、あるいはアジアの文化に共通に見られるのか、どのような背景によるものか、などは不明であり、非常に興味深い。今後さらに検討を加えるべき課題である。

evidenceの強さが低いといわれる断面的観察研究においても、国民標準値を用いることによって、有用な情報が得られる。すなわち、ある一点のQOL値を測定するだけでは、「測定のための測定」に終わってしまう。標準値に比べて、QOLの「どの側面」で、「どの程度異なるのか」が重要な情報を提供する。例えば、前述の腎移植患者を対象とした研究でも、研究を行う前には医師が予測しなかった結果が得られ、医師の視点と患者の視点の食い違いが明らかにされた。あるいは、医師がこの点に気付いていたとしても、これが定量化されなかったため、これらの患者の問題点がしっかりと認識されなかったということがある。QOLを定量的に測定することによって、このような問題点をより明確に認識でき、医療提供者側が医療の受け手である患者の健康感や日常生活機能を理解することができ、ひいては医師-患者関係を改善することにつながることも期待される。またわれわれは国民標準値の研究を3年間連続して行っており、一般住民のQOLの経時的な変化に関しても定量的なデータを得ることができた。このようなデータをレファレンスデータとして用いることによって、前述の保存期慢性腎不全患者の縦断研究のような研究が可能になる。無作為化比較試験(RCT)は最も強いevidenceを提供すると言われているが、倫理的な問題など、わが国では実施困難なことが多く、また外的妥当性に欠けるなどの批判もある。これらの問題点を補うために、quasi-experimentalな研究デザインが提唱されているが、経時的なQOLの変化に関する標準値は、このような研究において貴重なレファレンスデータとして活用されることが期待され

る。しかしながら、このような国民標準値と慢性疾患を有する患者のQOLを比較するときは、厳密に疫学的な立場からすると、source populationが異なる、あるいは測定時期が異なるなど、の点は否定できない。このような点による限界を認識しつつ解析を行う必要がある。

筆者らはQOL研究において、臨床応用研究はQOL研究の3つの大きな柱の1つであると認識している。これまでのQOL研究は、断面的な観察研究が多かったと言えるが、今後は縦断的観察研究が活発に行われることが期待される。1999年11月3～6日にスペインのバルセロナで第6回国際QOL研究学会年次集会が開催されるが、その副題はいみじくも、「Measurement, Validation and Interpretation of Changes in Health Related Quality of Life: Time for a Critical Review」である。あえてここでChangesとしたのは、横断的・観察研究が圧倒的に多かったQOL研究がいよいよ縦断研究の時代に入ったことを示唆している。縦断研究は、疾患を抱える患者のnatural historyに関する研究から、医療政策の改変がもたらす医療裨益者の健康度へのインパクトを測定・評価するbefore-after studyのようなヘルス・サービス研究にいたるまで、これらの研究結果から導き出される結果の臨床的、医療政策的な意義は計り知れない。また、QOL指標が、RCTなどのような臨床研究においてエンドポイントとして使われることも多くなることが予測される。多くの例ではsurrogate endpointであるが、疾患の種類や研究目的によってはprimary endpointにもなりうる。(例：保存期慢性腎不全における腎性貧血治療の評価研究など)

一方、現存の少数項目(回答者への負担を考慮し、50個以下の項目数のものが多い)のQOL尺度の限界の1つとして、個人レベルでの応用はまだ困難であることを認識しておかなければならない¹⁰⁾。このような問題点を克服するために、近年新しい手法を活用した測定方法の開発・研究が進んでおり、これらの研究の展開しだいによっては、将来的にQOL測定の臨床現場における個人応用が可能になることも期待され

る。

最後にSF-36を始めとする健康関連QOL指標は、アウトカム研究においてアウトカム指標として用いられるだけでなく、将来の健康アウトカムや医療資源利用率などの予測因子としても活用されている。さらにpopulation-based health monitoringや特定の慢性疾患のスクリーニングなどへの応用も試みられている事も付け加えておきたい。

お知らせ

(1) SF-36に関する最新の情報は、インターネットのホームページで入手することができます。home pageの<http://www.sf36.com/>をご覧ください。

(2) 国際QOL学会 (ISOQOL) が、1999年11月3～6日にスペイン・バルセロナで、2000年11月18～21日には日本で開催される予定です。詳細は、近日中に発表されます。<http://www.isoqol.org/>をご参照ください。

参考文献

- 1) Tarlov AR, Ware JE, Greenfield S et al. The Medical Outcomes Study, an application of methods for monitoring the results of medical care. JAMA. 1989; 262 [7]: 925-930
- 2) Ware JE, Snow K, Kosinski M, Gandek B: SF-36 Health Survey Manual & Interpretation Guide, The Health Institute, New England Medical Center, 1993
- 3) Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S, Gandek B, Wagner A, Aaronson N, Per Bech, Fukuhara S, Kaasa S, and Ware JE. Translating Health Status Questionnaires and Evaluation Their Quality: The IQOLA Project Approach, Journal of Clinical Epidemiology, 51, 11, 933-944, 1998
- 4) Keller S D, Ware JE, Gandek B, Aaronson N, Alonso J, Apolone G, Bjorner J, Brazier J, Bullinger M, Fukuhara S, Kaasa S, Leplege A, Sanson-Fisher R, Sullivan M, Wood-Dauphinee S: Testing the equivalence of translations of widely-used response choice labels: Result from the IQOLA Project. Journal of Clinical Epidemiology, 51, 11, 933-944, 1998
- 5) Fukuhara S, Bito S, Green J, Hsiao A, and Kurokawa K: Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. Journal of Clinical Epidemiology, 51, 11, 1037-1044, 1998
- 6) Gandek B, Ware JE, Aaronson N, Alonso J, Apolone G, Bjorner J, Bullinger M, Fukuhara S, Kaasa S, Leplege A, Sullivan M: Tests of data quality, scaling assumptions and reliability of the SF-36 in eleven countries: Results from IQOLA Project. Journal of Clinical Epidemiology, 51, 925-932, 11, 1998
- 7) Ware JE, Harris WJ, Gandek B, Rogers WH, Reese PR. Map-R for Windows; Multitrait/Multitem Analysis Program-Revised User's Guide: Boston, MA; Health Assessment Lab, 199
- 8) Fukuhara S, Ware JE, Kosinski M, Wada S, Gandek B, Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey, Journal of Clinical Epidemiology, 51, 11, 1045-1053, 1998
- 9) Bito S, Fukuhara S, Green J, Matsumura S, Kurokawa K: Comorbidity and Health-related Quality of Life. Qual of Life Res. vol. 6, no. 5, 43, 1997
- 10) 福原俊一, 高井一郎, 前田憲志, 平成8年度厚生科学研究長期慢性疾患対策事業(腎不全)「QOLに関する研究」報告書
- 11) Green J, Fukuhara S, Miura Y, Nomoto Y, et al: Method of dialysis and health-related quality of life in Japan: hemodialysis vs continuous ambulatory peritoneal dialysis. Qual Life Res 7: 601, 1998
- 12) Hayashi-Tsuji Y, Fukuhara S, Oshima S, Uchida K, Green J, Takai I, Maeda K: Health-related quality of life among renal transplant patients in Japan, Transplantation (in press), 1999
- 13) Fukuhara S, Rogers W, Marumo F, Maeda T and the Predialysis CRF study group: Health-related quality of life of predialysis patients with chronic renal failure. Nephrology. S507: 1742, 1997
- 14) 福原俊一, 日野邦彦, 加藤孝治, 富田栄一, 湯浅志郎, 奥新浩晃: C型肝炎ウイルスによる慢性肝疾患のHealth Related QOLの測定 肝臓38巻10号pp. 587-595 1997年
- 15) Hajiro T, Nishimura K, Tsukino M, Ikeda A, Izumi T. The St. George's Respiratory Questionnaire is more discriminatory than the SF-36 in assessing health-related quality of life in COPD patients. Am J Respir Crit Care Med 1998; 157(supple): A764.
- 16) McHorney CA, Tarlov AR. Individual-patient monitor in clinical practice: are available health status survey adequate? Qual of Life Res.; 4: 293-307. 1995