

QOL 研究における統計的方法

山岡和枝

国立保健医療科学院技術評価部

Statistical Methods and Related Topics in Health-Related Quality of Life Research

Kazue YAMAOKA

Department of Technology Assessment and Biostatistics, National Institute of Public Health

抄録

近年、患者の評価する健康関連クオリティ・オブ・ライフ (QOL) は、伝統的な疾患の生物学的マーカーに加えて、がんの臨床試験でのアウトカム指標として取りあげられるようになってきた。これらのアウトカム評価では、QOL を用いた分析が行われている。QOL は多様な要素をもつ概念であり、その測定や分析の実践に際し、尺度化をはじめ、要因との関連の分析においてさまざまな推定法や検定法が用いられている。本論ではまずその測定と尺度化の特徴として、尺度構成の観点からプロフィール型と効用型について述べる。また、具体例として日本人を対象として作成した HRQOL20 調査票を紹介する。この調査票はプロフィール型のようにいくつかのスキームについて 20 質問文で構成したものであるが、質問の選択課程での構造分析の結果、ほぼ一次元構造となることがわかり、一次元尺度として尺度化したものである。

次に QOL の論文を数多く掲載している Quality of Life Research 誌を取り上げ、2002.1.~2004.6 の間に掲載された論文計 255 編のうち、何らかの統計的記述のあった QOL に関する論文計 237 編について、用いられた統計的手法を数え上げた。その結果から、比較的手法が多様化し、経時データの取り扱いが多くなってきている傾向が伺えた。そして経時的データの分析では、横断的研究に比べて、用いられている手法も多く、「研究デザインに応じた分析手法を用いた研究」が広まりつつあることが示唆された。

以上を通して、データの取り扱いでの問題点と今後の課題についてまとめた。

キーワード：QOL, 統計的方法, 検定, 推定, 測定, QOLR 誌

Abstract

Recently, patient-reported health-related quality of life (QOL) has received increased attention as an outcome measure in cancer clinical trials, beyond the traditional biomedical markers of the disease. These outcome analyses have used various instruments to assess QOL as evaluation criteria. The QOL is a concept with various elements. In the process of measurement, scaling, and analysis, various statistical methods were used. In this paper, the characteristic feature of profile-based and utility-based scaling methods was discussed from the viewpoints of scale composition. Then, HRQOL20 questionnaire created for Japanese people was introduced. The HRQOL20 questionnaire was composed of 20 questions from several schemas, like profile-based questionnaire. From the pattern analysis of actual responses, one-dimensional structure was obtained. Thus, we used the one-dimensional scale for the measurement of QOL.

Since the Journal of Quality of Life Research has treated many papers related to QOL, it was taken up. Topics about statistical methods were taken up in the recent (2002.1.~2004.6) QOL papers (in total 237 articles which contained statistical description among 255 research and review articles in Volume 11.1 to 13.6). Numbers of the statistical methods were counted. By the analysis, a tendency using diverse statistical techniques was shown and the studies using data of longitudinal and repeated measurements increased. In the analysis of the longitudinal data, much more statistical models were used compared to the case of using cross-sectional data. It was suggested that "researches using the tools of analysis corresponding to the research design" has increased progressively.

Finally problems in the data handling and feature issue were summarized.

Keywords: QOL, statistical methods, statistical test, statistical inference, measurement, Journal of QOLR

1. はじめに

近年、保健・医療・福祉を考える上で、検査値や死亡率、罹患率に加え、健康関連クオリティ・オブ・ライフ (health-related quality of life; 以下、単に QOL と記す) や満足度などの主観的アウトカム指標が用いられ、QOL をより科学的に評価しようという様々な試みがなされている。QOL の評価方法としては、大きく分けて、①臨床検査値などの客観的データに基づく評価、②医師、パラメディカル・スタッフ、家族など第三者による評価、③患者自身による評価、の3通りの方法がある。患者の主観的アウトカムということを考えれば③を考えるのが自然であろう。ここでは③に関連した QOL 研究を取り上げる。このようなアウトカム指標としての QOL の利点としては、医療を受ける側の視点で捉えた健康度が評価できる、疾患の重症度中心の捉え方とは異なった次元での分類が行える、数量化が可能であり科学的取り扱いができるなどがある。しかし一方で、QOL 評価とその問題点として、定義のあいまいさと広義性による内容的妥当性の難しさ、絶対的な評価基準 (gold standard) の欠如による基準関連妥当性の切れの悪さ、最近では SF36 や EORTC などいくつかの標準的な測定手段が開発されているが、その国の医療や文化の違い等が影響する可能性があり、尺度の多文化的同等性¹⁾の検討もまだ課題として残っている。さらに QOL におよぼす要因の多様性 (身体的、心理的、環境など) や経時的データの評価の必要性とともにレスポンスシフト²⁾など反応性の問題についての検討が未だ必要と考える。

このような特徴をもつ QOL 研究では、その統計的取り扱いもさまざまである³⁾。本論の S2 では、QOL の測定の特徴としてプロファイル型と効用型でのスケールについて述べ、例として著者らが関わって開発した HRQOL20 調査票での検討結果を概説する。S3 では、QOL 研究を掲載する代表的な Journal である Quality of Life Research 誌を取り上げ、QOL 研究で用いられている統計的方法を数え上げ、最近の QOL 研究で用いられている方法を検討する。S4 ではデータの取り扱いでの問題点と今後の課題について述べる。

2. QOL の測定の特徴

2.1 プロファイル型と効用型 QOL の測定とスケール

健康関連 QOL の測定のための QOL 調査票はいくつかの捉え方があるが、スケールとしては尺度構成の観点からの分類を取り上げてみる。それは大きく分けて多次元的機能測定を行うプロファイル型 (profile based) と Quality-adjusted life-year など算定するなどの効用測定のための効用型 (utility based) に分類され、前者には SF36⁴⁾、SIP⁵⁾、EUROTC QLQ C-30⁶⁾などがあり、後者には EuroQOL⁷⁾などがあてはまる。プロファイル型のは心理学的尺度法に則ったスケールの信頼性や構成概念妥当性などが検討可能である。プロファイル型では、因子分析結果を基に質問項目の選定を行っているものが多く、プロファイルとしていくつ

かのサブスケール (次元) を構成し、それに総合的な QOL の評価を加えることが多い。たとえば EUROTC QLQ C-30 では、Physical functioning, Role functioning, Emotional functioning, Cognitive functioning, Social functioning などのサブスケールとグローバルスケールとで構成されている⁸⁾。これらのサブスケールは互いに独立なものや相関を持つものも混在しており、単純に足し合わせて QOL の総合得点を出すことは意味がない。そのためにグローバルスケールをサブスケールとは別に尋ねているのである。このようなプロファイル型の質問項目では、たとえば「潜在因子」という形で、因子分析を利用して先のサブスケールを構成する質問項目の分析を行うことが一般的である。また、この際、直交回転や斜交回転が利用されることが多い。一方、効用型ものは初めから Visual Analog Scale (VAS) などを用いて次元尺度として QOL の評価を受けており、サブスケールは考えていない。著者らの開発した HRQOL20 調査票は、数量化Ⅲ類という、もっと単純な項目間の関連構造から質問項目の選択を行っている。当初はプロファイル型としていくつかのスキームに対応する質問を考えたが、最終的に次元構造をもつことがわかり尺度化したものである。これについて次に概説する。

2.2 HRQOL20 調査票

HRQOL20⁹⁾⁻¹⁰⁾は患者の一般的な主観的 QOL を測定するために開発された。この構築にあたっては、パターン分類 (数量化Ⅲ類) の結果を利用した、次元構造に基づく質問項目の選択と得点の求め方に特徴を有する。

HRQOL20 は、QOL の定義を「いきがい」として、心理的要因、身体的要因、環境などに関する質問文で構成された自記式調査票である。質問項目は、大きく身体的機能 (食欲、睡眠、疼痛、苦痛など)、医療環境、人間性に関する環境、精神的・心理的状态、自己の病気に対する認識、という5つの観点 (スキーム) に分けられる。この意味では開始時点の視点はプロファイル型と同じ考え方であり、QOL 研究会 (代表 故林知己夫) において 1985 年より検討を進め、患者および一般人に対する 300 あまりの質問項目の分析を通して作成したものである⁹⁾。

調査票の作成にあたり、次のような国民性調査からみた日本人の特徴¹¹⁾を考慮して項目の選定や選択肢の決定を行った。たとえば義理人情である。これより、人間関係が大きなウェイトを占めることを配慮し、家庭生活や人間関係に関する項目を含めた。また、日本人は中間回答を好む傾向がある点がある。そのため、選択肢は中間を含めた3段階とした。一方、尺度化に関しては次のように行った。数量化Ⅲ類の第1軸および第2軸のカテゴリー値による二次元平面図で、視覚的ではあるが、カテゴリーの配置が positive なものがプラス側に、中間が中央に、negative なものがマイナス側に付置され、全体で U 字型を呈する U 字型構造 (ガットマンのスケログラム・アナリシス¹²⁾のように、例えば U 字のプラス側では、プラス側の反応の多い人のみが反応するような項目

が原点から離れ、マイナス側への反応の多い人が反応するような項目は原点付近に布置することを示す)がみられたことから、これら項目群の内容が一次元で表現できることがわかった。このような場合、全体の構造を表現するように項目を選択すれば少なくとも順序関係は保たれ、U字型構造に布置された全体の情報を、項目を減らしても比較的少ない情報の損失で再現することができる。得点化するには、必ずしもカテゴリー間の距離が等しくない、positive な回答傾向(プラス側)と negative な回答傾向(マイナス側)では疾患により反応が異なる可能性がある、という点を考慮し、HRQOL20ではプラス側得点、マイナス側得点と2種類のスケールを作成した⁸⁾。これらの得点とカテゴリーに付与された数値から得点を求めた場合との相関係数は0.8前後であり、比較的高かった¹⁰⁾。このような一次元構造を呈したのは「日本人とアメリカ人は欧州人に比べて健康状態と社会意識が未分化で、特に日本人は自覚症状と社会的不安が一体化している傾向がある」¹¹⁾という点からも頷ける結果であろう。

HRQOL20 調査票の評価として、信頼性と妥当性に関して検討した結果は次の通りである。信頼性は再テスト法により、健常人(女子学生)に対して1週間間隔で再現性を検討したところ、ピアソンの相関係数はプラス側得点で0.64、マイナス側得点で0.75であった。さらに、内的一貫性としてクロンバックの α 係数をみると、がん患者では0.72~0.82、鍼灸治療群では0.67~0.79、健常人では0.68~0.79であった。他方、判別的妥当性としてグループごとのプラス側得点とマイナス側得点を求め、Kruskal Wallis 検定によってグループ間の相違として検討したところ、男女とも非がん患者でプラス側得点、マイナス側得点とも低く、健常人で高い傾向を示していた(それぞれ $p < 0.001$ 、ただし、男性のプラス側得点を除く)¹⁰⁾。また、肺癌患者172名を対象に併存的妥当性を EORTC QLQ-C30 を用いて検討した結果、グローバルスケール(GL2)との相関係数は0.4~0.6程度であり、特にプラス側得点の相関が高かった。また、EORTC QLQ-C30の各次元との相関傾向もGL2との相関と比較的類似した関連を示していた。加えて、HRQOLが主観的尺度ということから性格特性の影響を受ける可能性があるため、その影響を分析し、影響することを示した⁹⁾。

以上、筆者が関わって行ったHRQOL20の例を示したが、これらの文献で用いた統計的手法としては、t検定、カイ2乗検定、Wilcoxon 順位和検定、Pearson 相関係数、Spearman 相関係数、クロンバックの α 係数、Kruskal Wallis 検定、数量化Ⅲ類、因子分析(探索的、主因子法)などの基本的手法が主である。構造の分析に関しては一般的には因子分析が利用されることが多いが、得点の考え方によっては必ずしも適切でない場合もある。著者らの例の場合では、リッカート・スケールを仮定していなかったことから数量化Ⅲ類(あるいはコレスポネンス分析、対応分析)などを用いて調査票の構造を把握することが有用な方法と考える。ガットマンのスケログラムは典型的な構造を仮定したものであるので、我々としてはHRQOL20は数量化Ⅲ類を用いた構造分析を利用

して構造化を図った。

尺度に関連した分析では、一定のパスあるいは潜在構造を仮定したパス解析あるいは潜在構造分析(latent structure model)、structural equation modeling (SEM)、共分散構造分析などとも呼ばれる相関構造を基にした確認的因子分析(confirmatory factor analysis)もHRQOLの分析において利用されている¹³⁾。そこで、S3では最近のQOL研究で用いられている統計的方法について概観してみた。

3. QOL 研究で用いられている統計的方法

QOL 研究ではさまざまな統計的方法が用いられており、それも時代と共に少しずつ変化してきている。QOL 研究の代表的な Journal として Quality of Life Research 誌があげられるが、ここでは Vol.11, No.1~Vol.13, No.6 (2002年1月から2004年6月)までに掲載された論文計257編のうち、review や総説論文を除いた何らかの統計的記述のあった研究論文計237編について、論文中で取り扱われている統計的手法をその内容ごとに数え上げた。数上げのルールは次の通りである。

まず、237論文すべてについて、掲載された統計的手法を調べた。主として“method”のセクションと図表を確認し、そこに記載されている方法を数え上げた。したがって、本文中にのみ記載された方法は除かれている可能性は否定できない。また、研究デザインにより用いられる統計的方法も異なってくるので、本来は研究デザインと用いられた統計的方法とを詳細に吟味して評価する必要があるが、ここでは手法の種類ということだけに着目して概観するに止めた。分類のカテゴリーは表1のとおりである。

これらの集計結果から、QOL 調査票の開発に関する論文が多いという Journal の性格上、クロンバックの α 係数が利用された論文は全体の33%を占めており、また、全体の半数以上で分散分析(回帰分析も含む)が用いられていたのが特徴的であった。一方、t検定やカイ2乗検定などの手法はそれぞれ全体の30%、22%で利用されていたが、最新号ではむしろ少なくなり、多変量解析の利用が多くなっている傾向が伺えた。ちなみにロジスティック回帰分析は平均9%で利用されていたが、それは2002年に2%であったのが、2004年には13%と増加していた。また、用いられる統計手法の種類も多様化し、Multilevel analysis などの利用もみられた。もちろん、研究デザインが洗練された確認的研究の場合には、単純な検定が大きな意味をもつことになるので、多ければいいというものではない。ここでは論文の詳細の検討は行っていないので、この点については言及できない。

統計的手法が多様化し経時データの取り扱いが多くなってきたのは、ある程度、QOL 評価法が確立し、それを用いた経時変化が分析されるようになったという面もあろう。そして経時的データの分析では、横断的研究に比べて用いられている手法も多く、分散分析などにおいても母数効果モデルを当てはめるだけでなく、変量効果モデルや混合モデルを用いた分析を用いるなど、研究デザインに応じた分析手法を

表 1 QOLR 誌 (2002.1-2004.6) の文献で用いられた統計的手法

PROCEDURE	Total	Year		
		2004	2003	2002
Articles Containing Statistical Description	237	97	81	59
Chi-square/Fisher's Exact / Test of Fit	53	18	21	14
Student's t-test / Paired t-test	70	24	22	24
Pearson Correlation Coefficient	68	28	24	16
Spearman Correlation Coefficient	46	21	13	12
Cronbach's Alpha Coefficient	79	32	23	24
Kappa Coefficient	10	1	4	5
Other Coefficient of Concordance	3	0	3	0
U-test / Wilcoxon / McNemar Test	45	17	16	12
Life Table (Kaplan-Meier)	4	1	1	2
Multiple Comparison	31	8	12	11
Intra-Class Correlation (ICC)	34	17	11	6
ANOVA/ANCOVA/Regression Model (Fixed Effect)	126	56	40	30
Other Regression Model (Random Effect Model, Mixed Model, etc.)	11	3	3	5
Kruskal-Wallis Test	18	7	7	4
Logistic Regression Model	21	13	7	1
Cox-proportional Hazard Model	3	0	0	3
Exploratory Factor Analysis / PCA	28	7	12	9
Confirmatory Factor Analysis / SEM, CSM	11	3	3	5
MANOVA/Multivariate Regression Model	6	2	4	0
Item-Response Theory / Rasch Model	17	4	10	3
Multitrait Analysis / Content Analysis	14	8	4	2
Multilevel Analysis	1	1	0	0
Monte Carlo Simulation	7	3	3	1
Other Methods	10	2	3	5

用いた研究が行われるようになってきたとも考えられる。

4. データの取り扱いでの問題点と今後の課題

統計学的立場から QOL 測定について検討した Cox ら¹⁴⁾も指摘するように、QOL は広く利用される反面、さらに評価尺度として解決していかねばならない課題が残っている。多国間比較や経時データを評価することが多くなってきた

現在では、QOL 調査票の信頼性・妥当性の検討として残された課題は、多文化的同等性¹⁾や、天井効果 (ceiling effect)、床効果 (floor effect)、レスポンスシフト²⁾の問題などを含めた反応性の検討であり、これらは未だ十分とはいえ、今後の発展が期待されよう。

さらに、対象の取り方など研究デザインに応じて multilevel analysis の適用など、最近比較的利用しやすいソ

フト (たとえば MLwin など) も開発されてきており, 新しい分析手法を活かし, データの性質に応じた分析を深めて行くことも意味があろう. 特に経時的なデータを取り扱う場合には欠損値の問題が生じる可能性が高い. このような不完全データの取り扱いも大きな課題である. Missing at Random (MAR) などの仮定をおきロジスティック回帰分析等を利用した multiple imputation などを含めた不完全データの解析^{15),16)}も行われている. しかし, 不完全データの取り扱いは簡単ではなく, 短絡的に MAR の仮定をおくことはできない. その仮定の吟味やデータの性質などについての検討が必要であり, 今後, 経時的データを取り扱う上での key point の1つとなろう. 今後はますます欠損値の取り扱いなども含めた, データの性質に応じた分析手法が求められるようになり, 結果の解釈と限界, 価値づけ, 評価をどのように行うかという点も含め, 注意深く検討していく必要がある.

QOL 研究は, 個々の患者の QOL を評価して患者の治療へ, あるいは保健政策などにフィードバックさせることをその目的の1つとしている. そのため診断指標としての QOL 調査票の有用性について, QOL の目的と定義に即して, 治療や政策等を行った時に, どのようなメリットが得られたかという結果までを含めて評価することが肝要と考える.

文献

- 1) Hui SL, Rosenberg SH. Multivariate slope ratio assay with repeated measurements. *Biometrics* 1985; 41(1): 11-8.
- 2) Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med.* 1999 Jun; 48(11): 1531-48.
- 3) 山岡和枝. 健康関連 QOL におよぼす性格特性の影響. 柳井晴夫, 岡太彬訓, 繁榎算男, 高木広文, 岩崎学, 編. 多変量解析実例ハンドブック. 東京: 朝倉書店; 2002. p.226-39.
- 4) Ware JE, Sherbourne CD. The MOS 36-item Short-form Health Survey (SF-36). 1. Conceptual framework and item selection. *Medical Care* 1992; 30: 473-83.
- 5) Bergner M, Babbitt RA, et al. The sickness impact profile: development and final revision of a health status measure. *Medical Care* 1981;19:787-8.
- 6) Aaronson NK, Ahmedzai S, et al. The European Organization for Research a Treatment of Cancer QLO-C30: A quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; 85: 365-76.
- 7) The EuroQOL Group. EuroQOL-a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199-208.
- 8) Yamaoka K, Hayashi F, et al. A Japanese version of the questionnaire for quality of life measurement. *Ann Cancer Res The* 1994; 3: 45-53.
- 9) Yamaoka K, Shigehisa, et al. Influence of personality on quality of life measurement. *QOLR* 1998; 7: 535-44.
- 10) Yamaoka K, Ogoshi K, et al. Validity of the Japanese version of the questionnaire for quality of life measurement (QOL20). *IMD* 1998; 5: 23-9.
- 11) 林知己夫, 櫻庭雅文. 数字が明かす日本人の潜在力. 東京: 講談社 SOPHIA BOOK: 2002.
- 12) Guttman L. The quantification of a class of attributes. A theory and method of scale construction. In: Horst P, edited. *The prediction of personal adjustment*. New York: Social Science Research Council; 1941.
- 13) Staquet MJ, Hays RD, et al. *Quality of life assessment in clinical trials*. Oxford: Oxford University Press; 1998.
- 14) Cox DR, Fletcher AE, et al. Quality-of-life assessment: Can we keep it simple? *J R Statist Soc A* 1992;155 Part 3: 353-93.
- 15) Rubin DR. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons; 1987.
- 16) 岩崎学. 不完全データの統計解析. 東京: エコノミスト社; 2002.